CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 120, No. 3 September 2006

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

Discrimination of near-field infrasound sources based on time-difference of arrival information

Curt A. L. Szuberla, Kenneth M. Arnoult, and John V. Olson
Geophysical Institute, University of Alaska Fairbanks, Fairbanks, Alaska 99775-7320
cas@gi.alaska.edu

Abstract: A computationally efficient method for discriminating between near- and far-field infrasound sources using array time-difference of arrival (TDOA) information is described. Rather than assess wave-front curvature, the discriminant quantifies the statistical departure of TDOA information from that of a plane wave passing the array. Since the method constrains neither the functional form nor the amplitude characteristics of a signal it is suited for discrimination of signals across large-aperture infrasound arrays. Experimental results confirm theoretical predictions to a range of order ten array apertures. The discriminant is applied to data from an Antarctic infrasound array.

PACS numbers: 43.60.Fg, 43.28.Dm [DKW]
Date Received: April 19, 2006 **Date Accepted:** August 14, 2006

1. Introduction

At present, more than half of the 60 planned infrasound arrays of the International Monitoring System (IMS) of the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) are operational.[1] These large aperture—$\mathcal{O}$(1 km)—arrays are designed to monitor and aid in the detection of impulsive events at large distances. A major challenge facing analysts of IMS infrasound array data is the ubiquitous presence of natural and manmade near-field sources, which are not generally of interest to the CTBTO. The presence of near-field signals exhibiting high signal-to-noise and spatiotemporal correlation can effectively saturate event detection algorithms with false alarms. There exists a distinct need for a discriminant between local and regional sources within the infrasound band.

In an effort to ameliorate this situation, a theoretical method is developed for discriminating between near- and far-field sources using time-difference of arrival (TDOA) information. The discriminant is coded in computationally efficient linear algebraic form, requiring only knowledge of TDOA information and the array geometry for calculation. The curvature of the wave front[2] need not be estimated. Since TDOA information is estimable via generalized cross correlation in the time domain, the method constrains neither the functional form, nor the amplitude characteristics of the signal present at the array. This is a distinct advantage for infrasound array data, as terrain and vegetation often cause significant departures from $1/r$ amplitude characteristics and infrasound signals are not often well represented by simple functional forms.

Recent work[3] has shown that IMS-type arrays may be used to accurately localize geophysical and manmade sources at very close range, thus the discriminant may also be used to eliminate long-range signals from consideration. Ice cliff avalanches and volcanic activity in the vicinity of the IMS array (I55US) at Windless Bight, Antarctica during the 2002–2003 austral summer are used to validate the performance of the discriminant. Sources that lie within a range of order ten times the array aperture are accurately characterized by the discriminant, as predicted theoretically. As a result of this demonstration, the discriminant has been put into use as a regular part of automated data processing for the infrasound arrays operated by the University of Alaska Fairbanks.

In this study the discriminant is applied to signals detected at the IMS array at Windless Bight, Antarctica (I55US). I55US is an eight-sensor infrasound array comprising Chaparral

Physics microphones at the vertices of a pentagon of 1-km radius and an inner triangle of 100-m radius. This original geometry has been distorted slightly since the array was originally constructed owing to the movements of the Ross Ice Shelf. Pressure data are sampled at 20 Hz from each of the microphones.

2. Near-field discriminant

The development of a near-field discriminant begins with the description of an idealized far-field signal incident on a d-dimensional array of n sensors. Using Dirac's *bra-ket* notation, TDOA information for the array is described by the vector $|\tau\rangle$. This vector represents the $N = n(n-1)/2$ unique intersensor time delays that comprise the TDOA information corresponding to a planar signal's passage across the array. The array coordinates are encoded in a $[N,d]$ matrix $\mathbb{X}$. A plane-wave arrival at the array is described by the model

$$|\tau\rangle = \mathbb{X}|\mathrm{s}\rangle + |\epsilon_\tau\rangle, \tag{1}$$

where $|\mathrm{s}\rangle$ is the slowness vector and $|\epsilon_\tau\rangle$ is some error introduced by noise. The vector $|\tau\rangle$ is estimated by generalized cross correlation and $|\mathrm{s}\rangle$ is solved for by minimizing the quantity $\langle\varepsilon_\tau|\varepsilon_\tau\rangle$ in the least squares sense. This solution of Eq. (1) via optimization is given by

$$|s\rangle = \mathbb{C}^{-1}\mathbb{X}^\dagger|\tau\rangle\,, \tag{2}$$

where $\mathbb{C}=\mathbb{X}^\dagger\mathbb{X}$. Under the assumption that the errors in the time delays are normally distributed and uncorrelated, then Eq. (2) is the maximum likelihood estimator of $|\mathrm{s}\rangle$ and exactly satisfies the Cramér-Rao bound.[4]

A statistical measure of the variance associated with the TDOA information $|\tau\rangle$ has been shown[5] to be

$$\sigma_\tau = \sqrt{\frac{\langle\tau|(\mathbb{I}-\mathbb{R})|\tau\rangle}{N-r}}, \tag{3}$$

where $\mathbb{I}$ is the identity matrix, $\mathbb{R}=\mathbb{X}\mathbb{C}^{-1}\mathbb{X}^\dagger$, and r is the rank of the idempotent matrix $\mathbb{R}$. Using this, Szuberla and Olson[6] demonstrated two properties of σ_τ for infrasound arrays. First, Eq. (3) represents the degree to which $|\tau\rangle$ differs from a physically realizable plane wave, modeled by Eq. (1), incident upon the array. Second, although σ_τ appears to be strictly a function of the TDOA information and the array geometry, the TDOA information in $|\tau\rangle$ is a function of azimuth, which leads to the correspondence $\sigma_\tau=\sigma_\tau(\theta)$.

The construction of a discriminant begins with an examination of the behavior of σ_τ as a function of range. These results are depicted in Fig. 1 for an ensemble of synthetic, noise-free $|\tau\rangle$ at the I55US array. When the array data are sampled at $f_s=20$ Hz, the 95% confidence limits of σ_τ versus source range are shown as a light gray band. Similarly, the dark band represents the limit of an infinite sampling rate. Dark lines indicate the ranges for which $\sigma_\tau=1/f_s$. Note that as the magnitude of σ_τ decreases from $1/f_s$ the character of the finite and infinite sampling results part company. The median value of σ_τ for the finite sampling case is shown in the figure as a white line.

The width of the band for continuous sampling in Fig. 1 is governed by spatial sampling (array geometry). In contrast, the finite sampling band in Fig. 1 is governed by spatiotemporal sampling. At close range the spatial aspect dominates, while in the far-field the temporal (sample rate) dominates. These characteristics are seen in Fig. 2, which depicts the behavior of σ_τ as a function of azimuth for an ensemble of synthetic, noise-free $|\tau\rangle$ at the I55US array at a set of fixed ranges. In each panel $\sigma_\tau(\theta)$ is normalized by the mean of σ_τ across all angles at that range. In the near-field (Fig. 2, left plot) the outer, fivefold, geometry of I55US dictates the appearance of σ_τ as a function of angle for both finite and infinite sampling. For more planar signals, arising from far-field sources (Fig. 2, right plot), σ_τ is no longer a strong function of angle in either sampling regime. Note that the "noise" in $\sigma_\tau(\theta)$ is not the result of a contaminated signal, since all the calculations were done in the noise-free case; rather, the effect is due

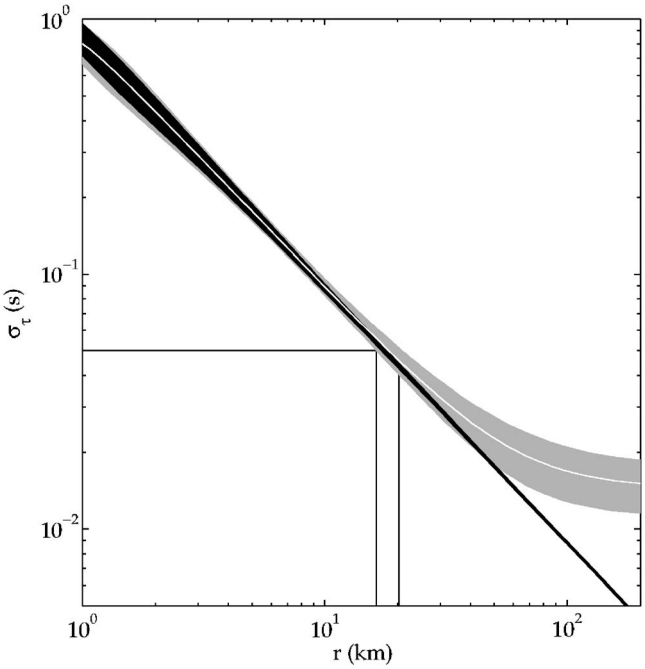


Fig. 1. Theoretical plot of σ_τ vs range for the I55US array. The light gray band depicts the 95% confidence limits of σ_τ for the array sampled at f_s=20 Hz; similarly, the dark band represents the limit of an infinite sampling rate. A white line indicates the median value of σ_τ for the array sampled at 20 Hz. Dark lines indicate the ranges for which $\sigma_\tau = 1/f_s$. The width of the dark band is strictly governed by spatial sampling; whereas, the width of the light band is the result of spatiotemporal sampling. For finite sampling, at close range ($\sigma_\tau > 1/f_s$), array geometry dominates the behavior of σ_τ; in the far-field ($\sigma_\tau < 1/f_s$), the temporal sample rate dominates.

to finite sampling of the data, which leads to a truncated representation of $|\tau\rangle$. Further, $\langle\sigma_\tau\rangle$ for finite sampling becomes greater than that for the infinite sampling case. For $\sigma_\tau \approx 1/f_s$ (center plot) there is a transition between the effects of spatial and temporal sampling. It is in this region, depicted by the dark lines of Fig. 1, that σ_τ forms the basis of a discriminant between near- and far-field sources.

For an array of sensors sampled at f_s a dimensionless measure of the deviation from signal planarity across the array can be constructed as

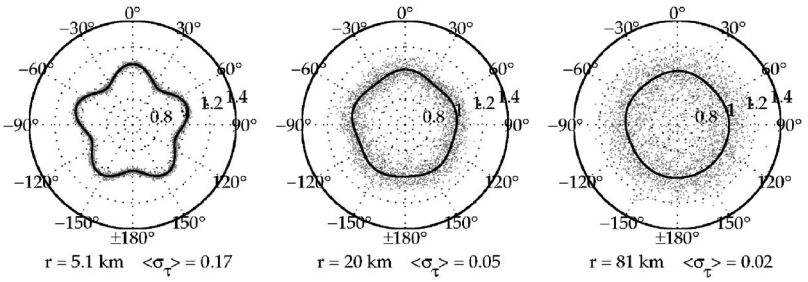


Fig. 2. Normalized polar plots of magnitude $\sigma_\tau(\theta)$ at ranges of 5, 20, and 80 km for the I55US array. In each panel the values of $\sigma_\tau(\theta)/\langle\sigma_\tau\rangle$ are depicted at the given range, where $\langle\sigma_\tau\rangle$ is the mean of σ_τ at that range. The solid lines represent the infinite sampling case, while the discrete dots represent f_s=20 Hz sampling in 0.1° increments. At close range (left plot) the outer, fivefold, geometry of I55US dictates the appearance of $\sigma_\tau(\theta)$ for both finite and infinite sampling. In the far-field (right plot), σ_τ is no longer a strong function of angle in either sampling regime. For $\sigma_\tau \approx 1/f_s$ (center plot) there is a transition between the effects of spatial and temporal sampling.

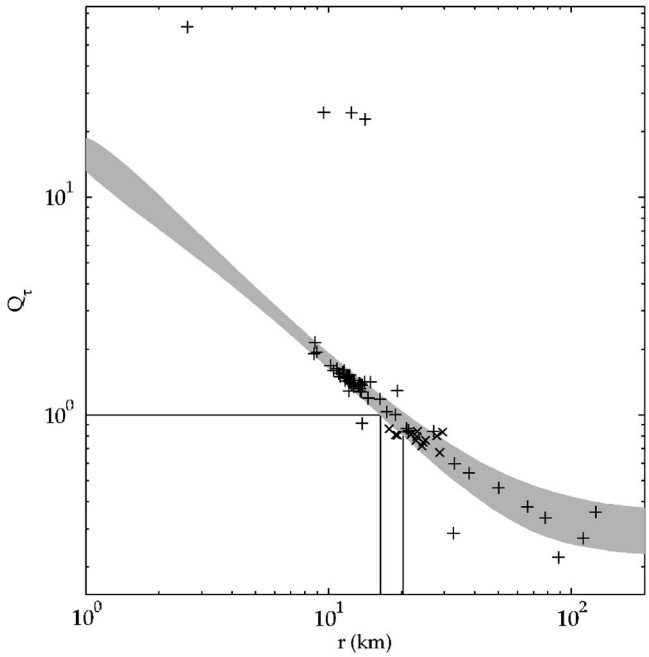


Fig. 3. Superposition of actual Q_τ data and theoretical noise-free predictions for the I55US array. The data are drawn from high signal-to-noise ratio signals detected during the 2002–2003 austral summer at Windless Bight, Antarctica. The light band and dark lines depict the 20 Hz sampling case as in Fig. 1. Mt. Erebus signals are shown with the (×) symbol and Ross Is. signals are shown with the (+) symbol.

$$Q_\tau = \sigma_\tau f_s. \tag{4}$$

It is this measure that forms a discriminant between near- and far-field sources. According to the formulation, a noise-free planar arrival at the array will give, identically, $Q_\tau = 0$. For signal sources close to an array $Q_\tau > 1$, while far-field sources will give $Q_\tau \to 0$.

3. Application and discussion

From November 2002 to January 2003, 76 signals exhibiting high signal-to-noise ratio (SNR $\gtrsim 15$ dB) were detected at the I53US array. The detections were all the result of an automated, correlation-based algorithm that is continuously applied to the I53US data. Follow-on analysis showed that these signals were associated with Mt. Erebus ($r \approx 25$ km) volcanic activity, Vee Cliffs ($r \in [8,30]$ km) ice and snow avalanches, and various glacial and sea ice sources ($r > 10$ km). All of the high SNR events during the period of study were selected without regard to frequency content or wave-form type in an effort to validate the theoretical predictions shown in Fig. 1. The validation is depicted in Fig. 3, where each of the events is superposed with the 20-Hz band of Fig. 1. Each of the range estimates for the individual events in Fig. 3 are the result of a source localization technique[3] (distinct from the Q_τ approach of this study in that it represents a TDOA information optimization procedure to estimate source locations) and not ground truth data. Subsequent experience with infrasound data from arrays in Alaska and Antarctica has shown that $Q_\tau \approx 1$ is a robust discriminant, one which reliably defines the boundary between near- and far-field for large-aperture infrasound arrays. For the I55US array $Q_\tau \approx 1$ when the source is located roughly ten array apertures ($r_{\text{source}} \in [15,20]$ km) distant (shown by the dark lines in Fig. 3). Coincidentally, this is roughly the limit of reliable near-field localizations via the algorithm used to estimate range in the figure.[3]

For arbitrary arrays, the range at which $Q_\tau = 1$ is trivially proportional to sample rate and it can be shown that it is also proportional to the square of the maximum aperture of the array parallel to the signal propagation path. Hence, the technique is ideally suited to the current generation of large-aperture IMS arrays. The discriminant is also applicable to smaller, portable arrays if the sample rate is increased appropriately.

The use of Q_τ as a range discriminant is not without limitations. While it is designed to discriminate between near- and far-field sources, it is not useful as a precise localization tool. As SNR decreases, the assumptions underlying the estimation of σ_τ become suspect. Even in the high SNR case there are several outliers shown in Fig. 3. The presence of the near-field outliers is likely due to misestimation of source range from corrupt TDOA information. Such corruption may arise from a variety of natural phenomena, including multipath, multiple signals and coherent noise fields; however, the exact cause for the near-field outliers of Fig. 3 is unknown. The far-field outliers of Fig. 3 are the result of insufficient range resolution in the localization algorithm for finite f_s.[3] Ground turbulence may affect the planarity of a signal across the array; however, even if the Gaussian, uncorrelated noise assumed to exist in $|\tau\rangle$ is replaced by pink noise, the results given here do not substantially change. As is often the case, simple bandpass filtration is sufficient to restore a reasonable facsimile of the assumptions underlying the estimation of σ_τ.[6]

4. Conclusions

A method for discriminating between near- and far-field sources has been detailed. The calculations are straightforward linear algebraic forms which generate little additional computational overhead. The bulk of the computational effort in arriving at Q_τ is spent estimating the TDOA information, which is already required to determine source azimuth and speed in the far field,[6] and also geographic location in the near-field.[3] While CTBTO-oriented analysts would prefer to reduce near-field clutter ($r < 100$ km), experience with the two University of Alaska Fairbanks/IMS arrays has shown that the Q_τ technique is most useful for discriminating sources at a range of roughly ten array apertures ($r \in [15, 20]$ km).

The technique has been put into use for all of the automatic data processing performed at the IMS arrays I53US and I55US. Since I53US is located nearby a population center, Fairbanks, Alaska, it has provided numerous events for which the ground truth is known. Unlike I55US, the sources for which ground truth is available at I53US do not cluster around the critical range at which $Q_\tau \sim 1$. The sources are either on campus ($r \sim 2$ km) or at larger distances (e.g., mine signals at $r \gtrsim 25$ km). While not conclusive, in these specific cases the I53US results validate the extremes of Fig. 3. Further work is required to study the discriminant at different IMS sites, each of which will have a different, characteristic noise field and topography. Presently, the discriminant is applied to infrasound data derived from large-aperture arrays; however, its development is quite general and a study of its utility in higher frequency, small-aperture applications is underway.

Acknowledgments

The authors thank the reviewers for their cogent and insightful comments.

References and links

[1]Complete information on the Prepatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) and the various monitoring technologies may be found at ⟨http://www.ctbto.org⟩

[2]B. G. Ferguson, "Variability in the passive ranging of acoustic sources in air using a wave-front curvature technique," J. Acoust. Soc. Am. **108**, 1535–1544 (2000).

[3]J. V. Olson, K. Arnoult, and C. A. L. Szuberla, "A method for locating nearby sources," J. Acoust. Soc. Am. **117**, 2610 (2005).

[4]S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall, Upper Saddle River, NJ, 1993).

[5]C. R. Rao, *Linear Statistical Inference and Its Applications*, Wiley Series in Probability and Mathematical Statistics, 2nd ed. (Wiley, New York, 1973).
[6]C. A. L. Szuberla and J. V. Olson, "Uncertainties associated with parameter estimation in atmospheric infrasound arrays," J. Acoust. Soc. Am. **115**, 253–258 (2004).

High-frequency measurements of blast wave propagation

Alexandra Loubeau and Victor W. Sparrow
Graduate Program in Acoustics, The Pennsylvania State University, 202 Applied Science Building, University Park, Pennsylvania 16802
aloubeau@gmail.com, vws1@psu.edu

Larry L. Pater
Construction Engineering Research Laboratory, U.S. Army Engineer Research and Development Center, 2902 Farber Drive, Champaign, Illinois 61821
Larry.L.Pater@erdc.usace.army.mil

Wayne M. Wright
Applied Research Laboratories, The University of Texas at Austin, P.O. Box 8029, Austin, Texas 78713-8029
wwright@arlut.utexas.edu

Abstract: Blast wave propagation measurements were conducted to investigate nonlinear propagation effects on blast waveform evolution with distance. Measurements were made with a wide-bandwidth capacitor microphone for comparison with conventional 3.175-mm (1/8-in.) microphones with and without baffles. It was found that the 3.175-mm microphone did not have sufficient high-frequency response to capture the actual rise times in some regions. For a source of 0.57 kg (1.25 lb) of C-4 plastic explosive, the trend observed is that nonlinear effects steepened the waveform, thereby decreasing the shock rise time, up to a range of 50 m. At 100 m, the rise times had increased slightly.

PACS numbers: 43.28.Mw, 43.25.Cb [VO]
Date Received: May 18, 2006 **Date Accepted:** August 14, 2006

1. Introduction

Much work has been done on blast wave research in the past century, including structural response to blasts and blast effects on human hearing. However, blast wave measurements continue to be important for assessing the impact on humans and wildlife. The finite-amplitude pressure and the transient nature of blast waves pose challenges for obtaining accurate measurements. In particular, resolving the shock structure requires special instrumentation.

Nonlinear steepening occurs in finite-amplitude waves generated by blasts because local wave speed increases with wave amplitude, and thus peaks of the waveform travel faster than troughs.[1] The interaction of nonlinearity and dissipation controls the rise portion of the shocks.[2–4] In particular, the combined effects of dissipation and dispersion due to molecular relaxation in air, which are more pronounced at higher frequencies, tend to increase shock rise times. Rise time is defined in this paper as the time it takes for the pressure to rise from 10 to 90 % of the maximum amplitude.

Work by Ford *et al.*[5] emphasized the importance of measuring blasts at medium distances to enhance understanding of blast waveform propagation in the transition area between the close-range region of very high pressure and the linear acoustics zone much farther away. Ford *et al.* acquired extensive data on blast wave propagation over concrete, grass, and water for small unconfined charges of plastic explosive. However, the instrumentation was insufficient to capture the short rise times of the waves. Another experimental study of blast noise that was performed at the U.S. Army's Fort Drum facility in New York was recently documented in a series of articles, but no high-frequency equipment was used.[6] In other research, wide-

bandwidth capacitor microphones have been used to measure intense acoustic impulses from sparks[7–9] and ballistic shock waves,[10] but no accounts of their use in measuring medium-range blast waves from explosives have been reported previously.

This paper describes blast wave propagation experiments conducted jointly by the U.S. Army and Penn State in November 2005. The main goal of the experiments was to capture the shock structure accurately using high-frequency instrumentation.

2. Microphone Equipment

Commercial 3.175-mm (1/8-in.) pressure microphones and a wide-bandwidth capacitor microphone were used to measure the blast wave. In addition, microphone baffles were used with some of the 3.175-mm microphones.

2.1 3.175-mm (1/8-in.) microphones

The 3.175-mm pressure microphones have a low sensitivity which allows for their use in finite-amplitude measurements. The dynamic range extends to 178 dB re 20 μPa with 3% distortion.[11] In addition, the nominal frequency response extends to 140 kHz. These microphones were chosen because their size is small compared to a wavelength for frequencies up to 100 kHz. They therefore have superior high-frequency response compared to 6.35-mm (1/4-in.) and 12.7-mm (1/2-in.) microphones. Typical preamplifiers[12,13] and power supplies[14] compatible with these condenser microphones were used in the blast experiments.

2.1.1 Unbaffled configuration

Unbaffled microphones were oriented with the microphone axis perpendicular to the blast wave direction; this will be referred to as the 90° orientation. This orientation gives a more accurate pressure measurement than a 0° orientation (microphone axis parallel to wave direction) because it minimizes diffraction effects.[15] All protection grids were removed to avoid any undesired acoustic resonances.[15]

2.1.2 Baffled configuration

Microphone baffles were also used to reduce diffraction artifacts of the microphone housing on the rise portion of the blast waves. The baffles used in the experiments were 20.32-cm (8-in.) squares, and the microphones fit flush with the front of the baffle. The microphones and baffles were oriented at 0° from the blast. Assuming a plane wave, a rigid circular baffle, and measurement at a point, there is pressure doubling at the microphone due to the superposition of the incident and reflected waves. The diffracted wave is delayed according to the radius, out of phase, and half the amplitude.[16] For the square baffles used in the experiments, the travel time from the midpoint of the baffle edge to the center is associated with the first arrival of the diffracted wave. The diffracted wave would first arrive with a delay of approximately 296 μs. Therefore, even though the blast waveform was distorted due to the finite size and square shape of the baffle, the rise portion was unaffected.

2.2 Wide-band microphone

The wide-band microphone and preamplifier[9,17,18] used in this study has a high-resonance frequency and therefore has a better high-frequency response than a conventional 3.175-mm microphone. This type of broadband electrostatic transducer, or "solid-dielectric," generally has a flat frequency response to above 500 kHz and can measure a rise time of less than 0.5 μs under ideal laboratory conditions. With the particular microphone used in the experiments, rise times of less than 1 μs were recorded for an acoustic impulse from a spark source in the laboratory.

The particular microphone used in the experiments had an active area diameter of 1.27 cm (0.5 in.) embedded in an insulator with a diameter of 3.81 cm (1.5 in.). This insulator acted as a built-in baffle when the microphone was oriented at 0° from the blast. Therefore a wave diffracted from the edge would arrive with a delay of 55.5 μs. Because this was outside the range of expected shock rise times, the rise portion of the waveform was not affected, and an

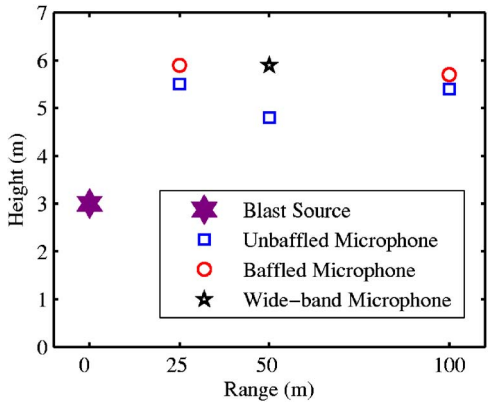


Fig. 1. (Color online) Experimental setup detailing relative locations of blast source and microphones. The 50 m baffled microphone is omitted because its measurements were unusable.

additional exterior baffle was not needed.

3. Experimental setup

Experiments were performed with the U.S. Army at the Edgewood site of Aberdeen Proving Grounds in Maryland. The explosive used was 0.57 kg (1.25 lb) of unconfined C-4 plastic explosive. Charges were detonated by a blasting cap at a height of 3 m. The experimental setup is shown in Fig. 1 . The 3.175-mm microphones were mounted on poles at distances of 25, 50, and 100 m from the blast. The baffled 3.175-mm microphones were colocated at the same distances of 25, 50, and 100 m; however, the 50 m baffled microphone failed. The single wide-band microphone was placed at 50 m. The microphones were mounted nominally 5 m above ground to minimize contamination from ground reflections. At this height, the time delay of the ground-reflected wave associated with the path length difference between the direct and reflected waves was sufficiently long to not affect the shock.

Blast waveforms were recorded on a digital oscilloscope at a sampling rate of 1 MHz. This high sampling frequency was chosen to give good resolution at the shock. Recording equipment was situated so that cable[19] lengths were minimized where possible. For example, the shortest cable lengths, approximately 22 m, were connected to the 50-m site. Minimizing cable lengths was a precaution taken to avoid any undesired signal distortion due to possible reflections in the cable.[20]

The experiment was conducted over two days, during which time the weather conditions varied. However no correlation was evident between the measured rise times and the measured local meteorological conditions.

4. Results

The measured waveforms for all 12 blasts were consistent for the 25 and 50 m sites. However, the waveforms measured at 100 m varied considerably. The shape of some peaks and the existence of double peaks in some waveforms suggest that turbulence likely affected these waveforms recorded at 100 m.[21]

Typical blast waveforms measured at 25, 50, and 100 m are presented in Figs. 2–4 . The asterisks denote the 10 and 90 % amplitude points used to calculate the rise time t_{rise}. The waveforms were measured at 1 μs intervals, and an additional 0.1 μs precision in the rise portion was deduced from interpolation.

For the particular blast shown in Fig. 2, the first 50 μs of the baffled and unbaffled waveforms are compared. The baffled rise time of 19.6 μs is nearly the same as the unbaffled rise time of 21.1 μs. Because of the pressure doubling at the baffled microphone, explained above in Sec. 2.1.2, the baffled response was halved. However, the peak amplitude does not

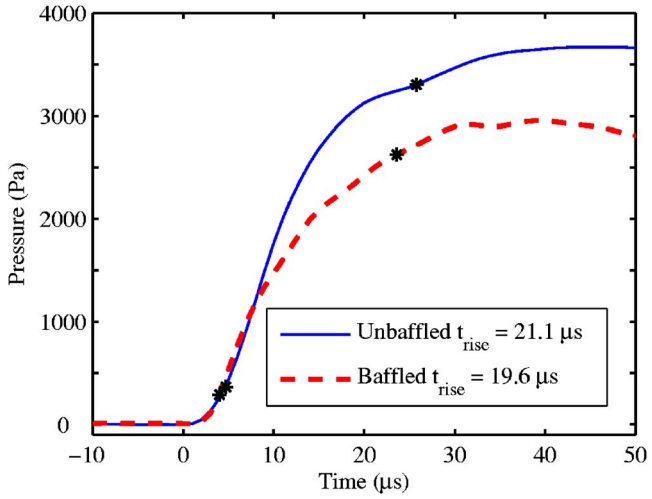


Fig. 2. (Color online) Typical waveforms measured at 25 m with unbaffled and baffled microphones. Asterisks denote the 10 and 90 % amplitude points used to calculate the rise time t_{rise}.

match that of the unbaffled microphone because the pressure doubling assumption is approximate. The experiments involved a square baffle of finite area. As a result, pressure doubling occurs only for frequencies with wavelengths that are small compared to the baffle size.

Because nonlinear steepening leads to decreased rise times, a decrease in rise times at 50 m was expected. The rise times of the waveforms at 50 m were indeed shorter. The experiment was designed to compare all three microphone types at 50 m. Unfortunately, as mentioned earlier, the baffled microphone at 50 m failed, and the only useful comparison at this location is between the unbaffled microphone and the wide-band microphone. Figure 3 shows that the wide-band response is clearly steeper than the unbaffled response. The rise time from the wide-band microphone was 4.5 μs, shorter than the unbaffled rise time of 6.1 μs.

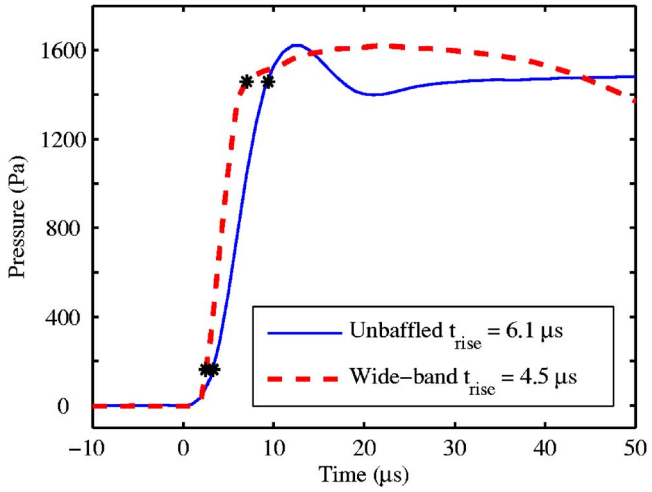


Fig. 3. (Color online) Typical waveforms measured at 50 m with unbaffled and wide-band microphones. Asterisks denote the 10 and 90 % amplitude points used to calculate the rise time t_{rise}. The maximum amplitude from the wide-band microphone is set to be the same as the maximum amplitude of the unbaffled response.

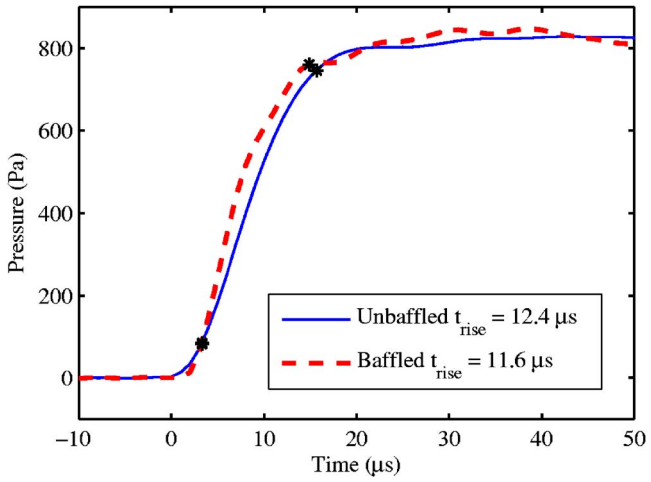


Fig. 4. (Color online) Typical waveforms measured at 100 m with unbaffled and baffled microphones. Asterisks denote the 10 and 90 % amplitude points used to calculate the rise time t_{rise}.

In Fig. 4, the 100 m unbaffled and baffled waveforms are similar, and the rise times were 12.4 and 11.6 μs. These rise times are longer than at 50 m. It is interesting to note that the difference in peak values observed at 25 m due to the pressure doubling approximation is not present in the 100 m measurement. The cause of this difference at 100 m is unknown.

So far, results for one representative waveform have been presented. Figure 5 presents the rise times as a function of range for all measurements. The rise times of the waveforms measured with unbaffled 3.175-mm microphones at 25 m varied from 17.5 to 21.3 μs. The baffled responses at 25 m had nominally the same rise times, ranging from 17.9 to 21.9 μs. The average shock overpressure was 3328.4 Pa.

The rise times of the waveforms measured with unbaffled microphones at 50 m were shorter, varying from 5.9 to 6.2 μs. At 50 m, the average shock overpressure was 1481.9 Pa.

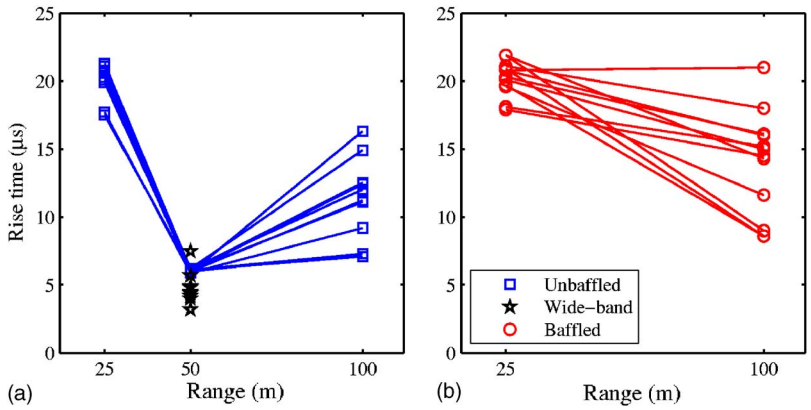


Fig. 5. (Color online) Comparison of rise times as a function of range computed for blast waves measured with different microphone types. (a) Rise times for unbaffled (25, 50, 100 m) and wide-band (50 m) microphones. At 50 m the square markers are located at the intersection of the solid lines. For clarity, the measured rise times at 50 m for the unbaffled microphone were 5.9, 6.2, 6.2, 6.0, 6.1, 5.9, 6.0, 6.1, 6.1, 6.0, 6.1, and 6.1 μs, and the wide-band microphone measured rise times of 4.4, 4.5, 3.2, 4.2, 4.8, 4.5, 4.9, 4.5, 4.5, 7.5, 4.0, and 5.7 μs. (b) Rise times for baffled microphones (25, 100 m). In both (a) and (b) the solid lines merely connect measurement points for individual blast events and do not represent additional data. For (b), there were no useful measurements obtained at 50 m.

The small variations in measured rise times at 50 m seem to indicate that a 6 μs rise time was the limit of the 3.175-mm microphone system in the current configuration. Therefore one cannot be sure if the true rise time of the wave was recorded.

It is because of this microphone limitation that a wide-band microphone was also used. At 50 m, the wide-band microphone measured rise times of 3.2 to 7.5 μs. As shown in Fig. 5, rise times from the wide-band microphone were less than 6 μs for all blasts except one. These rise times are within the capabilities of the microphone, so it is likely that these were the true rise times of the blast waves.

At 100 m, the rise times of the recorded waveforms were longer, with the unbaffled rise time measurements ranging from 7.1 to 16.3 μs and the baffled rise time measurements ranging from 8.6 to 21 μs. The average shock overpressure was 680.5 Pa. The differences between the unbaffled and baffled rise times at 100 m are probably due to some combination of the microphone housing diffraction effect and localized atmospheric turbulence. Overall, the experimental results show that between 50 and 100 m nonlinearity began to lose strength, and losses began to dominate. However, at 100 m the rise time was still shorter than at 25 m.

5. Conclusions

The blast wave measurements confirm that out to a distance of 50 m, where the sound pressure level (SPL) was around 154 dB, nonlinear effects had steepened the initial rise of the blast waveform. This resulted in a decrease in shock rise time and therefore an increase in high-frequency energy relative to low-frequency energy. Measurements with a wide-bandwidth microphone gave shorter rise times than measurements with conventional 3.175-mm microphones. Because of its extended high-frequency response, a wide-band microphone gives more confidence that one is observing true pressure variation and not the effect of the transducer. Therefore the results of this test suggest that wide-band microphones should be used for measuring rise times of finite-amplitude blast waves.

Acknowledgments

This work was supported by the U.S. Army Engineering Research and Development Center Construction Engineering Research Laboratory (ERDC-CERL). The authors appreciate the assistance of the Edgewood staff in coordinating the tests. The authors also gratefully acknowledge Tom Gabrielson and Tim Marston for advice on and construction of microphone baffles.

References and links

[1]D. T. Blackstock, M. F. Hamilton, and A. D. Pierce, "Progressive waves in lossless and lossy fluids," *in Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock, chapter 4 (Academic Press, San Diego, CA, 1998).

[2]H. E. Bass and R. Raspet, "Vibrational relaxation effects on the atmospheric attenuation and rise times of explosion waves," J. Acoust. Soc. Am. **64**, 1208–1210 (1978).

[3]H. E. Bass and J. Ezell, "Effect of vibrational relaxation on rise times of shock waves in the atmosphere," J. Acoust. Soc. Am. **74**, 1514–1517 (1983).

[4]H. E. Bass, R. Raspet, J. P. Chambers, and M. Kelly, "Modification of sonic boom waveforms during propagation from the source to the ground," J. Acoust. Soc. Am. **111**, 481–486 (2002).

[5]R. D. Ford, D. J. Saunders, and G. Kerry, "The acoustic pressure waveform from small unconfined charges of plastic explosive," J. Acoust. Soc. Am. **94**, 408–417 (1993).

[6]K. Attenborough, P. Schomer, F. van der Eerden, and E. Védy, "Special issue on blast sound propagation and reduction," Noise Control Eng. J. **53**, 69–114 (2005).

[7]M. O. Anderson, "The propagation of a spherical N wave in an absorbing medium and its diffraction by a circular aperture," Technical Report ARL-TR-74-25, Applied Research Laboratories, The University of Texas at Austin (1974).

[8]L. B. Orenstein, "The rise time of N waves produced by sparks," Technical Report ARL-TR-82-51, Applied Research Laboratories, The University of Texas at Austin (1982).

[9]W. M. Wright, "Propagation in air of N waves produced by sparks," J. Acoust. Soc. Am. **73**, 1948–1955 (1983).

[10]H. E. Bass, B. A. Layton, L. N. Bolen, and R. Raspet, "Propagation of medium strength shock waves through the atmosphere," J. Acoust. Soc. Am. **82**, 306–310 (1987).

[11]G.R.A.S. Sound & Vibration, *1/8-inch Pressure Microphone Type 40DP-Product Data and Specifications*.

[12]G.R.A.S. Sound & Vibration, *1/4-inch Preamps. Types 26AA*, *-AB*, *-AC*, *-AL*, *AN & -TC-Product Data and Specifications*.

[13]Brüel & Kjær, *1/4-inch Microphone Preamplifier Type 2639-Product Data*.
[14]Brüel & Kjær, *Battery Driven Power Supply Type 2804-Product Data*.
[15]T. B. Gabrielson, T. M. Marston, and A. A. Atchley, "Nonlinear propagation modeling: Guidelines for supporting measurements," in *Proc. NOISE-CON 2005* (Institute of Noise Control Engineering of the USA, Washington, D.C., 2005).
[16]D. T. Blackstock, *Fundamentals of Physical Acoustics* (Wiley, New York, 2000).
[17]W. M. Wright, "Studies of N waves from weak sparks in air: Final report," Technical Report NR-384–321, Office of Naval Research (1971).
[18]W. M. Wright, "High-frequency electrostatic transducers for use in gases," Technical Report NR-384–903, Office of Naval Research (1962).
[19]Belden CDT, *1032A Paired 300V Power-Limited Tray Cable-Detailed Specifications & Technical Data*.
[20]P. L. Walter, "Air-blast and the science of dynamic pressure measurements," Sound Vib. **38**, 10–17 (2004).
[21]S. Ollivier and P. Blanc-Benon, "Model experiments to study acoustic N-waves propagation through turbulent medium," in *10th AIAA/CEAS Aeroacoustics Conference*, AIAA 2004–2921 (American Institute of Aeronautics and Astronautics, Reston, VA, 2004).

Mechanisms of irregular vibration in a physical model of the vocal folds

David A. Berry, Zhaoyan Zhang, and Juergen Neubauer
The Laryngeal Dynamics Laboratory, David Geffen School of Medicine at UCLA, 1000 Veteran Avenue, Suite 31-24, Los Angeles, California 90095-1794
daberry@ucla.edu

Abstract: Previous investigations have shown that one mechanism of irregular vocal fold vibration may be a desynchronization of two or more vibratory modes of the vocal fold tissues. In the current investigation, mechanisms of irregular vibration were further examined using a self-oscillating, physical model of vocal fold vibration, a hemi-model methodology, and high-speed, stereoscopic, digital imaging. Using the method of empirical eigenfunctions, a spatiotemporal analysis revealed mechanisms of irregular vibration in subharmonic phonation and biphonation, which were not disclosed in a standard acoustic spectrum.

PACS numbers: 43.70.Aj, 43.70.Bk [AL]
Date Received: May 8, 2006 **Date Accepted:** August 15, 2006

1. Introduction

Over the past decade, the method of empirical eigenfunctions (EEFs) has been used to examine spatiotemporal structures in vocal fold vibration (Berry *et al.*, 1994; Alipour *et al.*, 2000; Svec *et al.*, 2000; Berry *et al.*, 2001; Neubauer *et al.*, 2001; Zhang and Jiang, 2005; Doellinger *et al.*, 2005). Herzel *et al.* (1994) hypothesized that "normal phonation corresponds to an essentially synchronized motion of all vibratory modes. A change of parameters … may lead to a desynchronization of certain modes resulting in bifurcations and chaos." Berry *et al.* (1994) confirmed this hypothesis using a finite element simulation of vocal fold vibration. In that study, the spatial EEFs (derived from the nodal displacements of the finite element model) were similar for both simple and complex vibrations patterns, and the desynchronization of the EEFs was the primary mechanism of irregular vibration. Neubauer *et al.* (2001) further substantiated this hypothesis using high-speed digital imaging of vocal fold vibration in human subjects. Desynchronization of anterior-posterior EEFs (derived from glottal contours) distinguished two classes of biphonation: one based on left-right asymmetries, and one based on anterior-posterior asymmetries. In the current study, this desynchronization hypothesis will be subjected to more stringent testing through the use of a self-oscillating, physical model of vocal fold vibration. That is, the laboratory investigations will not be limited by the mathematical simplifications of the previous theoretical studies. Moreover, in contrast to the clinical study which only permitted vocal fold imaging from a superior aspect, the laboratory investigations will yield a systematic study of the medial surface dynamics of the vocal folds. In particular, medial surface dynamics will be extracted from the model using a hemi-model methodology and high-speed, stereoscopic, digital imaging. As a function of subglottal pressure, both simple and complex oscillatory regimes will be imaged, and medial surface coordinates will be extracted. Using the method of EEFs, a spatiotemporal analysis of the medial surface dynamics will evaluate the desynchronization hypothesis.

2. Methods

A physical replica of the human vocal system was constructed using a rubber model of the vocal folds and a uniform (PVC) tracheal tube (with an inner diameter of 2.54 cm and a length of approximately 60 cm) connected upstream to an expansion chamber, simulating the subglottal system. The expansion chamber had an inner cross-section of 23.5×25.4 cm and was 50.8 cm

long. The inside of the expansion chamber was lined with a 2.54-cm-thick layer of fiberglass. The expansion chamber was connected to the air flow supply through a 15.2-m-long rubber hose, reducing possible flow noise from the air supply. Because the isotropic, single-layered, physical model of the vocal fold has been described previously (Thomson *et al.*, 2005), few details will be given here. However, in the current study, the Young's modulus was approximately 3 kPa for strains in the range of 0–20%, as compared to 13.7 kPa in the previous study. As described elsewhere (Berry *et al.*, 2001; Doellinger *et al.*, 2005), the medial surface dynamics of the folds were imaged using a hemi-model methodology, stereoscopic imaging (through means of a glass prism), and a high-speed digital camera (Fastcam-Ultima APX, Photron Unlimited, Inc.). A frame rate of 2000 Hz was used with a spatial resolution of 1024×1024 pixels per image.

Prior to imaging, graphite powder was sprinkled on the medial surface of the vocal fold to form random dot patterns. In the post-processing stage, a multipass, time-series cross-correlation analysis was performed on the medial surface images using the image-processing package DaVis (LaVision Inc.). In contrast to previous methods for tracking vocal fold displacements which were semiautomatic (Berry *et al.*, 2001; Doellinger *et al.*, 2005), the DaVis software package was fully automated. The lack of automation in previous imaging studies precluded the study of irregular vibrations.

The subglottal acoustic pressure in the tracheal tube was monitored using a probe microphone (B&K 4182), which was mounted flush with the inner wall of the tracheal tube, 5 cm upstream from the vocal fold plates. The mean transglottal pressure was measured using a pressure transducer (Baratron type 220D), which was also mounted flush with the inner wall of the tracheal tube, 2 cm upstream from the vocal fold plates. Analog-to-digital conversion of the signal was performed using a United Electronic Industries Powerdaq board (model no. PD2-MFS-8-500/16), with 16 bit resolution over a ±10 V measurement range at a sampling rate of 50 kHz.

3. Results and discussion

During the experiment, the subglottal pressure was increased from zero to a maximum value in discrete increments, and then decreased back to zero in discrete decrements. At each step, measurement was delayed for an interval of approximately 4–5 s after the subglottal pressure changed, allowing the flow field to stabilize. The power spectra of the microphone signal at each of these constant subglottal pressure values are summarized in Fig. 1 . Inspection of this figure suggests that several distinct oscillatory regimes appeared in the physical model: "1" denotes biphonation at a subglottal pressure of approximately 2.1 kPa, "2" denotes a quintupling bifurcation (creating a period-5 subharmonic) at a subglottal pressure of approximately 1.8 kPa, and "3" denotes periodic phonation, or a limit cycle, at a subglottal pressure of approximately 0.8 kPa.

The acoustic spectra for each of these vibratory regimes are shown in Fig. 2 . In Fig. 2(a), the acoustic spectrum for periodic phonation is shown with a fundamental frequency of approximately 200 Hz, and two overtones. In Fig. 2(b), the acoustic spectrum of a subharmonic phonation is shown, with a fundamental frequency of approximately 231 Hz, and subharmonics occurring at integer multiples of approximately 46.2 Hz, corresponding to a period-5 or quintupling of the fundamental frequency. The entrainment, or integer ratio, of these two dominant frequencies is captured by the discrete, finite number of spectral lines in Fig. 2(b). On the other hand, the dense spectral line structure in Fig. 2(c) suggests a lack of entrainment, or a noninteger ratio, of two independent frequencies.

Using the method of EEFs, a spatiotemporal analysis was performed on the three distinct phonatory regimes. The EEFs were computed from high-speed recordings over a time window of 150 ms, centered at the three time locations shown in Fig. 1. The EEFs for periodic phonation are illustrated in the two left-most columns of Fig. 3 . The first column illustrates the spatial EEFs (only the medial-lateral component is shown for comparison with lateral modes of vibration introduced previously in the literature). The second column illustrates the spectra of the corresponding temporal EEFs. Spatial EEFs are often considered the buildings blocks of

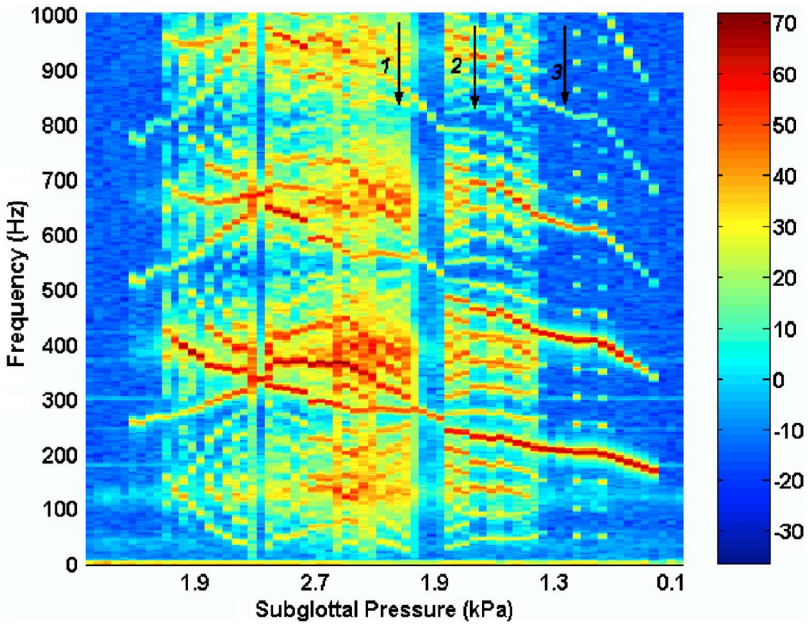


Fig. 1. (Color online) A spectral bifurcation diagram produced by the physical model of vocal fold vibration [in SPL (dB)]. The numbers 1, 2, and 3, denote the instances at which biphonation, subharmonic phonation (quintupling), and periodic phonation occurred, respectively.

vibration. For example, for a linear system, the spatial EEFs would correspond to the normal modes of the system. In column 1, the first spatial EEF (e.g., illustrated in the top subplot) appears to be qualitatively similar to a x-11 mode (according to nomenclature used previously by Titze and Strong (1975), and Berry and Titze (1996)), in which superior and inferior portions of the medial surface vibrate 180 deg out-of-phase with each other. These out-of-phase regions are illustrated by the adjacent red and blue regions found between 0 and 5 mm along the x axis (anterior-posterior direction) and near 0 mm along the y axis (inferior-superior direction). The second spatial EEF (e.g., the subplot directly below EEF1) appears qualitatively similar to a x-10 mode, which captures the in-phase lateral vibrations of the fold (as marked by the blue region, between 0 and 5 mm on the x axis). As shown in the second column, the corresponding temporal EEFs entrained at a fundamental frequency of 200 Hz. As shown in Table 1 these first

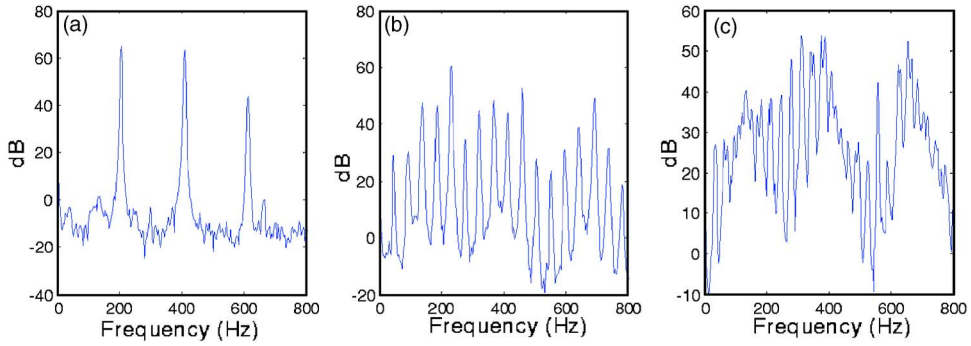


Fig. 2. (Color online) Acoustic spectra corresponding to the numbers 3, 2, and 1 in Fig. 1, which represent the following phonation types, respectively: (a) periodic phonation; (b) subharmonic phonation (quintupling); and (c) biphonation.

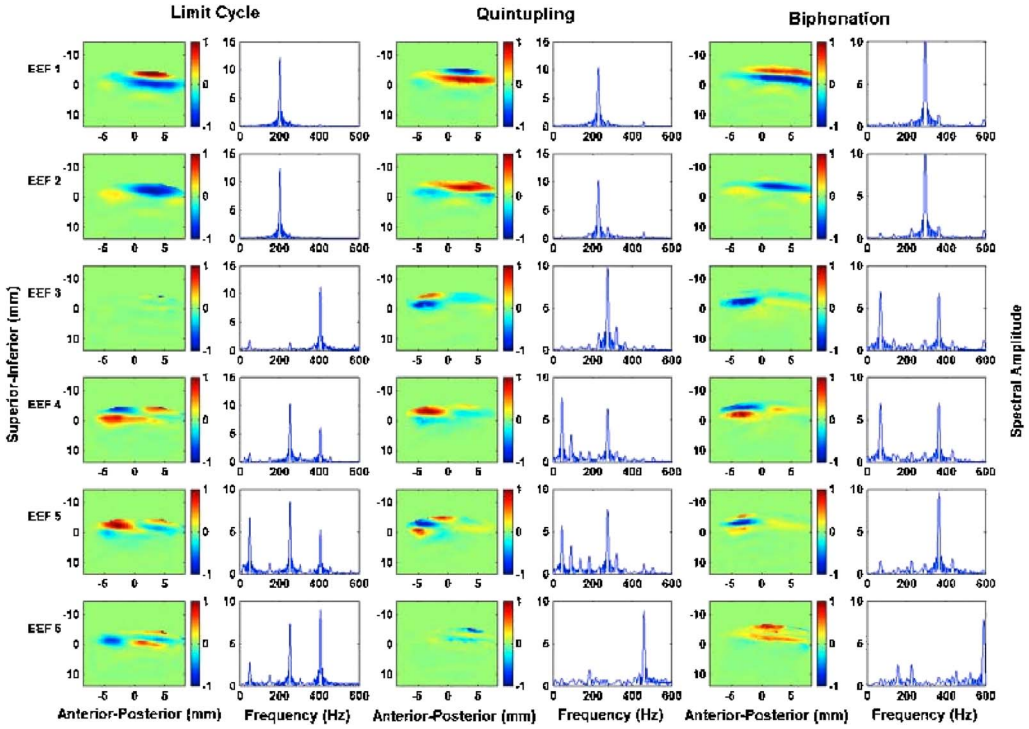


Fig. 3. (Color online) Columns (1), (3), and (5): the normalized medial-lateral components of the first six spatial EEFs for periodic phonation (limit cycle); subharmonic phonation (quintupling); and biphonation, respectively. As shown to the left of the figure, the *y* axis spans the "inferior-superior" direction. Columns (2), (4), and (6): the acoustic spectra of the first six temporal EEFs for periodic phonation (limit cycle); subharmonic phonation (quintupling); and biphonation, respectively. As shown to the right of the figure, the *y* axis captures the "spectral amplitude" (arbitrary units).

two EEFs captured 96.2% of the variance of the medial surface data. Spatial EFFs 3–6 captured higher-order modes in both the anterior-posterior and inferior-superior directions. These higher-order EEFs exhibited both harmonic and nonharmonic frequency components, with respect to the fundamental frequency of 200 Hz.

The EEFs extracted from the subharmonic (quintupling) regime are illustrated in the two central columns of Fig. 3. The first two spatial EEFs entrain at a fundamental frequency $f_1 = 231$ Hz, and are qualitatively similar to the first two EEFs from periodic phonation (the switching of blue and red from superior to inferior positions, as compared to column one, is not significant: Either case demonstrates that the inferior and superior regions are 180 deg out-of-phase with each other). Spatial EEFs 3–6 capture higher-order modes.

Table 1. Percentage variances (or weights) of the first six EEFs.

	Limit cycle weight (%)	Quintupling weight (%)	Biphonation weight (%)
EEF1	58.6	60.5	45.6
EEF2	37.6	27.7	30.1
EEF3	0.76	3.3	6.1
EEF4	0.71	2.3	4.4
EEF5	0.54	1.6	3.9
EEF6	0.42	0.7	1.7

Table 2. Dot products to compare the spatial correlation of the first six EEFs of periodic phonation (P), quintupling (Q), and biphonation (B).

	EEF1	EEF2	EEF3	EEF4	EEF5	EEF6
(P_i, Q_i)	91.2	78.7	4.6	7.6	4.7	41.7
(P_i, B_i)	72.5	66.6	9.0	50.7	23.5	14.6
(Q_i, B_i)	76.8	79.6	41.4	42.1	58.0	19.4

While the EEFs 1–2 entrained with each other, the higher-order EEFs also entrained with the subglottal acoustic resonance f_{sg} of approximately 138 Hz, as estimated by the quarter-length resonance of the subglottal tube. For example, EEF3 (which possessed a dominant frequency component at f_2=277 Hz), exhibited 2:1 entrainment with f_{sg}. EEF4 had significant frequency components at approximately 46, 93, 139, 185, 231, and 277 Hz, etc. All these frequency components were linear combinations of the two primary frequencies, f_1 and f_2 [i.e., $f(n,m)=nf_1+mf_2$, where m and n are integers, e.g., $f(-1,1)=-231$ Hz+277 Hz=46 Hz].

From Table 1, the first two EEFs capture 88.2% of the variance, in comparison to 96.1% of the variance captured by the first two EEFs of periodic phonation. Six EEFs from this quintupling regime were required to capture 96.1% of the variance. In Table 2 , the spatial EEFs from periodic phonation and the quintupling regime are compared using the dot product (the Euclidean scalar product of two vectors). A value of 100% corresponds to identical vectors, or to identical spatial EEFs. A value of 0% corresponds to perpendicular vectors, or to EEFs which are spatially uncorrelated. The dot products of the first three EEFs of the periodic and quintupling regimes were 91.2, 78.7, and 4.6%, denoting a relatively high-spatial correlation between the first two EEFs of the two vibratory regimes. Because EEFs 3 and higher were only weakly excited in the periodic regime, the higher-order EEFs may have captured noise patterns rather than physically meaningful information. Thus, one would not necessarily expect to see a correspondence between the higher-order EEFs of these two oscillation types.

The EEFs extracted from the example of biphonation are illustrated in the two rightmost columns of Fig. 3. The first two spatial EEFs entrain at a fundamental frequency of f_1 =296 Hz, and are qualitatively similar to the first two EEFs from periodic phonation. Spatial EEFs 3–6 capture higher-order modes. EEFs 3–4 have two dominant frequency components at f_2=69 Hz (which exhibited a 1:2 entrainment with the subglottal acoustic resonance of 138 Hz), and 365 Hz. EEF5 also has a frequency component at 365 Hz, and EEF6 has a frequency component at 592 Hz. Thus, again, the higher-order EEFs entrained not only with other tissue modes, but also with the subglottal acoustic resonance. Also, similar to the quintupling regime, all frequency components were linear combinations of f_1 and f_2. Thus, while the central importance of these two frequencies was not apparent in the corresponding acoustic spectra of Fig. 2(c) the spatiotemporal decomposition using EEFs revealed considerable information regarding mechanisms of irregular vibration.

From Table 1, the first two EEFs captured 75.7% of the variance, in comparison to 88.2% for the quintupling regime, and 96.1% for periodic phonation. Six EEFs from this quintupling regime were required to capture 91.8% of the variance. In Table 2, the spatial EEFs from periodic phonation, quintupling, and biphonation regime were compared. For the first spatial EEF, the dot product across the three regimes ranged between 72.5 and 91.2%; and for the second EEF, the dot produced ranged between 66.6 and 79.6%. For the higher-order EEFs the dot products continued to decrease. However, the dot products between the quintupling regime and biphonation (Q_i, B_i) did not diminish as quickly for the higher-order EEFs, perhaps because the higher-order modes were relatively strong and thus essential descriptors of these more complex phonation types, and less significant for periodic phonation.

Dynamical complexity of the vocal fold system not only complicated the entrainment patterns of the EEFs, but also increased the effective number of spatial EEFs more dramatically than observed in a previous theoretical study (Berry *et al.*, 1994). In part, this may have been

because the parameter variation implemented in the current study was subglottal pressure, which naturally excites additional modes when increased. In contrast, the previous study varied mucosal stiffness.

One may argue that the bifurcations observed in this paper may not be related to human phonation because of the relatively long subglottal tube length of 60 cm. However, the entrainment of the vocal folds to acoustic resonances of the supraglottal tract in human subjects has been reported previously (Mergell and Herzel, 1997). Moreover, previous investigators have observed an intimate connection between subglottal acoustic resonances and vocal registers (Austin and Titze, 1997). Also, given the range of first resonances of the human subglottal system reported in the literature (approximately 400–500 Hz, see Ishizaka *et al.*, 1976), some of the frequency components of the EEFs for irregular phonation reported by Neubauer *et al.*, (2001, e.g., in the range of 197–338 Hz), may have been exhibiting 1:2 entrainment patterns with sub- or supra-glottal acoustic resonances. Thus, while future laboratory studies which more accurately model the human subglottal system may be needed to further investigate mechanisms of irregular vibration, the entrainment patterns observed between tissue modes and acoustic resonances in this study appear to be characteristic of source-resonator systems, and are consistent with previous observations in voice literature.

4. Conclusions

Mechanisms of irregular vibrations were explored in a self-oscillating, physical model of vocal fold vibration. Medial surface dynamics of a vocal fold were obtained using a hemi-model methodology together with high-speed, stereoscopic, digital imaging. Using the method of EEFs, a spatiotemporal analysis was performed on the medial surface vibrations of the folds. Mechanisms of irregular vibration for subharmonic phonation (quintupling) and biphonation were revealed by the entrainment patterns of the spatial EEFs. While stronger, lower-order EEFs tended to entrain with each other, weaker, higher-order EEFs also sometimes entrained with subglottal acoustic resonances. These mechanisms of irregular vibration were not disclosed by a spectral analysis of the acoustic signal, but were revealed through a spatiotemporal analysis of vocal fold oscillations using the method of EEFs. Using these same laboratory techniques, future studies are envisioned which will investigate mechanisms of irregular vibration across many phonatory conditions of both the normal and pathological voice. Such studies may employ the following models: multilayered physical models, excised animal larynges, excised human larynges, and *in vivo* animal larynges. While high-speed observation of the medial surface of the vocal folds is not currently available in the clinic, as the temporal and spatial resolution of ultrasonic imaging methods continues to increase, at some point in the future it may be possible to have clinical access to the medial surface of the vocal folds in order to assess mechanisms of irregular vibration. Already, Neubauer *et al.* (2001) have demonstrated that some mechanisms of irregular vocal fold vibration may be investigated clinically using a spatiotemporal decomposition of the superior view provided by high-speed, digital, endoscopic imaging.

Acknowledgment

This investigation was supported by Grant No. R01 DC03072 from NIDCD.

References and links

Austin, S. F. and Titze, I. R. (**1997**). "The effect of subglottal resonance upon vocal fold vibration," J. Voice **11**, 391–402.

Alipour, F., Berry, D. A., and Titze, I. R. (**2000**). "A finite element model of vocal fold vibration," J. Acoust. Soc. Am. **108**, 3003–3012.

Berry, D. A., Herzel, H., Titze, I. R., and Krischer, K. (**1994**). "Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions," J. Acoust. Soc. Am. **95**, 3595–3604.

Berry, D. A., Montequin, D. W., and Tayama, N. (**2001**). "High-speed, digital imaging of the medial surface of the vocal folds," J. Acoust. Soc. Am. **110**, 2539–2547.

Berry, D. A., and Titze, I. R. (**1996**). "Normal modes in a continuum model of vocal fold tissues," J. Acoust. Soc. Am. **100**, 3345–3354.

Doellinger, M., Berry, D. A., and Berke, G. S. (**2005**). "Medial surface dynamics of an *in vivo* canine vocal fold during phonation," J. Acoust. Soc. Am. **117**, 3174–3183.

Herzel, H., Berry, D. A., Titze, I. R., and Saleh, M. M. (**1994**). "Analysis of vocal disorders with methods from nonlinear dynamics," J. Speech Hear. Res. **37**, 1008–1019.
Ishizaka, K., Matsudaira, M., and Kaneko, T. (**1976**). "Input acoustic-impedance measurement of the subglottal system," J. Acoust. Soc. Am. **60**, 190–197.
Mergell, P., and Herzel, H. (**1997**). "Modeling biphonation — The role of the vocal tract," Speech Commun. **22**, 14l–154.
Neubauer, J., Mergell, P., Eysholdt, U., and Herzel, H. (**2001**). "Spatio-temporal analysis of irregular vocal fold oscillations: biphonation due to desynchronization of spatial modes," J. Acoust. Soc. Am. **110**, 3179–3192.
Svec, J. G., Horacek, J., Sram, F., and Vesely, J. (**2000**). "Resonance properties of the vocal folds: in vivo laryngoscopic investigation of the externally excited laryngeal vibrations," J. Acoust. Soc. Am. **108**, 397–407.
Thomson, S. L., Mongeau, L., and Frankel, S. H. (**2005**). "Aerodynamic transfer of energy to the vocal folds," J. Acoust. Soc. Am. **118**, 1689–1700.
Titze, I. R., and Strong, W. J. (**1975**). "Normal modes in vocal fold tissues," J. Acoust. Soc. Am. **57**, 736–744.
Zhang, Y., and Jiang, J. J. (**2005**). "Spatiotemporal chaos in excised larynx vibrations," Phys. Rev. E **72**, Art. No. 035201R, 1–4.

Imaging marine geophysical environments with vector acoustics

Dennis Lindwall
Marine Geosciences Division, Naval Research Laboratory, Stennis Space Center, Mississippi 39529
lindwall@nrlssc.navy.mil

Abstract: Using vector acoustic sensors for marine geoacoustic surveys instead of the usual scalar hydrophones enables one to acquire three-dimensional (3D) survey data with instrumentation and logistics similar to current 2D surveys. Vector acoustic sensors measure the sound wave direction directly without the cumbersome arrays that hydrophones require. This concept was tested by a scaled experiment in an acoustic water tank that had a well-controlled environment with a few targets. Using vector acoustic data from a single line of sources, the three-dimensional tank environment was imaged by directly locating the source and all reflectors.

PACS numbers: 43.60.Lq, 43.20.Ye, 43.30.Xm [DKW]
Date Received: December 19, 2005 **Date Accepted:** August 16, 2006

Vector acoustic sensors, sometimes called acoustic intensity sensors, are compact devices that measure the direction of motion of an acoustic wave.[1–3] They generally measure either the particle motion or the pressure gradient caused by the passing waves and produce data as a three-component vector. Three-dimensional vector acoustic data does not have the axisymmetric directional ambiguity of hydrophone array data. Vector sensors have been used to study the directional nature of noise in the ocean as well as to track whales.[4,5] Using a vector sensor with a controlled pulse source, one can determine both range and bearing for the sound source as well as scattering and reflection points. In the case of nonoverlapping reflections, the vector time series from a single source pulse recorded at a single position can provide a three-dimensional acoustic picture of the local environment. In practice, much of the data has overlapping signals from two or more reflections or scattering targets. Identification of reflecting surfaces requires multiple source or receiver positions. If multiple reflections are strong enough, they will appear as ghosts behind the real reflectors.

Data from linear hydrophone arrays used for geoacoustic surveys or seismic profiling are mapped onto half planes that are either the (usually flat) seafloor on one side of the ship track or vertically below the ship track. Three-dimensional (3D) structures in the environment that can be correctly imaged with vector sensor data will be incorrectly located in a hydrophone-based 2D image. This paper presents how a vector acoustic sensor may be used for marine geoacoustic applications and shows that vector acoustic data can be acquired with existing sensors. This is verified by an experiment using an accelerometer-based vector acoustic sensor in a water tank with a short-pulse source and passive scattering targets.

Vector sensing seismometers have been used to measure local particle motion in the solid earth since the start of observational seismology. These seismometers usually measure displacement or acceleration and with three sensors arranged orthogonally, produced a vector description of the local motion. One significant practical difficulty with seismometers is the coupling between the solid earth and the sensors as well as the device resonance.[5] Using seismometers in water posses even more difficult coupling problems but new developments in miniature accelerometers have solved the largest part of this problem by designing a small, rigid, neutrally buoyant sensor. Moving an accelerometer-based sensor through the water would require a more practical suspension solution while minimizing the flow noise.

Pressure gradient sensors have been around for many years, particularly in air acoustics[1] but there have been few demonstrations in water acoustics and they have never been used for geoacoustic survey applications. A typical pressure gradient sensor consists of six hydrophones

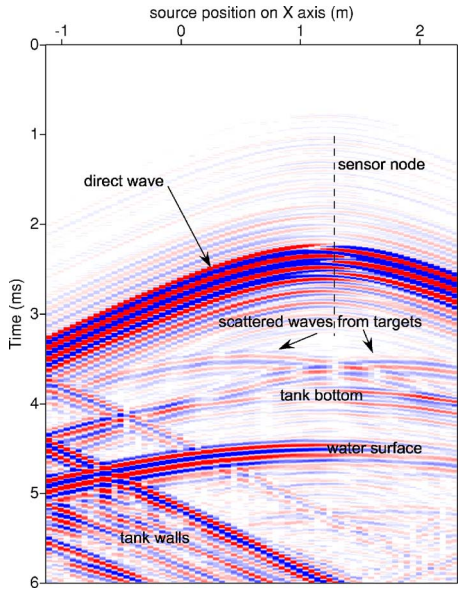


Fig. 1. (Color online) Time-series offset plot of the *x*-axis accelerometer. This display is in the seismic sense with time increasing downward and red and blue colors denoting positive and negative amplitudes. The direct arrival from the source is the most prominent signal and is hyperbolic with the *x*-axis node and its phase shift at the hyperbola's apex. The main pulse of the signal has a delay of 15 ms from the time break. The sandy bottom of the tank and the water surface reflections are visible as hyperbolic reflections with their apexes at the same offset as the direct arrival and times of 3.9 and 4.5 ms. The scattered waves from target floats in the tank show up as similar hyperbolic reflections with apexes that are offset to the sides and beginning times of about 3.5 s and having about 0.01 the amplitude of the direct wave.

arranged in three orthogonal pairs or four hydrophones on the vertices of a tetrahedron. The hydrophones must be exactly matched in phase but may result in a more robust package than a three-axis accelerometer.

We used an accelerometer-based vector sensor designed specifically for use in water by Roger Richards and others of the NAVSEA division of the Naval Underwater Warfare Center in conjunction with Wilcoxon Research Inc. This sensor is called the TV-001 Miniature Vector Sensor, has a sensitivity of 1.0 V/g and a frequency response of about 3 to 9 kHz, has the shape of a cylinder with two hemispherical end caps, and is 71 mm long and 41 mm in diameter with a mass of 54 g. This sensor is neutrally buoyant in water to produce the best motion coupling with the acoustic waves in the water. The sensor was suspended near the center of a pool of water with five, thin, elastic strands. This allowed the sensor to move freely in all directions and had a resonant frequency of approximately one hertz. The electrical cables were left slack so as to minimize their motion influence on the sensor. The sound source was a single transducer using a single cycle sine wave at 8 kHz. The vector sensor remained stationary and the source transducer was moved along several lines so as to cover multiple directions and distances from the sensor and targets. After shooting several reference lines, we placed two medium-sized floats and two small floats in the water as highly reflective targets.

Figure 1 is a time-series offset plot of the *x*-axis accelerometer. The reflections from objects that are broadside to the source line appear as hyperbolas and the reflections from objects beyond the ends of the source line appear as straight lines. The gain on this display is very high, clipping the direct wave, so that the much weaker reflections are visible. The weak precursor to the main signal pulse, which is one of the noise sources, is also visible. The coherency of the main pulse as well as the precursor across all source positions is some indication of the consistent receiver response over a wide range of angles.

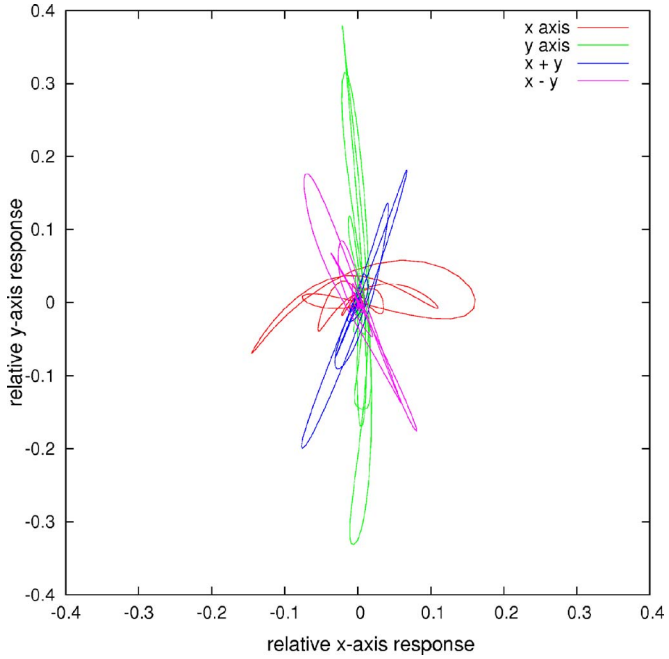


Fig. 2. (Color online) Hodogram of the x and y components of the accelerations from direct waves from the source. Two of the signals are aligned with the x and y axes and two are at 45-deg angles from the axies. The y-axis accelerometer of this sensor is about twice as sensitive as the x-axis accelerometer. The 45-deg off-axis signals appear to be about 30 deg from the y axis because there have been no amplitude corrections applied to these data.

Figure 2 illustrates the directional sensitivity of our accelerometer-based vector sensor and also shows some of the inaccuracies of the sensor and of the experimental setup. This is a hodogram of the raw data from the x and y components of the accelerations from direct source signals at four different locations. Two of the signals are aligned with the x and y axes and two are at 45-deg angles from the axes. The y-axis accelerometer of this sensor is about twice as sensitive as the x-axis accelerometer. The sensor motion from the source in the y-axis direction was aligned well with the y axis but the sensor motion from the source in the x direction deviated substantially from the axis. We speculate that the x-component deviations were caused either by an internal defect in the sensor or by motion restrictions from the electrical cables. The 45-deg off-axis signals however do not show such interference and indicate a direction of about 30 deg from the y axis because they have not been corrected for the lower sensitivity of the x-axis accelerometer. The principle of vector acoustic surveys can still be demonstrated with these data in spite of the different sensitivities and the nonlinear x-direction response.

A controlled source can be located with a single vector sensor measuring the bearing and the delay time. The source position vector $\boldsymbol{S}$ is

$$\boldsymbol{S} = \boldsymbol{G} + \hat{\boldsymbol{d}}tc, \tag{1}$$

where $\boldsymbol{G}$ is the receiver position vector. The direction vector $\hat{\boldsymbol{d}}$ is the unit vector in the direction of the accelerometer response, t is the travel time, and c is the sound speed. Locating a single scattering target with a single vector sensor can be done if the source location and pulse timings are known. Even if the source timings are unknown, the target bearings are known and there is a linear relation for the source and target ranges. Multiple scattering targets can be located with a single vector sensor provided that the individual scattered pulses are separated in time at the sensor. The target position vector $\boldsymbol{T}$ is calculated by

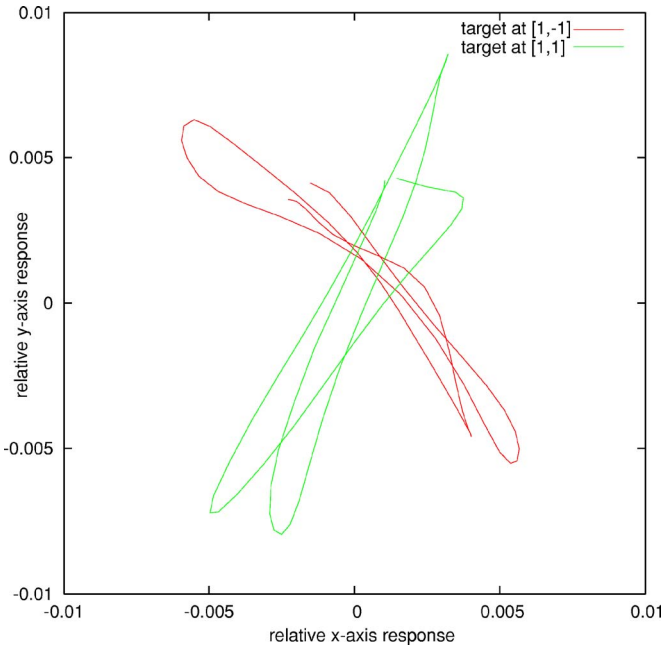


Fig. 3. (Color online) Hodogram of the x and y components of the accelerations due to waves scattered from the target floats. The signals were time windowed and selected from a position that had the best separation from other signals.

$$\boldsymbol{T} = \boldsymbol{G} + \hat{\boldsymbol{d}}r, \tag{2}$$

$$r = \frac{s(s - tc)}{b\cos^2(\theta/2) - s}, \tag{3}$$

$$s = \frac{tc + b}{2}, \tag{4}$$

$$\theta = \cos^{-1}\left(\frac{\boldsymbol{d} \bullet \boldsymbol{S}}{b|\boldsymbol{d}|}\right). \tag{5}$$

r is the range to the target, b is the source to receiver range, $\boldsymbol{d}$ is the target vector, θ is the angle between the receiver-source and the receiver-target vectors, and $||$ denotes the norm, or length, of the vector. The large dot denotes the vector dot product.

A hodogram of the motions in the x-y plane of the scattered signals from two target floats graphically illustrates that an acoustic environment and not just the sources can be directly sensed with vector acoustic data (Fig. 3). The signals were time windowed and from a source location that gave the best separation from other signals. These data have not been corrected for the different sensitivity of the x and y axes accelerometers.

If scattered signals from several targets or surfaces overlap, such as in our data, then they cannot always be located with a single source and receiver location. Locating each target within a field of scattering targets depends on target spacing relative to the signal wavelength. For most cases, a single line of either source or vector receiver locations is sufficient.

The data from our acoustic tank experiment were mapped into a volume image by binning the target vectors and summing the signal amplitudes in each bin. Figure 4 shows three slices through the z axis of the volume map showing the two float targets as well as more distant walls.

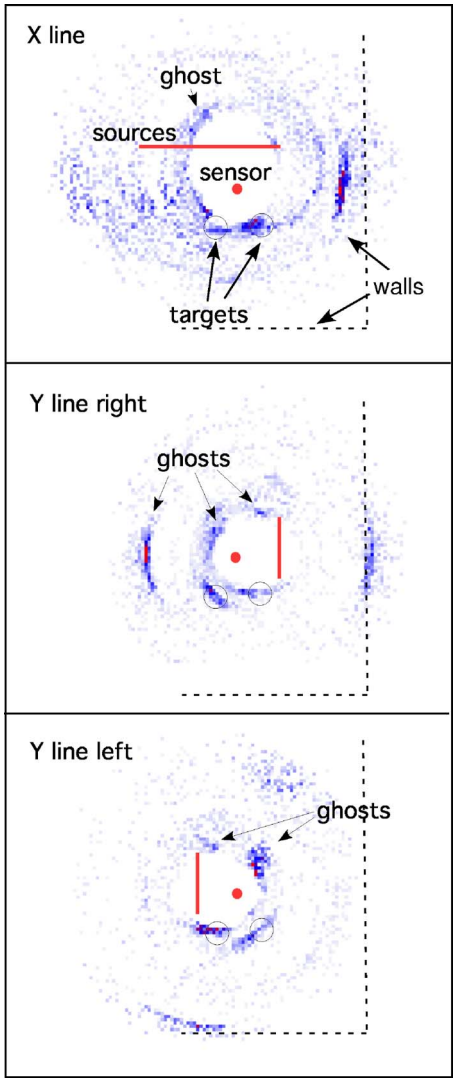


Fig. 4. (Color online) Horizontal slices of the volume image of the tank environment from three different source lines. The sensor is at the red dot in the center, the sources are every 5 cm along the red line, and the targets are located at the centers of the black circles. The source lines have a minimum distance of one meter from the source. These images ware made by calculating the range and bearing for each time sample using Eq. (2) and adding the magnitude of the signal to the corresponding grid point. The ghosts are due to the unrectified signals.

Since the scattered target position vector is calculated with a different equation than the source position vector, a time window was applied that eliminated most of the direct wave. The targets and walls are well positioned except for the wall in the top figure. Most of the targets and walls have ghosts caused by the unrectified negative parts of the signals. The noise causes uncertainty in the direction to the targets but not for most of the ranges. The data were corrected for different sensitivities of the axies before imaging.

The imaging of these vector sensor data can probably be improved substantially by incorporating existing techniques such as migration and phase coherence. These existing processing techniques can reduce the noise and may be able to separate overlapping arrivals from different targets such as seen in Fig. 1.

In conclusion, this work demonstrates the application of a vector acoustic sensor to marine geoacoustic surveys. The accelerometer data show the direct source wave as well as the target scattered waves and reflections from the nearby water surface, tank bottom, and sides. Vector data from single shots show that the wave motion direction can be readily determined for both direct waves and scattered waves. Without resorting to the usual methods of imaging used in seismic exploration, which in this case would have only been two dimensional and relied entirely on the use of a synthetic source aperture, the three-dimensional volume of the tank environment was imaged. Vector acoustics in marine environments may be applicable to a wide range of problems ranging from long-range ocean acoustics to subseafloor seismic exploration surveys.

Acknowledgments

This work would not have been possible without the loan of a TV-100 sensor as well as the technical expertise of Roger Richards from the NAVSEA division of the Naval Underwater Warfare Center. The experimental setup regarding the tank configuration and data acquisition was done by Harry Roberts and Roger Volk of NRL. This work was funded by the Office of Naval Research program element No. 61153N.

References and links

[1]M. J. Berliner and J. F. Lindberg, *Acoustic particle velocity sensors: Design, performance, and applications* (AIP Conf. Proc. **368**, 1996).

[2]F. J. Fahy, "Measurement of acoustic intensity using the cross-spectral density of two microphone signals," J. Acoust. Soc. Am. **62**, 1057–1059 (1977).

[3]F. J. Fahy, *Sound Intensity* (Elsevier Science, New York, 1989).

[4]G. L. D'Spain, W. S. Hodgkiss, and G. L. Edmonds, "Energetics of the deep ocean's infrasonic sound field," J. Acoust. Soc. Am. **89**, 1134–1158 (1991).

[5]G. L. D'Spain, W. S. Hodgkiss, and G. L. Edmonds, "The simultaneous measurement of infrasonic acoustic particle velocity and acoustic pressure in the ocean by freely drifting swallow floats," IEEE J. Ocean. Eng. **16**, 195–207 (1991).

ACOUSTICAL NEWS—USA

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

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

Report of the Auditor

Published herewith is a condensed version of our auditor's report for calendar year ended 31 December 2005.

Independent Auditors' Report

To the Executive Council
Acoustical Society of America

We have audited the accompanying statements of financial position of the Acoustical Society of America as of 31 December 2005 and 2004 and the related statements of activities and cash flows for the years then ended. These financial statements are the responsibility of the Society's management. Our responsibility is to express an opinion on the financial statements based on our audits.

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

In our opinion, the financial statements referred to above present fairly, in all material respects, the financial position of the Acoustical Society of America as of 31 December 2005 and 2004 and the changes in its net assets and its cash flows for the years then ended in conformity with accounting principles generally accepted in the United States of America.

O'CONNOR DAVIES MUNNS & DOBBINS, LLP
April, 24, 2006
New York, NY

ACOUSTICAL SOCIETY OF AMERICA
STATEMENTS OF FINANCIAL POSITION
31 DECEMBER 2005 AND 2004

	2005	2004
Assets		
Cash and cash equivalents	$1,658,462	$1,654,787
Accounts receivable	412,950	337,150
Marketable securities	6,872,236	6,458,334
Furniture, fixtures, and equipment—net	71,962	44,792
Other assets	403,516	451,527
	$9,419,126	$8,946,590
Liabilities		
Accounts payable and accrued expenses	$ 332,808	$ 315,778
Deferred revenue	1,513,580	1,604,183
Total liabilities	$1,846,388	$1,919,961
Net assets		
Unrestricted	$6,536,395	$5,993,687
Temporarily restricted	669,084	666,043
Permanently restricted	367,259	366,899
Total net assets	$7,572,738	$7,026,629
	$9,419,126	$8,946,590

 0001-4966/2006/120(3)/1133/20/$22.50

ACOUSTICAL SOCIETY OF AMERICA
STATEMENT OF ACTIVITIES
FOR THE YEARS ENDED 31 DECEMBER 2005 AND 2004

	2005	2004
Changes in Unrestricted Net Assets		
Revenues:		
Dues	$ 664,866	$ 629,000
Publishing—JASA	2,285,814	2,156,994
Standards	360,312	303,051
Spring Meeting	264,622	414,755
Fall Meeting	272,659	223,959
Other member services revenue	14,733	12,715
Other	121,436	101,708
Net assets released from restrictions	71,371	69,970
	$4,055,813	$3,912,152
Expenses:		
Publishing	$1,817,053	$1,531,446
Standards	388,934	411,942
Spring Meeting	300,916	467,159
Fall Meeting	302,798	227,258
Member Services	215,866	246,391
Other	438,758	648,521
Administration	485,329	484,939
	$3,949,654	$4,017,656
Net (loss) income from operations	$ 106,159	($ 105,504)
Non-operating activities:		
Interest and dividends	$177,702	$ 124,903
Realized gain (loss) on investments	106,985	69,697
Unrealized gain on investments	151,862	264,511
	$ 436,549	$ 459,111
Increase in unrestricted net assets	$ 542,708	$ 353,607
Changes in Temporarily Restricted Net Assets		
Contributions	$ 2,800	$ 21,855
Interest and dividends	25,682	20,708
Realized gain	18,983	12,368
Unrealized gain	26,947	46,934
Net assets released from restrictions	(71,371)	(69,970)
Increase in temporarily restricted net assets	$ 3,041	$ 31,895
Changes in Permanently Restricted Net Assets		
Contributions	$ 360	$ –
Increase in permanently restricted net assets	$ 360	$ –
Increase in net assets	$ 546,109	$ 385,502
Net assets at beginning of year	7,026,629	6,641,127
Net assets at end of year	$7,572,738	$7,026,629

Annual Reports of Technical Committees

(See October issue for additional report)

Acoustical Oceanography

Fall 2005 Meeting (Minneapolis, MN). The Technical Committee on Acoustical Oceanography (AO) sponsored two special sessions: (1) "Inversion Using Ambient Noise Sources," organized by Peter Gerstoft (Marine Physical Laboratory, Scripps); and (2) "Ocean Ecosystem Measurements," cosponsored by Animal Bioacoustics (AB) and organized by Whit Au (Hawaii Inst. of Marine Biology) and Van Holliday (BAE Systems). The Best Student Paper Awards in Acoustical Oceanography went to Michael Morley from the University of Victoria (First Prize winner) for his paper: "Estimating geoacoustic properties of marine sediments by matched field inversion using ship noise as a sound source," and Julie Oswald from Scripps (Second Prize winner), for her paper: "New tool for real-time acoustic species identification of delphinid whistles." Martin Siderius (HLS, San Diego) represented AO at the Technical Program Organizing Meeting.

Spring 2006 Meeting (Providence, RI). The 2006 Medwin Prize in Acoustical Oceanography was awarded to John Horne of the University of the Washington School of Aquatic and Fishery Science, Seattle, "for research on the acoustic identification of fish species and applications to fisheries resource management." Dr. Horne presented a highly entertaining Medwin Prize Lecture entitled "Acoustic species identification: When biology collides with physics." AO sponsored two special sessions, and contributed a paper in the Hot Topics in Acoustics special session. Ralph Stephen (Woods Hole Oceanographic Institution) organized a special session entitled "Ultra-Low and Low Frequency Marine Seismo-Acoustics." David Bradley (Penn State) and Whit Au (Hawaii Inst. of Marine Biology) organized a special session jointly with Animal Bioacoustics entitled "Exploitation of Sound by Marine Mammals." Karim Sabra of the Marine Physical Laboratory presented the AO Hot Topics lecture: "Extracting coherent information from cross correlations of random wave fields." Matthias Meyer (Free University of Brussels) won first prize in the competition for the Best Student Paper Award in Acoustical Oceanography for his paper entitled "Adjoint approach to the physical characterization of a shallow water environment," and Jason Holmes (Boston University) was awarded second prize for his paper entitled "Low frequency sound reflection and conversion to Darcy-type diffusion waves at bottom interfaces with marine sediments." Andone Lavery (Woods Hole Oceanographic Institution) and Mohsen Badiey (University of Delaware) represented AO at the Technical Program Organizing Meeting. AO also arranged for the speaker for the Tutorial Lecture at the Providence Meeting. The lecture, "The 2004 Sumatra earthquake and tsunami: Multidisciplinary lessons from an ocean monster," was given by geophysicist Emil Okal of Northwestern University.

Information on these and related matters is available on the new Acoustical Oceanography Technical Committee website, for which we are indebted to Kate Brooks of the Centre for Earth and Ocean Research at the University of Victoria. It can be reached through the ASA web page by clicking on "Committees." Please e-mail any material for or comments on the web site directly to chapman@uvic.ca.

I want to close by thanking all the volunteers who have contributed to the success of the activities of the AO Technical Committee during this past year. All of the AO members were greatly saddened by the passing of Hank Medwin earlier this year. Hank was one of the founding members of the AO technical committee, and his advice and encouragement have been important to all of us.

N. ROSS CHAPMAN
Chair

Animal Bioacoustics

The Animal Bioacoustics (AB) Technical Committee (TC) had a busy and productive year. Members of the TC as well as Acoustical Society of America (ASA) members citing animal bioacoustics as either a primary or a secondary interest were well represented at both the fall (Minneapolis) and spring (Providence) meetings.

We take special pride in acknowledging our most recent award recipients. At the Minneapolis meeting, James A. Simmons of Brown University received the Silver Medal in Animal Bioacoustics ''for his contributions to understanding bat echolocation.'' Also at this meeting, Alison Stimpert, from the Hawaii Institute of Marine Biology, received the Best Student Paper Award in Animal Bioacoustics for her paper "Sound production patterns from humpback whales in a high latitude foraging area." In Providence, we had two student paper award winners, Charlotte Kotas (Georgia Institute of Technology) for her paper "Are acoustically induced flows relevant in fish hearing?" and Anthony Petrites (Brown University) for his paper "Echolocating big brown bats shorten interpulse intervals when flying in high-clutter environments." Lee Miller (University of Southern Denmark) and Bertel Møhl (Aarhus University) were both elevated to Fellow status.

AB sponsored or cosponsored three special sessions at the Minneapolis meeting. These include Cognition in the Acoustic Behavior of Animals organized by Caroline Delong and cosponsored with Psychological and Physiological Acoustics (P&P); Ocean Ecosystem Measurements organized by Whitlow Au and Vance Holliday with Acoustical Oceanography (AO) as the primary sponsor; Temporal Patterns of Sounds by Marine Mammals organized by Jeanette Thomas; and Frequency Weighting for Animal Species organized by Larry Pater and Ann Bowles and cosponsored by P&P and Standards. A full session of contributed papers, chaired by Michael Ferragamo, was also held. The TC meeting attracted over 30 attendees, with most discussion centering around issues arising from the session on Frequency Weighting and Associated Standards topics.

The Providence meeting was very heavily attended, with so many AB sessions that overlap in scheduling could not be avoided. Joshua Schwartz organized a day-long session on Acoustic Interactions in Animal Groups that attracted many non-ASA members. Along with AO and Underwater Acoustics (UW), AB cosponsored the Joe Blue Memorial Session on Shallow Water and Marine Animal Acoustics, chaired by Ed Gerstein and George Frisk. Arthur Popper and Richard Fay organized and cochaired a session on Effects of Anthropogenic Sounds on Fishes, also cosponsored with AO and UW. Jim Simmons organized a session on Mechanisms of Biosonar, cosponsored with SP and EA. Whit Au and David Bradley organized and cochaired a day-long session on Exploitation of Sound by Marine Mammals, cosponsored by AO. Neil Todd organized a session titled Is Hearing All Cochlear, cosponsored by P&P. In addition, there were three contributed paper sessions, chaired by Peter Scheifele, John Buck, and Jennifer Miksis-Olds. AB organized a special art/soundscape performance by the local artist China Blue that explored urban bioacoustics. This performance was also heavily attended. Approximately 50 people participated in the ABTC meeting, where discussion centered on publication issues.

Planning continues for the Second International Workshop on Acoustic Communication by Animals, to be held in Oregon in summer 2008, under the leadership of David Mellinger.

AB owes a great debt to our members for their efforts in making our meetings successful and for their participation in the overall affairs of our Society. Whitlow Au finished his year as Vice President-Elect and now will serve as Vice President. Whit and Jim Simmons continue as section editors for AB for *JASA*, and Cynthia Moss continues as AB editor for *JASA Express Letters*. Ann Bowles serves as the AB representative to ASACOS and Dave Mellinger continues to keep our website (http://cetus.pmel.noaa.gov/Bioacoustics.html) up to date. Michael Ferragamo volunteered as our TPOM for the Minneapolis meeting, while John Buck and Andrea Simmons shared Technical Program Organizing Meeting duties for the Providence meeting. John, Jim Simmons, Peter Scheifele and Andrea Simmons all participated on the Local Organizing Committee for this meeting. We also acknowledge Mardi Hastings (Executive Council), Charles Greene (Medals and Awards), and Darlene Ketten (Membership) for the completion of their outstanding service on these important committees. We congratulate Marla Holt, our outgoing student representative, on the award of her Ph.D. Alison Stimpert will be replacing Marla on the Student Council. The terms of the following ABTC members ended at the Providence meeting: Ann Bowles, John Buck, Charles Greene, Mardi Hastings, Van Holliday, David Mann and Sam Ridgway. Their participation and assistance on the TC are much appreciated. We welcome our new ABTC members, Robert Dooling, Dave Mellinger, Arthur Popper and Annemarie Surlykke, and we are grateful to Whit Au, Christopher Clark, and William Cummings for agreeing to continue for another term.

Finally, this is my last report as ABTC chair. I thank AB members and members of the Technical Council for their support and encouragement during my term. It has been an honor to work with all of you. My successor is Richard Fay (Loyola University).

ANDREA M. SIMMONS
Chair 2003–2006

Biomedical Ultrasound/Bioresponse to Vibration

The Biomedical Ultrasound/Bioresponse to Vibration Technical Committee (BBTC) has been very active for the period of this report—Fall 2005 (Minneapolis) and Spring 2006 (Providence) meetings. Meetings have been well attended with many special sessions, awards, and Fellowship promotions.

Special sessions (and organizers) at the Minneapolis meeting included: Topical Meeting on Imaging and Control of High Intensity Focused Ultrasound-induced Lesions (Emad Ebbini), Medical Applications of Time Reversal Acoustics (Armen Sarvazyan), and Acoustic Radiation Force Methods for Medical Imaging and Tissue Evaluation (Mostafa Fatemi). Contributed sessions (and chairs) included: Ultrasound Contrast Agents (John Allen), Transducers and Imaging (Robert McGough). A sampling of topics covered was surgery through the skull with ultrasound, a 40 MHz array that has improved eye imaging, and preliminary investigations of tissue engineering of vocal-fold tissue.

Two other people put special effort in on behalf of the BBTC. Carr Everbach gave the hot topics presentation on ultrasound therapy and its integration with ultrasonic imaging, and presented a well-attended tutorial lecture on diagnostic imaging in biomedical ultrasound. Emad Ebbini served as the Technical Program Organizing Meeting representative.

For several years, the BBTC has tried to emphasize the spring meeting. Special sessions (and organizers) at the Providence meeting included Brown Tadpoles and Red Herrings: Boiling, Inertial Cavitation and Nonlinearity in High Intensity Focused Ultrasound Lesion Formation (Glynn Holt and Peter Kaczkowski), Memorial Session for Frederic Lizzi (Jeff Ketterling and Ernest Feleppa) Sensing and Imaging Using Light and Sound (Todd Murray and Ron Roy), Targeted Contrast Agents (Tyrone Porter and Christy Holland), and Celebration Session for Edwin Carstensen (Diane Dalecki, Lawrence Crum, Leon Frizzell, and Fred Kremkau). The contributed session was Imaging and Wave Propagation (Yuri Pishchalnikov). The tadpole session was an extended debate and discussion to explain how thermal necrosis of tissue (e.g., of cancer tumors) occurs, can be accelerated, is monitored, and may be controlled. Talks highlighted Dr. Lizzi's diverse and influential career by discussing for example, the use of ultrasound to differentiate cancerous from healthy tissue, development of a therapeutic ultrasound system used clinically to treat the eye, and adding ultrasound to enhance the delivery of drugs. Targeted agents, primarily tiny stabilized bubbles, are used to deliver drugs to a tissue region where ultrasound then activates the drug's release. Dr. Carstensen has led many in the ASA to help establish safe guidelines for diagnostic ultrasound, has brought nonlinear acoustics to the field of medical ultrasound, and established bioeffects, particularly near gas pockets in the body. Robin Cleveland and Glynn Holt were the TPOM representatives.

Several members of the BBTC received awards at the Providence meeting. Lawrence Crum received the Student Council Mentoring Award. Purnima Ratilal received the R. Bruce Lindsay Award "for contributions to the theory of wave propagation and scattering through a waveguide, and to the acoustic remote sensing of marine life." Some of this work has been applied to medical imaging. Mathias Fink received the Helmholtz-Rayleigh Interdisciplinary Silver Medal "for contributions to the understanding of time reversal acoustics." His work has led to many new medical imaging and therapy techniques, including a really gee-whiz way of supersonically generating shear waves used to image the stiffness of a tumor. Jamie Collin of Oxford University and Matthew Urban, working at the Mayo Clinic, were the BBTC Student Paper Award winners in a competition that included 20 excellent presentations by students. At both meetings, several people from the BBTC were promoted to Fellow.

On the whole, the BBTC is growing and membership has been very active. Technical committee meetings have been attended by over 50 people and long lists of special sessions and invited speakers are developed each time. Shira Broschat continues to maintain the BBTC website (http://moab.eecs.wsu.edu/~shira/asa/bubv.html). As evidenced by the list of awards and Fellows, Lawrence Crum, as representative to the Medals and Awards Committee, and Carr Everbach, who succeeded Tony Brammer as representative to the Membership Committee, have kept busy. This year, Paul Barbone took over for Jeff Ketterling as the ASA representative to the Biomedical Imaging Research Opportunities Workshop (BIROW). Neil Owen finished his term as the BBTC representative to the ASA Student Council at the Minneapolis meeting, and Michael Canney of the University of Washington was elected to succeed him. All are looking forward to a full meeting of at least six special sessions with our Japanese colleagues in Honolulu in the Fall of 2006 and to a special session on bioresponse to vibration at the Salt Lake meeting in Spring 2007.

MICHAEL R. BAILEY
Chair

Engineering Acoustics

The Engineering Acoustics Technical Committee (EATC) met at each of the two meetings of the Society, Minneapolis, MN in the Fall, and Providence, RI in the Spring.

In Minneapolis, the EATC sponsored one special session and one session of contributed papers. The title of the special session and organizer was: An ANSI standard for measuring *in-situ* directivity of hearing aids in 3-dimensions—Steve Thompson. In addition, EATC cosponsored several other sessions with the various technical committees. The Committee offers many thanks to Steve Thompson, who served as the representative to the Minneapolis Technical Program Organizing Meeting. The EATC met in Minneapolis on the evening of Tuesday, October 18.

In Providence, the EATC sponsored two special sessions and one session of contributed papers. The special session topics and organizers were as follows:

- Joe Blue Memorial Session II: Transduction, Linear and Nonlinear—Thomas G. Muir and Joseph F. Zalesak
- Joe Blue Memorial Session III: Transducers—David A. Brown and Edmund R. Gerstein.

The Committee offers many thanks to David A. Brown, Jeffrey E. Boisvert, and Thomas R. Howarth who served as EATC representatives to the Providence TPOM.

The EATC met in Providence on the evening of Tuesday, June 6.

At the Providence meeting, James E. West, a longstanding member of the Engineering Acoustics Technical Committee received the Society's Gold Medal award. His citation reads "for the development of polymer electret transducers, and for leadership in acoustics and the Society." A well deserved congratulations to Jim!

The recipients of the Best Student Paper awards for Engineering Acoustics were: in Minneapolis—Miguel A. Horta, "Sonic gas analyzer for microbiological metabolic measurements," (3aEA9), and in Providence—Aaron M. Kyle, "Wave propagation in fluid-filled tubes: measurements and model predictions," (4pEA8).

Prior to the Providence meeting, Dennis Jones was elected as the new Chair of EATC. Also, many thanks to our past student representative, Alexandra Loubeau, who was an integral member of the EATC. Her replacement will be Doug Wilcox, also from Pennsylvania State University.

The outgoing Chair of EATC would like to thank all the people who have worked to make the activities of EATC successful this year, and to invite all interested parties to attend EATC meetings to be held at future meetings of the Society.

KIM C. BENJAMIN
Chair 2003–2006

Musical Acoustics

During 2005–2006 the Technical Committee on Musical Acoustics (TCMU) was chaired by James Beauchamp, although Peter L. Hoekje substituted for him at the Providence meeting in June 2006. Representatives to ASA committees were as follows: James P. Cottingham, Membership; Uwe Hansen, Medals and Awards; Ian M. Lindevald, ASACOS; and James W. Beauchamp, Technical Council. Associate Editors were Diana Deutsch and Neville H. Fletcher. Technical Program Organizing Meeting (TPOM) representatives were Ian M. Lindevald (Minneapolis) and Courtney B. Burroughs (Providence). Those appointed or reappointed for 2006–2009 terms as members of TCMU are Rolf Bader, Xavier Boutillon, Jonas Braasch, Murray D. Campbell, Rene E. Causse, Antoine J. Chaigne, Neville H. Fletcher, Nicholas J. Giordano, J. M. Harrison, William M. Hartmann, William L. Martens, James M. Pyne, Daniel A. Russell, Punita G. Singh, Sten O. Ternstrom, Paul A. Wheeler, and Shigeru Yoshikawa. Also, two TCMU members were recently elected Fellows of the ASA: Anders G. Askenfelt, a musical acousti-

cian with the Royal Institute of Technology in Stockholm, and James Cottingham, previous TCMU Chair and professor of physics at Coe College, Cedar Rapids, Iowa.

TCMU presented four special sessions at the Minneapolis meeting in October, 2005: "Patents in Musical Acoustics," organized and cochaired by George Brock-Nannestand and George Augspurger; "Nonlinear Vibrations of Strings," organized and chaired by Antoine Chaigne; "Acoustics of Choir Singing I & II" (cosponsored by the Technical Committee on Architectural Acoustics), organized and cochaired by Sten Ternstrom, Thomas Rossing, and Anthony Hoover; and "Music Information Retrieval," organized and chaired by James Beauchamp. In addition, Ian Lindevald chaired the session "General Topics in Musical Acoustics," which consisted of contributed papers not related to the special sessions. The "Acoustics of Choir Singing" session was split into morning and afternoon sessions, and the afternoon session was followed by a panel discussion and choir concert at the Central Lutheran Church in Minneapolis. The panel, consisting of the session organizers and Dr. John Ferguson of St. Olaf College, Northfield, MN, discussed "Acoustical issues relevant to choral singing." This was followed by a service especially designed for the Acoustical Society with "Reflections" given by Pastor Bruce Benson and a choir concert by the St. Olaf Cantorei directed from the organ by John Ferguson. The concert, a definite highlight of the Minneapolis meeting for TCMU, featured "Magnificat," an original composition by Dr. Ferguson.

Three special sessions were presented at the Providence meeting in June, 2006: "Scaling of Musical Instrument Families," organized and chaired by George Bissinger; "Finite Element and Finite Difference Methods in Musical Acoustics I & II," organized and chaired by Rolf Bader and Uwe Hansen; and "Human-Computer Interfaces," organized and chaired by Jonas Braasch and William Martens. Also, Peter Hoekje chaired "Topics in Musical Acoustics," consisting of contributed papers. In addition, TCMU was a cosponsor of the special sessions "Surround Sound Essentials I & II," organized and chaired by Alexander Case and Anthony Hoover of the TCAA; and "Composed Spaces" and "Composed Spaces Loudspeaker Concert I & II," both organized and chaired by Alexander Case.

TCMU continues to promote student involvement in musical acoustics. Since the Vancouver (May 2005) meeting Brian Monson has been TCMU's representative to the Student Council, and he also has been chair of that group. Also, we have continued to sponsor ASA Best Student Paper Awards in Musical Acoustics. There were six entries for the competition at the Minneapolis meeting and nine entries for the Providence meeting. The winners were Harald Jers (Germany), first place at Minneapolis for "Multitrack analysis of amateur and professional choirs;" Jyri Pakarinen (Finland), second place at Minneapolis for "Modeling of tension-modulated strings using finite difference and digital waveguide techniques;" Andrey Ricardo da Silva (Canada), first place at Providence for "Benchmarking the lattice Boltzmann method for the determination of acoustic impedance of axisymmetric waveguides;" and Jacob Skubal (U.S.), second place at Providence for "Tuning parameters of a Nigerian slit gong."

Recently, new subject classifications (PACS) for Musical Acoustics replaced ones, which had been in place for many years. The new classifications incorporate suggestions provided by a subcommittee of TCMU. These are scales, intonation, vibrato, composition; music perception and cognition; bowed stringed instruments; woodwinds; brass instruments and other lip-vibrated instruments; plucked string instruments; drums; bells, gongs, cymbals, mallet percussion, and similar instruments; free reed instruments; pianos and other struck string instruments; pipe organs; reed woodwind instruments; flutes and similar wind instruments; singing; musical performance, training, and analysis; electroacoustic and electronic instruments; electronic and computer music; automatic music recognition, classification, and information retrieval; instrumentation and measurement methods for musical acoustics; analysis, synthesis, and processing of musical sounds.

A good number of papers on musical acoustics topics were published in *JASA* during 2005–2006. There were papers on such topics as timbre spaces, wind instrument and pipe organ analysis/synthesis, piano acoustics and modeling, vocal and choral acoustics, meter induction, music performance expression, and analysis of ethnic instruments and singing. Also, James Beauchamp recently published an article on brass acoustics in *Acoustics Today* (April, 2006), ASA's popular acoustics magazine.

TCMU sometimes supports extra-society events. Last year, ASA, in response to a request from James Cottingham, the previous TCMU Chair, cosponsored "Octet 2005," the First International Convention of the New Violin Family Association. The convention was held in Ithaca, New York in November 2005 with Carleen Hutchins, inventor of the string octet family, former TCMU member, and 1981 recipient of ASA's Silver Medal in Musical Acoustics, as the ASA contact person.

JAMES W. BEAUCHAMP
Chair

Noise

The Fall 2005 meeting of the ASA in Minneapolis was a special occasion for the Technical Committee on Noise: This meeting was held jointly with the NOISE-CON 2005 conference, organized by the Institute of Noise Control Engineering (INCE). Many members of TC-Noise are also members of INCE. It was a time to showcase all aspects of noise and its control. Forty-one special sessions included both Noise and NOISE-CON as sponsors. TC-Noise took the lead on five of these special sessions. These were "Special Session in Honor of William W. Lang" (organized by Paul Schomer and George Maling), "Workshop on Methods for Community Noise and Noise Policy" (Brigitte Schulte-Fortkamp and Bennett Brooks), "Specifying Uncertainties in Acoustic Measurements, I and II" (William Murphy and Ralph Muehleisen), and "Hospital Interior Noise Control" (Ilene Busch-Vishniac and James West). Mike Stinson was the Noise representative at the Technical Program Organizing Meeting.

The Spring 2006 Providence ASA meeting was not as noise-centric but was still host to an interesting array of special sessions. TC-Noise was lead organizer for "Audio-Visual Design in Soundscapes, I and II" (Brigitte Schulte-Fortkamp and Bennett Brooks), "Essential Acoustical Curriculum for Noise Control Techniques for Engineers, Industrial Hygienists, and Architects" (William Murphy), "New Loudness Standard" (Rhona Hellman), "Fifty Years of Speech Privacy, I and II" (Greg Tocci), and "Heating, Ventilation and Air-Conditioning Noise Control" (Dan Raichel and Bennett Brooks). Nancy Timmerman was the Noise representative at the Technical Program Organizing Meeting.

On the classroom acoustics front, the ASA has been having constructive discussions with the Director of the Modular Building Institute. The Collaborative for High Performance Schools (CHPS) has been convinced of the need for good acoustics in schools—they are promoting ANSI S12.60 and their actions are being held as an example for schools elsewhere in the US. ASA President Bill Yost issued a position statement stressing that sound amplification should not routinely be used in the classroom.

The contributions of several Noise people were recognized at these two meetings. For the Minneapolis meeting, we had two winners of the Noise Young Presenter award, Courtney McGinnes for her talk "An environmental and economical solution to sound absorption using straw" and Steven Ryherd for his talk "Acoustical prediction methods for heating, ventilating, and air-conditioning (HVAC) systems." In Providence, the Noise Young Presenter award went to Kent Gee for his talk "Analysis of high-amplitude jet noise using nonlinearity indicators." We have a new Fellow of the ASA, Samir Gerges. And in Providence, James West was presented with the ASA Gold Medal "for development of polymer electret transducers, and for leadership in acoustics and the Society."

The efforts of several volunteers should be recognized. Nancy Timmerman is the Noise representative on the Medals & Awards Committee, John Erdreich is our representative on the Membership Committee, and Richard Peppin is our representative on the ASA Committee on Standards. Connor Duke is our Student Council representative and, following the practice initiated a few meetings ago, the secretary for the TC-Noise meetings. (Connor was unable to attend Providence due to his recent marriage; we thank Matt Green for ably substituting.) The Noise web page (http://www.nonoise.org/quietnet/tcn/) is maintained by Les Blomberg. Ralph Muehleisen is coordinator for the Noise Young Presenter Awards. Serving as Associate Editor for *JASA Express Letters* is Mike Stinson and as *JASA* Associate Editors are Keith Attenborough, Kenneth Cunefare, Vladimir Ostashev, and Brigitte Schulte-Fortkamp.

My term as Chair of TC-Noise is complete. It has been a fun three years, in large part because of the dedication, cooperation, and good cheer that exists within the Technical Committee on Noise. I am pleased to welcome our incoming Chair, Brigitte Schulte-Fortkamp.

MICHAEL R. STINSON
Chair 2003–2006

Physical Acoustics

It is truly an honor and a privilege for me to have the opportunity to serve the Physical Acoustics (PA) community as your new Chair of the Physical Acoustics Technical Committee of the ASA. Tom Matula, our past Chair (from 2003–2005), has provided the committee with excellent leadership, blending past success with new initiatives in the long-range planning of special sessions among other activities. It is the hope of the new chair to provide good leadership, move forward with new initiatives, and most importantly to get PA committee members, along with other Society members, involved.

To quote Tom from his 2004–2005 Annual Report, "We've been redefining ourselves over the past few years … ." This is very evident at both the 150th Minneapolis and 151st Providence meetings.

While funding in sonoluminescence and thermoacoustics might have leveled, new avenues of interest are emerging in thermoacoustics, infrasound, nonlinear acoustics including (a) nonlinear techniques in biomedical ultrasonic imaging and therapy (along with acoustic radiation force methods), (b) high intensity focused ultrasound (HIFU) and boiling cavitation, and (c) nonlinear and time reversal techniques in acoustic-seismic landmine detection, atmospheric acoustics, and nanoacoustics.

The tutorial lectures "Diagnostic imaging in biomedical ultrasound," by E. Carr Everbach (Minneapolis) and "The 2004 Sumatra earthquake and tsunami: Multidisciplinary lessons from an ocean monster," by Emile Okal (Providence) were excellent and highly attended.

The Minneapolis ASA Meeting was held jointly with NOISE-CON 2005. PA planned one large (two part) special session entitled: "Thermoacoustics: What Our Customers Want and Why They Want It," organized by David Gardner and cosponsored by Engineering Acoustics which featured invited papers by Swift, Corey, Garrett (*et al.*), Spoor, Poese, Mozurkewich, Backhaus, and Kotsubo. There was standing room only for all of the 16 total papers. Physical Acoustics and Biomedical Ultrasound/Bioresponse to Vibration had participation in Hot Topics in Acoustics. Carr Everbach spoke on biomedical ultrasonic therapy, imaging, and therapeutic applications of High Intensity Focused Ultrasound. Thanks again, Carr.

Physical Acoustics cosponsored two special sessions: (a) NOISE-CON's, "Advances in Military Jet Noise Modeling," (co-organizers Sparrow and Downing) featuring nonlinear propagation; and (b) BB's "Acoustic Radiation Force Methods for Medical Imaging and Tissue Evaluation," (organized by Fatemi) featuring Rudenko's "Principles of radiation force," and radiation pressure papers by Marston, Ostrovsky, and Silva. Other PA sessions were "Topics in Atmospheric Acoustics" and "Topics in Seismic Acoustics." "Atmospheric Acoustics" papers featured shock propagation, volcanic infrasound, high altitude propagation, scattering, windscreen effects, ground impedance, and phased-array antennas. The "Seismic Acoustics" papers appeared to have roots going back to Sabatier's research, National Center for Physical Acoustics workshops, or his outreach and covered impedance, signature human footsteps, soil and landmine nonlinearity, resonances, granular-plate interaction and acoustic vibrometry.

Education in Acoustics had PA representation with Korman, Keolian, and Loubeau's "Acts of Sound"—Hands-on Workshop for High School Students," and "Acoustics Demonstrations," by Poese (demo apparatus workshop), Garrett, "Hey kid! Wanna build a loudspeaker?" and Korman and Bond's (landmine detection demo).

PA had two special sessions in Providence (with ~1400 registered). Joseph Turner and Donna Hurley organized an excellent session entitled "Acoustic Microscopy at the Nanoscale," with invited papers from Germany, Switzerland and Japan. The "Celebration of the Work of Brown University" (organizers Letcher, Maris and Korman) featured Robert Beyer (who received a standing ovation), Nyborg, Chick, Muir, Elder, Butler, Rogers, Foote, Brown, among talks by the organizers. Works of Lindsay, Williams, Beyer, and Westervelt were admired.

PA cosponsored the following special sessions: (a) Architectural Acoustics, Noise and Engineering Acoustics:" "Microperforated Acoustical Absorbing Materials," co-organizers Nocke and Xiang; (b) Biomedical Ultrasound/Bioresponse to Vibration and Signal Processing: "Sensing and Imaging Light and Sound," (co-organizers Roy and Murray); (c) Engineering Acoustics: "Joe Blue Memorial Session II: Transduction, Linear and Nonlinear," (co-organizers Gerstein, Muir, and Zalesak), featuring Joe's colleagues from Office of Naval Research, Naval Sea Systems Command, Naval Undersea Warfare Center, Georgia Tech, US Naval Academy, Boston University, Woods Hole, National Center for Physical Acoustics, and U. of Del., along with talks by the organizers; (d) Biomedical Ultrasound/Bioresponse to Vibration: "Celebration Session for Edwin Carstensen, Parts I and II (co-organizers Crum, Dalecki, Kremkau, and Frizzell), highlighted HIFU therapy, sound-induced lung hemorrhage, Carstensen's nonlinear contributions, medical imaging, and thermal effects. Part II included bubble phenomenon in cavitation and lithotripsy.

PA had four other sessions, (e) "Outdoor Sound Propagation," covering propagation under cloud cover, over irregular terrain, by turbulence, nonlinear and magnetic sound in the atmosphere, seismic pulses, blast waves, aeroacoustics, and sound absorption on Mars; (f) "Nonlinear Acoustics, Flow and Miscellaneous Topics," involving material damage, elastodynamic field resonances, interactions in elastic media, finite-amplitude and thermal gradient effects, time reversal, oscillations of a bias-flow aperture, flow around a porous screen, diffraction effects and transient evanescent waves; (g) "Landmine Detection; Scattering Phenomena," included aspects of linear and nonlinear acoustic detection, time reversal, and ultrasonic vibrometry, scattering problems included a wedge, a cylinder over an absorbing boundary, cluster of rigid rods, super-radiant modes in distorted hexagonal clusters, caustics from backscattering and coherent backscattering; and (h) "Thermoacoustics and Resonating Systems," discussing the ultrasonic analog for a laser, time reversal focusing, periodic binary systems and bandgap effects (engineering, transmission through periodic arrays of hollow cylinders, and bubbly media), particle history in fields, thin film resonators as mass sensors, thermoacoustic topics (infrasonic engines, miniature coolers and hot wire anemometry temperature measurements). This session had papers from Belgium, France, Mexico and Spain. All said, the Minneapolis and Providence meetings were big successes for PA.

Technical Program Organizing Meeting representatives were Joseph Turner, (Minneapolis) and Ron Roy, Tom Muir and Charles Thomas (Providence). Thank you all for a great job! At Minneapolis, Charlie Church and Paul Johnson became new fellows. Congratulations!

After Providence, Anthony Atchley became our new President and Gilles A. Daigle our new President-Elect. Outstanding and congratulations! Anthony and Gilles are flanked by our own PATC members Vic Sparrow and Wayne Wright who serve on the Executive Council.

In Minneapolis Wheeler Howard passed the torch to Todd Hay, our new student council rep. Thanks Wheeler and good luck to Todd.

It is time to acknowledge members of PA who volunteer their services on various committees: Robert Keolian who has served on the Medals and Awards Committee for over three years, will pass the baton to Tom Matula, who will officially start in Honolulu. Jim Sabatier was Robert's predecessor and both have worked hard to do a great job for the Society. Steve Garrett serves on the Membership Committee; Phil Marston serves on the Books$^+$ Committee, and Sameer Madanshetty serves on ASACOS. This year Mack Breazeale, Logan Hargrove, and James Miller reached 50 years of ASA service. Wow!

Other items: David Blackstock is extremely active in the Student Council and the program involving "Take a Student to Lunch." Ron Roy's tenure at *ARLO* is now over and Keith Wilson takes over as editor of *JASA Express Letters*. Ron also received the Eastman Fellowship. His one year sabbatical is in Oxford. The 2006 Physical Acoustics Summer School, PASS 2006, took place from 18–25 June 2006, at Sunrise Springs, La Cienega, New Mexico. The program is jointly run by NCPA (Hank Bass) and Penn State (Anthony Atchley), with strong ASA ties.

The 4th Joint Meeting of the ASA and Acoustical Society of Japan (ASJ) will take place in Honolulu, 28 November–2 December 2006. There are five excellent special sessions (see the most recent call for papers) that were planned during Tom Matula's reign. Your new Chair has met with Dr. Yoîti Suzuki (current President of ASJ and the ASJ Chair of the Technical Program) along with Dr. Hiroshi Sato (Secretary of the ASJ Technical Program Committee) in Minneapolis to develop a well thought out program involving all the Technical Committees.

Tom, I am grateful for all your help and guidance. Those are big shoes to fill. Thank you very much. You have been an inspiration to us all and an excellent Chair.

MURRAY S. KORMAN
Chair

Psychological and Physiological Acoustics

Reflecting Psychological and Physiological's (P&P's) strategy of emphasizing participation at ASA's spring meetings, the 150th Meeting of the Acoustical Society of America—the fall meeting in Minneapolis—was sparsely attended by the members of P&P. P&P sponsored one session, and was a cosponsor on four special sessions. We thank Magda Wojtczak for organizing the sessions for this meeting. Although P&P policy is to deemphasize fall meetings, the meeting in Hawaii next fall will be something of an exception, due to it being a joint meeting with the Acoustical Society of Japan.

The spring meeting in Providence was buzzing with activity. There were six sessions sponsored by P&P, one of which was cosponsored by the ASA Committee on Standards and by Noise. Of the six sessions, two were special sessions. We extend our thanks to the organizers of the special sessions: Sharon Kujawa and Lynne Marshall (Individual Susceptibility to Noise-Induced Hearing Loss) and Frederick J. Gallun (Characterizing Auditory Attention). P&P also contributed as the cosponsor of three additional special sessions. We thank Laurie Heller for accomplishing the challenge of organizing the P&P sessions for the Providence meeting.

The P&P open meeting included an update of the Student Council report by student representative Suzy Carr. Brenda Lonsbury-Martin informed the group about *JASA Express Letters*, the replacement for *ARLO*. Most of the discussion in the open meeting centered on the relatively low impact factor of *JASA*, the negative effect this is having on some members of the Society, and ways for trying to deal with this problem. The P&P Associate Editors provided an overview of their efforts, and encouraged the members of ASA to agree to review papers, and complete the reviews in a timely manner. We thank the P&P Associate Editors, Brenda L. Lonsbury-Martin, William P. Shofner, John H. Grose, Gerald D. Kidd, Armin Kohlrausch, Robert A. Lutfi, and Andrew J. Oxenham, for their efforts on our behalf. We express our thanks to Lynne Marshall, Don Sinex, and Magda Wojtczak for organizing the ballot for election to the Technical Committee, and, thanks to members of the P&P community who so willingly added their names to the ballot. The newly elected members of the Technical Committee are Sid Bacon, Qian-Jie Fu, Kim Schairer, Chris Shera, Ed Walsh, and Beverly Wright. We thank our outgoing members, Michelle Hicks, Lynne Marshall, Chris Plack, Don Sinex, and Magda Wojtczak.

Two of P&Ps Associate Editors, Gerald D. Kidd and Armin Kohlrausch, completed successful terms this year. Their efforts on our behalf are greatly appreciated; the work of our associate editors is important but also time consuming and difficult. Their willingness to contribute, and the contributions of the other associate editors, is highly valued by the community. Our two new Associate Editors are Rich Freyman and Ruth Litovsky. We also thank Joe Hall, III, our outgoing representative to the Medals and Awards Committee, and Les Bernstein, our outgoing representative to the Membership Committee.

The P&P Technical Initiatives continue unchanged. The initiatives include travel support for invited speakers, student receptions, and homepage maintenance. Suggestions for uses of funds, including innovations such as workshops, satellite meetings, etc., are welcome (estrick@purdue.edu).

ELIZABETH A. STRICKLAND
Chair

Speech Communication

The Speech Communication Technical Committee (SCTC) supports the activities, meetings, publications, etc. for the largest technical area in the Society. This report covers the meetings in Minneapolis, MN, and Providence, RI. The current members of the Committee are Jean Andruski, Norma Barroso, Patrice Beddor, Lynne Bernstein, Ocke-Schwen Bohn, Suzanne E. Boyce, Ann Bradlow, Dani Byrd, Roger Chan, Robert Fox, Alexander Francis, Bruce Gerratt, Kenneth Grant, Helen Hanson, Diane Kewley-Port, Jody E. Kreiman, Anders Lofqvist, Andrew Lotto, Benjamin Munson, Terrence Nearey, Peggy Nelson, Douglas O'Shaughnessy, Dwayne Paschall, Joseph Perkell, Astrid Schmidt-Nielsen, and Gary G. Weismer. *Ex-officio* members include Abeer Alwan (Membership Committee), Shrikanth Narayanan (ASACOS), Fredericka Bell-Bert (Medals and Awards Committee), and Jennell Vick (Student Council). The continuing Associate Editors for speech production are Anders Lofqvist and Brad Story; for speech perception they are Ann Bradlow, Kenneth Grant, Paul Iverson, Mitchell Sommers, and Joan Sussman. The continuing Associate editor for Speech Processing is Douglas O'Shaughnessy. Christine Shadle is the newly appointed Associate Editor for Speech Production.

We are grateful for the help of our members this year: (1) Our paper sorters, who arranged the technical programs at meetings: Ben Munson and Arlene Carney in Minneapolis; and Doug Whalen and Harriet Magen in Providence and (2) Coordinator for student judging: Khalil Iskarous. We also are very happy to welcome four new fellows to the ASA from the speech communication TC: Peggy Nelson, Kenneth Grant, Carole Espy-Wilson, Shri Narayanan,

The Klatt Award has been awarded to Rajka Smiljanic at the University of Minnesota.

Student Activities

The Committee sponsored two student activities at each meeting including a competition with a cash award for best student presentation and an evening reception. The evening reception, which is sponsored jointly with other technical committees, is intended to allow students to meet more senior ASA members informally. The reception was well attended. The student papers were judged by SCTC members and the winners were awarded $300 for first prize and $200 for second prize. In Vancouver the first- and second-place winners were Asaf Bacharach of Massachusetts Institute of Technology and Amanda Miller of Advanced Technological Research, Japan. In Minneapolis the first and second place winners were Tarun Pruthi of the University of Maryland, and Byron Erath of Purdue University. In Providence, a student ice-breaker social was initiated and was very well received.

Special Sessions, Special Talks and Workshops

Speech Communication has sponsored many interesting special sessions during the past year. In Minneapolis Pat Keating organized a special session "In honor of the 80th birthday of Peter Ladefoged," who passed away shortly thereafter. In Providence Yi Xu and Carlos Gussenhaven organized a special session entitled "Phonetic Enhancement in Speech: Evidence and Mechanisms."

MAUREEN L. STONE
Chair

Structural Acoustics and Vibration

Although structural acoustics and vibration play an important part in many areas of acoustics and noise, many of those involved with the Acoustical Society of America (ASA) through the Structural Acoustics and Vibration Technical Committee (SAVTC) have been historically involved in research for the US Navy. As Navy research funding in structural acoustics has decreased, so has the activities of SAVTC in ASA. From July 2005 to June 2006, there were two special technical sessions in the two ASA meetings, and 28 papers in structural acoustics published in JASA. During the same period ten years ago, there were six special technical sessions at the two ASA meetings and 63 structural acoustics papers published in JASA. Although this indicates that structural acoustics and vibration remain active, it also implies that challenges remain if the level of activities of SAVTC in ASA is to rise to the level it once enjoyed.

During the year from July 2005 to June 2006, there was one new Fellow elected from the SAVTC, Greg McDaniel. Jerry Ginsberg was presented the Trent-Crede award at the Minneapolis meeting in October 2005. The two special technical sessions were Experimental Modal Analysis, organized by Jerry Ginsberg at the Minneapolis meeting, and Ultrasonic Waveguides for Structural Monitoring, organized by Joel Garrelick at the Providence meeting in June 2006. Two student paper awards were presented at each of the ASA meetings. In Minneapolis, Noah Schiller won first place and Benjamin Doty, second place. In Providence, Michael Pedrick won first place and Christopher Dudley, second place. Two Associate Editors for structural acoustics were appointed; David Feit and Linda Franzoni. There are now six JASA Associate Editors in structural acoustics. The TPOM representative for SAVTC for the Minneapolis meeting was Courtney Burroughs and for the Providence meeting, Jeff Boisvert.

Courtney Burroughs completed his three-year term as chair of the SAVTC at the end of the Providence meeting. Sean Wu was elected as the new chair of SAVTC.

COURTNEY B. BURROUGHS
Chair 2003–2006

Underwater Acoustics

The year began with preparation for the Fall 2005 meeting in Minneapolis, for which we thank Dave Dowling and Dezhang Chu for representing the Underwater Acoustics Technical Committee (UWTC) at the Technical Program Organizing Meeting. In Minneapolis the UWTC sponsored special sessions entitled "Head waves and interface waves," organized by Peter Dahl, and "Sonar performance and signal processing in uncertain environments" organized by Lisa Zurk and David Dowling. We are also pleased to report that Henrik Schmidt received the Pioneers of Underwater Acoustics Medal in Minneapolis and we congratulate UWTC members Ralph Stephen and Dajun Tang on the occasion of their election to Fellowship in the ASA at the Minneapolis meeting and Jixun Zhou who was elected in Vancouver.

As always, there was healthy number of papers submitted to the student paper competition for the Minneapolis meeting under the auspices of the UWTC. The first prize went to Weichang Li (Massachusetts Institute of Technology) for the paper entitled: "*Identification of rapidly time-varying acoustic communication channels*." The second prize went to Jason Holmes (Boston University) for his paper entitled: "*An autonomous underwater vehicle technique for in-situ waveguide characterization*."

Finally, the period between the Minneapolis meeting and Providence meeting was marked by the sad passing in January of Prof. Hank Medwin, a pioneer in underwater acoustics and the founder of our allied Technical Committee on Acoustical Oceanography.

The spring 2006 meeting in Providence had special significance for Underwater Acoustics given the long standing involvement of local New England, and Navy institutions, in the field of underwater sound. A special thanks is given to Jim Miller (meeting chair) and Jim Lynch (technical chair), and to Kathleen Wage and Gopu Potty for representing the UWTC at the Technical Program Organizing Meeting.

In Providence, the UWTC sponsored sessions entitled "High-Frequency ambient noise," organized by Juan Arvelo, "Scattering of sound at the sea surface," organized by Duncan Williams, and "High frequency acoustic propagation and applications," organized by James Preisig and Mohsen Badiey. Paul Hines gave the Hot Topic presentation on behalf of the UWTC on the subject of vector sensors. In Providence the first prize for the student paper competition went to Kevin R. James (University of Michigan) for his paper entitled "*Approximating acoustic field uncertainty in underwater sound channels.*" The second prize went to Jason Holmes (Boston University) for his paper entitled: "*Shallow water waveguide characterization using an autonomous underwater vehicle towed hydrophone array.*"

We are very pleased to announce that UWTC member Purnima Ratilal of Northeastern University was the recipient of the R. Bruce Lindsay Award in Providence. We also congratulate UWTC members Charles Holland, Kevin LePage, and AOTC members Mark Trevorrow, David Palmer and Peter Rona, who were elected Fellows in Providence.

Shortly into 2006, the UWTC held its election for Chair and I am pleased to announce that Kevin LePage takes over from me. It has been a pleasure and honor to serve as Chair of the UWTC.

PETER H. DAHL
Chair 2003–2006

USA Meetings Calendar

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

2006

17–21 Sept.	INTERSPEECH 2006 (ICSLP 2006), Pittsburgh, PA[Web: www.interspeech2006.org]
28 Nov.–2 Dec.	152nd Meeting of the Acoustical Society of America joint with the Acoustical Society of Japan, Honolulu, Hawaii [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].

2007

4–8 June	153rd Meeting of the Acoustical Society of America, Salt Lake City, Utah [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].
27 Nov.–2 Dec.	154th Meeting of the Acoustical Society of America, New Orleans, Louisiana (note Tuesday through Saturday) [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; Web: http://asa.aip.org].

2008

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

Cumulative Indexes to the Journal of the Acoustical Society of America

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

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

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

Members of Technical and Administrative Committees of the Acoustical Society of America

The Technical and Administrative Committees listed below have been appointed by the Executive Council. These appointments, with such changes as may be made by the President from time to time, will be in effect until the Spring meeting of the Society in 2007.

Technical Committees 2006–2007

Acoustical Oceanography

N. Ross Chapman, Chair to 2007

Term to 2009
Mohsen Badiey
Michael J. Buckingham
Dezhang Chu
John A. Colosi
Christian de Moustier
Stan E. Dosso
Kenneth G. Foote
D. Vance Holliday
Andone C. Lavery
Zoi-Heleni Michalopoulou
Jeffrey A. Nystuen
David R. Palmer
Simon D. Richards
Martin Siderius
Aaron M. Thode

Term to 2008
Daniela Di Iorio
Gerald L. D'Spain Gary J. Heald
Jean-Pierre Hermand
David P. Knobles
Timothy G. Leighton
James H. Miller
Daniel Rouseff
Emmanuel K. Skarsoulis
Jerome A. Smith
Dajun Tang

Term to 2007
Kyle M. Becker
Grant B. Deane
Christopher Feuillade
Peter Gerstoft
Oleg A. Godin
John K. Horne
Bruce M. Howe
Anthony P. Lyons
Ralph A. Stephen
Kathleen E. Wage
Peter F. Worcester

Ex officio:
James F. Lynch, member of Medals and Awards Committee
Mohsen Badiey, member of Membership Committee
Anthony P. Lyons, member of ASACOS
Lora J. Van Uffelen, member of Student Council

Animal Bioacoustics

Richard R. Fay, Chair to 2009

Term to 2009
Christopher W. Clark
William C. Cummings
Robert R. Dooling
David K. Mellinger
Arthur H. Popper
Annemarie Surlykke

Term to 2008
Whitlow W.L. Au
Kelly J. Benoit-Bird
W. Tecumseh Fitch
Dorian S. Houser
Lee A. Miller
Larry L. Pater
Hiroshi Riquimaroux
Peter M. Scheifele
James A. Simmons
Edward J. Walsh

Term to 2007
Sheryl L. Coombs
Edmund R. Gerstein
Seth S. Horowitz
Cynthia F. Moss
Jeannette A. Thomas

Ex officio:
James A. Simmons, member of Medals and Awards Committee
Andrea M. Simmons, member of Membership Committee
Ann E. Bowles, member of ASACOS
Marla M. Holt, member of Student Council

Architectural Acoustics

Lily M. Wang, Chair to 2007

Term to 2009
Nils-Ake Andersson
C. Walter Beamer, IV
Leo L. Beranek
Sergio Beristain
Jim X. Borzym
Erica E. Bowden
David T. Bradley
David Braslau
Todd L. Brooks
Courtney B. Burroughs
Paul T. Calamia
Alexander U. Case
William J. Cavanaugh
Dan Clayton
Jessica S. Clements
Elizabeth A. Cohen
David A. Conant
Damian Doria
John Erdreich

Robin S. Glosemeyer
Timothy E. Gulsrud
Byron W. Harrison
Robert D. Hellweg
Murray R. Hodgson
Ian B. Hoffman
Jin Jeon
James A. Johnson
Jian Kiang
Bertram Y. Kinzey, Jr.
Mendel Kleiner
Alexis D. Kurtz
Timothy W. Leishman
Jerry G. Lilly
Edward L. Logsdon
Peter A. Mapp
David E. Marsh
Gregory A. Miller
Hideo Miyazaki
Matthew A. Nobile
Christian Nocke
Bruce C. Olson
Cornelius H. Overweg
Richard J. Peppin
Stephen D. Pettyjohn
Scott D. Pfeiffer
Norman H. Philipp
James E. Phillips
Joseph Pope
Jens Holger Rindel
Carl J. Rosenberg
Kenneth P. Roy
Hiroshi Sato
Melvin L. Saunders
Ron Sauro
Paul D. Schomer
Kevin P. Shepherd
Yasushi Shimizu
Gary W. Siebein
Abigail E. Stefaniw
Christopher A. Storch
Jason E. Summers
Louis C. Sutherland
Jiri Tichy
Nancy S. Timmerman
Brandon D. Tinianov
Gregory C. Tocci
Rendell R. Torres
Alfred C.C. Warnock
George P. Wilson
Ning Xiang

Term to 2008
Wolfgang Ahnert
Christopher N. Blair
John S. Bradley
Christopher N. Brooks
Angelo J. Campanella
Quinsan Ciao
Robert C. Coffeen
Peter D'Antonio
Felicia M. Doggett
William Dohn
Timothy J. Foulkes
Richard D. Godfrey
Tyrone Hunter
Clare M. Hurtgen
J. Christopher Jaffe
Mendel Kleiner
Jeff P. Kwolkoski
Brad W. Lewis
Stephen J. Lind
David Lubman
Ralph T. Muehleisen
Michael T. Nixon
Boaz Rafaely
Daniel R. Raichel
Jack E. Randorff
Jonathan Rathsam
H. Stanley Roller
Steven R. Ryherd
B. Schulte-Fortkamp
Noral D. Stewart
Michael Vorländer
Ning Xiang

Term to 2007
Warren E. Blazier
Joseph F. Bridger
Norm Broner
Bennett M. Brooks
Steven M. Brown
Todd A. Busch
Richard H. Campbell
F. M. del Solar Dorrego
Erin L. Dugan
M. David Egan
Jesse J. Ehnert
Michael Ermann
Adam R. Foxwell
Ronald R. Freiheit
Klaus Genuit
Matthew V. Golden
Kenneth W. Good, Jr.
Brad N. Gover
Mark A. Holden
K. Anthony Hoover
Jerald R. Hyde
David W. Kahn
Martha M. Larson
Gary S. Madaras
Benjamin E. Markham
Charles T. Moritz
David L. Moyer
Edward T. Nykaza
Paul B. Ostergaard
Dennis A. Paoletti
Stephen W. Payne
Benjamin C. Seep
Neil A. Shaw
Ryan G. Sieler
Rose Mary Su
Jeff D. Szymanski
Richard H. Talaske
Michelle Vigeant

Ewart A. Wetherill
George E. Winzer
Michael R. Yantis

Ex officio:
Ewart A. Wetherill, member of Medals and Awards
Gregory C. Tocci, member of Membership Committee
George E. Winzer, member of ASACOS
Michelle C. Vigeant, member of Student Council

Biomedical Ultrasound/Bioresponse to Vibration

Michael R. Bailey, Chair to 2008

Term to 2009
Emmanuel Bossy
Anthony J. Brammer
Robin O. Cleveland
Paul A. Dayton
Mostafa Fatemi
Yuri A. Ilinskii
Peter J. Kaczkowski
Jeffrey A. Ketterling
Vera A. Khokhlova
Oliver D. Kripfgans
James C. Lacefield
James A. McAteer
Yuri A. Pishchalnikov
Tyrone M. Porter

Term to 2008
John S. Allen
Whitlow W.L. Au
Paul E. Barbone
Charles C. Church
Gregory Clement
Floyd Dunn
E. Carr Everbach
Mark S. Hamilton
Christy K. Holland
R. Glynn Holt
Elisa E. Konofagou
Subha Maruvada
Wesley L. Nyborg
James A. Simmons
Andrew J. Szeri

Term to 2007
Constantin-C. Coussios
Diane Dalecki
J. Brian Fowlkes
Kullervo H. Hynynen
T. Douglas Mast
Thomas J. Matula
Robert J. McGough
Douglas L. Miller
Ronald A. Roy
Thomas J. Royston
Kendall R. Waters
Pei Zhong

Ex officio:
Lawrence A. Crum, member of the Medals and Awards Committee
E. Carr Everbach, member of the Membership Committee and member of ASACOS
Neil R. Owen, member of Student Council

Engineering Acoustics

Dennis F. Jones, Chair to 2009

Term to 2009
Stanley L. Ehrlich
Gary W. Elko
Robert D. Finch
Guillermo C. Gaunaurd
Thomas R. Howarth
Dehua Huang
Sung Hwan Ko
Victor Nedzelnitsky
James M. Powers
P. K. Raju
Stephen C. Thompson
James E. West
George S. K. Wong

Term to 2008
Steven R. Baker
David A. Brown
Stephen C. Butler
Robert D. Corsaro
Stephen E. Forsythe
Brian H. Houston
W. Jack Hughes
Robert M. Koch
L. Dwight Luker
Arnie L. Van Buren
Kenneth M. Walsh
Daniel M. Warren
Joseph F. Zalesak

Term to 2007
Mahlon D. Burkhard
James Christoff
Fernando Garcia-Osuna
Charles S. Hayden
Jan F. Lindberg
Yushieh Ma
Elizabeth A. McLaughlin
Alan Powell
Roger T. Richards
Kenneth D. Rolt
Neil A. Shaw
James F. Tressler

Ex officio:
Mahlon D. Burkhard, member of Medals and Awards Committee and member of ASACOS
Thomas R. Howarth, member of Membership Committee
Alexandra Loubeau, member of Student Council

Musical Acoustics

James W. Beauchamp, Chair to 2008

Term to 2009

Rolf Bader
Xavier Boutillon
Jonas Braasch
Murray D. Campbell
Rene E. Causse
Antoine J. Chaigne
Neville H. Fletcher
Nicholas J. Giordano
J. M. Harrison
William M. Hartmann
William L. Martens
James M. Pyne
Daniel A. Russell
Punita G. Singh
Sten O. Ternstrom
Paul A. Wheeler
Shigeru Yoshikawa

Term to 2008

George A. Bissinger
Annabel J. Cohen
James P. Cottingham
Diana Deutsch
Neville H. Fletcher
Roger J. Hansen
Uwe J. Hansen
Peter L. Hoekje
James H. Irwin
Ian M. Lindevald
Stephen E. McAdams
Gary P. Scavone
Chris E. Waltham

Term to 2007

R. Dean Ayers
Judith C. Brown
Courtney B. Burroughs
John R. Buschert
Thomas M. Huber
Bozena Kostek
Barry Larkin
Daniel O. Ludwigsen
Thomas D. Rossing
David B. Sharp
Julius O. Smith
William J. Strong

Ex officio:

Uwe J. Hansen, member of Medals and Awards Committee
James P. Cottingham, member of Membership Committee
Ian M. Lindevald, member of ASACOS
Brian B. Monson, member of Student Council

Noise

Brigitte Schulte-Fortkamp, Chair to 2009

Term to 2009

Sergio Beristain
Susan B. Blaeser
Erica E. Bowden
Bennett M. Brooks
Ilene J. Busch-Vishniac
Angelo J. Campanella
William J. Cavanaugh
Gilles A. Daigle
Patricia Davies
Damian J. Doria
Connor R. Duke
Jesse J. Ehnert
Tony F. W. Embleton
John Erdreich
David J. Evans
Bradford N. Gover
Robert D. Hellweg
Tyrone Hunter
William W. Lang
Richard H. Lyon
Alan H. Marsh
Ralph T. Muehleisen
William J. Murphy
Joseph Pope
Daniel R. Raichel
Kenneth P. Roy
Kevin P. Shepherd
Scott D. Sommerfeldt
Kerrie G. Standlee
George S.K. Wong

Term to 2008

Elliott H. Berger
Ann E. Bowles
Frank H. Brittain
Steven M. Brown
Mahlon D. Burkhard
Robert D. Collier
Lawrence S. Finegold
Samir N. Y. Gerges
Richard D. Godfrey
Matthew V. Golden
Murray R. Hodgson
Jerry G. Lilly
Stephen J. Lind
David Lubman
George A. Luz
Matthew A. Nobile
Richard J. Peppin
Robert A. Putnam
Jack E. Randorff
Stephen I. Roth
Paul D. Schomer
Michelle E. Swearingen
Nancy S. Timmerman
Brandon D. Tinianov

Gregory C. Tocci
Lily M. Wang

Term to 2007
Martin Alexander
Brian E. Anderson
Keith Attenborough
John P. Barry
Leo L. Beranek
Arno S. Bommer
James O. Buntin
John C. Burgess
Jim R. Cummins
Kenneth A. Cunefare
Paul R. Donavan
Ronald R. Freiheit
Klaus Genuit
Michael L. Gross
David C. Haser
Gerald C. Lauchle
George C. Maling
Thomas R. Norris
John P. Seiler
Noral D. Stewart
Louis C. Sutherland
Jiri Tichy
D. Keith Wilson
Ning Xiang

Ex officio:
Nancy S. Timmerman, member of Medals and Awards Committee
John Erdreich, member of Membership Committee
Richard J. Peppin, member of ASACOS
Connor R. Duke, member of Student Council

Physical Acoustics
Murray S. Korman, Chair to 2008

Term to 2009
Anthony A. Atchley
Henry E. Bass
Yves H. Berthelot
James P. Chambers
Charles C. Church
Kenneth G. Foote
Mark F. Hamilton
David I. Havelock
Philip L. Marston
Peter H. Rogers
Ronald A. Roy
James M. Sabatier
Philip S. Spoor
Larry A. Wilen
D. Keith Wilson
Evgenia A. Zabolotskaya

Term to 2008
Robert T. Beyer
Robin O. Cleveland
Lawrence A. Crum
Kenneth E. Gilbert
Robert A. Hiller
R. Glynn Holt
Bart Lipkens
Thomas J. Matula
Ralph T. Muehleisen
Harry Simpson
John S. Stroud
Richard L. Weaver
Preston S. Wilson

Term to 2007
David T. Blackstock
David A. Brown
John A. Burkhardt
Kerry W. Commander
Bruce C. Denardo
Logan E. Hargrove
D. Kent Lewis
Julian D. Maynard
George Mozurkewich
Lev A. Ostrovsky
Andrea Prosperetti
Neil A. Shaw
Victor W. Sparrow
Richard Stern
Roger M. Waxler

Ex officio:
Thomas J. Matula, member of Medals and Awards Committee
Steven L. Garrett, member of Membership Committee
Sameer I. Madanshetty, member of ASACOS
Todd A. Hay, member of Student Council

Psychological and Physiological Acoustics
Elizabeth A. Strickland, Chair to 2008

Term to 2009
Katherine H. Aerhart
Amy R. Horwitz
Glenis R. Long
Enrique A. Lopez-Poveda
Robert S. Schlauch
Stanley E. Sheft

Term to 2008
Michael A. Akeroyd
Alain de Cheveigne
Brent W. Edwards
John H. Grose
Gerald D. Kidd, Jr.
Armin Kohlrausch
Brenda L. Lonsbury-Martin
William P. Shofner
Lynne A. Werner

Term to 2007
David A. Eddins
Lawrence L. Feth

Hedwig E. Gockel
Jennifer Lentz
Mario A. Ruggero

Ex officio:
William M. Hartmann, member of the Medals and Awards Committee
Lynne A. Werner, member of Membership Committee
Brent W. Edwards, member of ASACOS
Suzanne P. Carr, member of Student Council

Signal Processing in Acoustics

David H. Chambers, Chair to 2009

Term to 2009
James V. Candy
William M. Carey
Leon Cohen
Geoffrey S. Edelson
Stanley L. Ehrlich
Brian Ferguson
Paul J. Gendron
Peter Gerstoft
William M. Hartmann
Kevin D. Heaney
William S. Hodgkiss
Paul D. Hursky
John M. Impagliazzo
Patrick J. Loughlin
Jens M. Meyer
Hassan Namarvar
Joe W. Posey
James C. Preisig
Brian D. Rapids
Edmund J. Sullivan

Term to 2008
Frank A. Boyle
Joe A. Clark
R. Lee Culver
David J. Evans
David M. Fromm
Howard A. Gaberson
David I. Havelock
Jean-Pierre Hermand
George E. Ioup
Juliette Ioup
Matti A. Karjalainen
Sean K. Lehman
Lance L. Locey
Zoi-Heleni Michalopoulou
Brian B. Monson
Joseph Pope
Leon H. Sibul
Randall W. Smith
Krykidos Tsiappoutas
James E. West
Gary R. Wilson
George S. K. Wong
Ning Xiang

Term to 2007
Max Deffenbaugh
Alireza A. Dibazar
Gary W. Elko
Alan W. Meyer
Daniel J. Sinder
David C. Swanson
Robert C. Waag
Preston S. Wilson
Lixue Wu

Ex officio:
Leon H. Sibul, member of Medals and Awards Committee
David I. Havelock, member of Membership Committee
Charles F. Gaumond, member of ASACOS
John H. Camin, member of Student Council

Speech Communication

Maureen L. Stone, Chair to 2007

Term to 2009
Jean E. Andruski
Lynne E. Bernstein
Ocke-Schwen Bohn
Suzanne E. Boyce
Ann R. Bradlow
Bruce R. Gerratt
Kenneth W. Grant
Paul E. Iverson
Benjamin R. Munson
Peggy B. Nelson
Mitchell S. Sommers
Brad H. Story
Joan E. Sussman

Term to 2008
Norma S. Barroso
Fredericka Bell-Berti
Ann R. Bradlow
Dani M. Byrd
Roger W. Chan
Alexander L. Francis
Kenneth W. Grant
Anders Lofqvist
Terrance M. Nearey
Douglas D. O'Shaughnessy
Joseph S. Perkell
Astrid Schmidt-Nielsen
Joan E. Sussman
Gary G. Weismer

Term to 2007
Patrice S. Beddor
Melissa A. Epstein
Robert Allen Fox
Helen M. Hanson
Diane Kewley-Port

Jody E. Kreiman
Andrew J. Lotto
Shrikanth S. Narayanan
Dwayne Paschall

Ex officio:
Fredericka Bell-Berti, member of Medals and Awards Committee
Abeer Alwan, member of Membership Committee
Shrikanth S. Narayanan, member of ASACOS
Jennell Vick, member of Student Council

Structural Acoustics and Vibration
Sean F. Wu, Chair to 2009

Term to 2009
Joseph M. Cuschieri
David Feit
Sabih I. Hayek
Philip L. Marston
James E. Phillips
Earl G. Williams

Term to 2008
Dean E. Capone
Joel Garrelick
Peter C. Herdic
Teik C. Lim
Thomas J. Royston
Angie Sarkissian
Richard L. Weaver
Jeffrey S. Vipperman

Term to 2007
Jeffrey E. Boisvert
Stephen C. Conlon
Linda P. Franzoni
Robert C. Haberman
Rudolph Martinez
Koorosh Naghshineh
Carl Pray
Michael F. Shaw

Ex officio:
Courtney B. Burroughs, member of Membership Committee
Mauro Pierucci, member of Medals and Awards Committee
Sabih I. Hayek, member of ASACOS
Micah Shepherd, member of Student Council

Underwater Acoustics
Kevin LePage, Chair to 2009

Term to 2009
Ralph N. Baer
John R. Buck
Chi-Fan Cheng
David R. Dowling
Roger C. Gauss
Frank S. Henyey
Paul C. Hines
Chen-Fen Huang
Marcia J. Isakson
Finn B. Jensen
Sunwoong Lee
James H. Miller
John R. Preston
Purnima Ratilal
Karim G. Sabra
John B. Schneider
Ralph A. Stephen
Alexander G. Voronovich
Kevin L. Williams
Lisa M. Zurk

Term to 2008
Juan I. Arvelo, Jr.
Pierre-Philippe Beaujean
Shira L. Broschat
Geoffrey F. Edelmann
Peter Gerstoft
John H. Glattetre
Brian T. Hefner
Jean-Pierre Hermand
Charles W. Holland
John C. Osler
Kevin B. Smith
Brian J. Sperry
Christopher T. Tindle
Alexandra I. Tolstoy

Term to 2007
David C. Calvo
Jee Woong Choi
Christian P. de Moustier
Stan E. Dosso
Nicholas C. Makris
Zoi-Heleni Michalopoulou
Tracianne B. Neilsen
Robert I. Odom
Marshall H. Orr
Gregory J. Orris
James C. Preisig
Martin Siderius
Kathleen E. Wage
Jixun Zhou

Ex officio:
Henrik Schmidt, member of Membership Committee
Eric I. Thorsos, member of Medals and Awards Committee
Joseph F. Zalesak, member of ASACOS
Andrew Ganse, member of Student Council

Administrative Committees 2006–2007

Archives and History
Julian D. Maynard, Chair to 2007

Term to 2009
Jont B. Allen

Ralph R. Goodman
David I. Havelock
Wesley L. Nyborg
Richard J. Peppin
William J. Strong

Term to 2008
Anthony A. Atchley
Leo L. Beranek
William J. Cavanaugh
Steven L. Garrett
Logan E. Hargrove
Allan D. Pierce
Victor W. Sparrow

Term to 2007
Henry E. Bass
David T. Blackstock
E. Carr Everbach
William W. Lang
David L. Moyer
Richard Stern
Rosalie M. Uchanski

Audit Committee

George V. Frisk, Chair to 2006

Term to 2007
Judy R. Dubno

Term to 2008
Diane Kewley-Port

***Books*⁺**

David L. Bradley, Chair to 2008

Term to 2009
Juan I. Arvelo
Jerry H. Ginsberg
Philip L. Marston

Term to 2008
James P. Cottingham
Nancy S. McGarr
Jeffrey A. Nystuen
Neil A. Shaw
Emily A. Tobey

Term to 2007
Stanley L. Chin-Bing
Robert C. Spindel

Ex officio:
Allan D. Pierce, Editor-in-Chief

College of Fellows

Janet M. Weisenberger, Chair to 2007

Term to 2009
Thomas J. Matula
Scott D. Sommerfeldt
Stephen C. Thompson
Beverly A. Wright

Term to 2008
Stanley L. Ehrlich
E. Carr Everbach

Term to 2007
Peter G. Cable
M. David Egan
Uwe J. Hansen
Diane Kewley-Port
Thomas D. Rossing

Ex officio:
William J. Cavanaugh, past Chair
Richard H. Lyon, past Chair

Education in Acoustics

James M. Sabatier, Chair to 2009

Term to 2009
William A. Ahroon
Takayuki Arai
Anthony A. Atchley
Fredericka Bell-Berti
Suzanne E. Boyce
Robert D. Celmer
Annabel J. Cohen
E. Carr Everbach
Thomas B. Gabrielson
Steven L. Garrett
Kent L. Gee
Uwe J. Hansen
Katherine S. Harris
Elizabeth S. Ivey
Joie P. Jones
Maria B. Mody
Amy T. Neel
P. K. Raju
Deborah M. Rekart
Daniel A. Russell
M. Roman Serbyn
Victor W. Sparrow
Emily A. Tobey

Term to 2008
David T. Blackstock
Courtney B. Burroughs
Robin O. Cleveland
Kenneth A. Cunefare
D. Michael Daly
Mary Florentine
Logan E. Hargrove
Mardi C. Hastings
Peter L. Hoekje
Darrell R. Jackson
Michel T. T. Jackson
Murray S. Korman
Luc Mongeau
Neil A. Shaw
Kevin B. Smith
Ralph A. Stephen
James E. West
Wayne M.Wright

Term to 2007
George A. Bissinger
David A. Brown
Robert D. Collier
Corinne M. Darvennes

Margaritis S. Fourakis
Carole E. Gelfer
Daniel O. Ludwigsen
Sharon Y. Manuel
Philip L. Marston
Ralph T. Muehleisen
Andrew A. Piacsek
Daniel R. Raichel
Thomas D. Rossing
Ronald A. Roy
Dawn R. Schuette
Scott D. Sommerfeldt
William Thompson, Jr.
Robert A. Walkling
George S. K. Wong

Ethics and Grievances

Janet M. Weisenberger, Chair to 2008

Term to 2008
William J. Cavanaugh
Gerald L. D'Spain
Barbara G. Shinn-Cunningham

International Research and Education

Gilles A. Daigle, Chair
Sergio Beristain
Lawrence A. Crum
Malcolm J. Crocker
Samir N. Y. Gerges
Vera A. Khokhlova
William M. Hartmann
Konstantin A. Naugolnykh
Oleg Sapozhnikov
Brigitte Schulte-Fortkamp
Michael Vorlander
Suk Wang Yoon

Investments

Ilene J. Busch-Vishnaic, Chair to 2007

Term to 2009
William A. Yost

Term to 2008
Lawrence A. Crum
Richard H. Lyon

Ex officio:
David Feit, Treasurer

Medals and Awards

David L. Bradley, Chair to 2007

Term to 2009

Lawrence A. Crum	Biomedical Ultrasound/Bioresponse to Vibration
William M. Hartmann	Psychological and Physiological Acoustics
Thomas J. Matula	Physical Acoustics
Nancy S. Timmerman	Noise

Term to 2008

James A. Simmons	Animal Bioacoustics
Mauro Pierucci	Structural Acoustics and Vibration
Eric I. Thorsos	Underwater Acoustics
Ewart A. Wetherill	Architectural Acoustics

Term to 2007

Uwe J. Hansen	Musical Acoustics
Fredericka Bell-Berti	Speech Communication
Mahlon D. Burkhard	Engineering Acoustics
James F. Lynch	Acoustical Oceanography
Leon H. Sibul	Signal Processing in Acoustics

Meetings—June 2006–November 2006

Clark S. Penrod, Chair to 2008

Whitlow W.L. Au, Fall 2006, Honolulu and ASA Vice President-Elect
Fred C. DeMetz, Fall 2007, New Orleans
George V. Frisk, Vice President
Murray S. Hodgson, Spring 2005, Vancouver
James H. Miller, Spring 2006, Providence
Elaine Moran, ASA Office Manager, ex officio
Peggy B. Nelson, Fall 2005, Minneapolis
Charles E. Schmid, Executive Director, ex officio
Scott D. Sommerfeldt, Spring 2007, Salt Lake City

Meetings—November 2006–June 2007

Clark S. Penrod, Chair to 2008

Whitlow W.L. Au, Fall 2006, Honolulu and ASA Vice President-Elect
Fred C. DeMetz, Fall 2007, New Orleans
George V. Frisk, Vice President
Murray S. Hodgson, Spring 2005, Vancouver
James H. Miller, Spring 2006, Providence
Elaine Moran, ASA Office Manager, ex officio
Peggy B. Nelson, Fall 2005, Minneapolis
Charles E. Schmid, Executive Director, ex officio
Scott D. Sommerfeldt, Spring 2007, Salt Lake City
William A. Yost, Spring 2008, Paris

Membership

Peter H. Rogers, Chair to 2009

Term to 2009

Abeer Alwan	Speech Communication
Lynne E. Werner	Psychological and Physiological Acoustics
James P. Cottingham	Musical Acoustics
Thomas R. Howarth	Engineering Acoustics
Andrea M. Simmons	Animal Bioacoustics

Term to 2008

E. Carr Everbach	Biomedical Ultrasound/Bioresponse to Vibration
Courtney B. Burroughs	Structural Acoustics and Vibration
Burton G. Hurdle	International Members
John Erdreich	Noise

Term to 2007

Mohsen Badiey	Acoustical Oceanography
Steven L. Garrett	Physical Acoustics
David I. Havelock	Signal Processing in Acoustics
Henrik Schmidt	Underwater Acoustics
Gregory C. Tocci	Architectural Acoustics

Prizes and Special Fellowships

Wayne M. Wright, Chair to 2007

Term to 2009
Fredericka Bell-Berti
James E. West

Term to 2008
Uwe J. Hansen

Term to 2007
Anthony A. Atchley
Constantine Trahiotis

Public Relations

Geoffrey E. Edelmann, Chair to 2009

Term to 2009
Paul A. Baxley
Ann E. Bowles
Paul D. Hursky
Jack E. Randorff
Barbara J. Sotirin
Lora J. Van Uffelen
Kathleen E. Wage

Term to 2008
Katherine H. Kim
Ellen S. Livingston
Andrew A. Piacsek
Brigitte Schulte-Fortkamp

Term to 2007
Kelly J. Benoit-Bird
E. Carr Everbach
Christy K. Holland
Joe W. Posey
Stephen C. Thompson

Ex officio:
Allan D. Pierce, Editor-in-Chief
Elaine Moran, ASA Office Manager
Charles E. Schmid, Executive Director
Thomas D. Rossing, Echoes Editor

Publication Policy

Mark F. Hamilton, Chair to 2009

Term to 2009
Jont B. Allen
David I. Havelock
Brenda L. Lonsbury-Martin

Term to 2008
Diane Dalecki
James F. Lynch

Term to 2007
Charles C. Church
Mark F. Hamilton
Mardi C. Hastings

Ex officio:
Gilles A. Daigle, President-Elect
Allan D. Pierce, Editor-in-Chief

Regional Chapters

Juan I. Arvelo, Cochair to 2008
Elizabeth A. McLaughlin, Cochair to 2008

Brian B. Monson	Brigham Young Univ. Student Chapter
Angelo J. Campanella	Central Ohio
Robert M. Keolian	Central Pennsylvania
Ernest M. Weiler	Cincinnati
Rebecca Mercuri	Delaware Valley
Gary W. Siebein	Florida
Timothy J. Foulkes	Greater Boston
Michael J. Anderson	Inland Northwest
Neil A. Shaw	Los Angeles
Hari S. Paul	Madras, India
Sergio Beristain	Mexico City
Roger T. Richards	Narragansett
Jonathan Rathsam	Univ. of Nebraska, Student Chapter
Richard F. Riedel	New York
George A. Bissinger	North Carolina
Peter F. Assmann	North Texas
James R. Angerer	Northwest
David Lubman	Orange County
Paul A. Baxley	San Diego
David Braslau	Upper Midwest
Juan I. Arvelo	Washington, D. C.
Thomas M. Disch	Wisconsin

Ex officio:
James M. Sabatier, Chair, Education in Acoustics
David Feit, Treasurer
Connor Duke, Student Council representative

Rules and Governance

William M. Hartmann, Chair to 2008

Term to 2009
Elaine Moran
Charles E. Schmid

Term to 2008
Ilene J. Busch-Vishnaic
Tony F. W. Embleton
Richard H. Lyon

Term to 2007
William J. Cavanaugh
Floyd Dunn

Standards

Executive Committee
Paul D. Schomer, Chair (Standards Director)
Robert D. Hellweg, Vice Chair
Susan B. Blaeser, Standards Manager, ex officio

S1 Representation
John P. Sieler, Chair S1 and ASA rep. on S1
George S. K. Wong, Vice Chair S1 and ASA alternate rep. on S1

S2 Representation
Ronald L. Eshleman, Chair S2
Ali T. Herfat, Vice Chair
Sabih I. Hayek, ASA rep. on S2
Bruce E. Douglas, ASA alternate rep. on S2

S3 Representation
Craig A. Champlin, Chair S3 and ASA rep. on S3
Robert F. Burkard, Vice Chair S3 and ASA alternate rep. on S3

S12 Representation
Robert D. Hellweg, Chair S12
William J. Murphy, Vice Chair S12
Bennett M. Brooks, ASA rep. on S12
David Lubman, ASA alternate rep. on S12

International TAGs (ex officio)
Paul D. Schomer, Chair, U. S. TAG for ISO/TC 43 and ISO/TC 43/SC1
David J. Evans, Chair, U. S. TAG for ISO/TC 108
Victor A. Nedzelnitsky, U. S. Technical Advisor for IEC/TC 29

ASA Technical Committee Representatives
Whitlow W. L. Au, Chair of ASA Technical Council, ex officio
Anthony P. Lyons, Acoustical Oceanography
Ann E. Bowles, Animal Bioacoustics
George E. Winzer, Architectural Acoustics
E. Carr Everbach, Biomedical Ultrasound/Bioresponse to Vibration
Mahlon D. Burkhard, Engineering Acoustics
Ian M. Lindevald, Musical Acoustics
Richard J. Peppin, Noise
Sameer I. Madanshetty, Physical Acoustics
Brent W. Edwards, Psychological and Physiological Acoustics
Charles F. Gaumond, Signal Processing in Acoustics
Shrikanth S. Narayanan, Speech Communication
Sabih I. Hayek, Structural Acoustics and Vibration
Joseph F. Zalesak, Underwater Acoustics

ASA Officers
David Feit, Treasurer, ex officio
Charles E. Schmid, Executive Director, ex officio

Past Chair of ASACOS (ex officio)
Tony F. W. Embleton
Associate Editors for Standards News—JASA (ex officio)
Susan B. Blaeser
George S. K. Wong

Student Council

Brian B. Monson	Chair and Musical Acoustics
Jennell C. Vick	Speech Communication
Suzanne P. Carr	Psychological and Physiological Acoustics
Michelle C. Vigeant	Architectural Acoustics
Connor R. Duke	Noise and Regional Chapters Com. Liaison
Andrew Ganse	Underwater Acoustics
Marla M. Holt	Animal Bioacoustics
Todd A. Hay	Physical Acoustics
Lora J. van Uffelen	Acoustical Oceanography
Alexandra Loubeau	Engineering Acoustics
John Camin	Signal Processing in Acoustics
Michael Canney	Biomedical/Bioresponse
Micah Shepherd	Structural Acoustics and Vibration

Tutorials
Lily M. Wang, Chair to 2009

Term to 2009
Gerald D. Kidd

Term to 2008
Kenneth A. Cunefare
David R. Dowling
Barbara G. Shinn-Cunningham

Term to 2007
Ann R. Bradlow
James V. Candy
James P. Chambers

Ex officio:
Charles E. Schmid, Executive Director

Women in Acoustics
Lisa M. Zurk, Chair to 2009

Term to 2009
Kathryn W. Hatlestad
Carolyn J. Richie

Term to 2008
Mardi C. Hastings
Brigitte Schulte-Fortkamp
Sophie Van Parijs
Lily M. Wang

Term to 2007
Sarah Hargus Ferguson
Benjamin R. Munson
Donna L. Neff
Nancy S. Timmerman

Ex officio:
George V. Frisk, Vice President-Elect

Acoustics Today Editorial Board
Richard Stern, Chair
Elliott H. Berger
Ilene J. Busch-Vishniac
Carol Espy-Wilson
K. Anthony Hoover
James F. Lynch
Allan D. Pierce
Thomas D. Rossing
Brigitte Schulte-Fortkamp

JASA Editorial Board

Term to June 2009
S. L. Broschat, Underwater Sound
J. A. Colosi, Underwater Sound
D. R. Dowling, Underwater Sound
R. F. Freyman, Psychological Acoustics
R. C. Gauss, Undewater Sound
M. C. Hastings, Bioacoustics—Animal
R. Y. Litovsky, Psychological Acoustics
A. Lofqvist, Speech Production
B. L. Lonsbury-Martin, Physiological Acoustics
T. D. Mast, Ultrasonics and Physical Acoustics
J. J. McCoy, Mathematical Acoustics
E. Moran, Acoustical News—USA
T. D. Rossing, Education in Acoustics
C. H. Shadle, Speech Production
V. W. Sparrow, Education in Acoustics
E. J. Sullivan, Acoustic Signal Processing
R. Stern, Electronic Archives
A. I. Tolstoy, Underwater Sound
S. F. Wu, General Linear Acoustics

Term to June 2008
K. A. Attenborough, Noise
S. B. Blaeser, Acoustical News—Standards
D. S. Burnett, Computational Acoustics
W. M. Carey, Signal Processing in Acoustics
K. A. Cunefare, Noise, Its Effects and Control
D. Deutsch, Music and Musical Instruments
D. Feit, Structural Acoustics and Vibration
N. H. Fletcher, Music and Musical Instruments
K. G. Foote, Underwater Sound

K. W. Grant, Speech Perception
P. E. Iverson, Speech Perception
G. C. Lauchle, Atmospheric Acoustics and Aeroacoustics
R. A. Lutfi, Psychological Acoustics
P. L. Marston, Acoustical Reviews-Books
W. G. Mayer, Acoustical News-International
V. E. Ostashev, Atmospheric Acoustics and Aeroacoustics
L. D. Rice, Acoustical Reviews-Patents
B. Schulte-Fortkamp, Noise: Effects and Control
W. P. Shofner, Physiological Acoustics
M. Sommers, Speech Perception
R. A. Stephen, Underwater Sound
B. H. Story, Speech Production
J. E. Sussman, Speech Perception
A. J. Szeri, Ultrasonics and Physical Effects of Sound
R. M. Waxler, General Linear Acoustics
G. S. K. Wong, Acoustical News—Standards
A. J. Zuckerwar, Applied Acoustics; Transduction; Acoustical Measurements

Term to June 2007
W. W. L. Au, Bioacoustics-Animal
P. E. Barbone, Ultrasonics; Physical Effects of Sound
Y. H. Berthelot, Ultrasonics, Physical Effects of Sound
A. R. Bradlow, Speech Perception
C. C. Church, Bioacoustics-Biomedical
A. J. M. Davis, General Linear Acoustics
F. Dunn, Bioacoustics-Biomedical
L. P. Franzoni, Structural Acoustics andVibration
J. H. Grose, Psychological Acoustics
J. H. Ginsberg, Structural Acoustics and Vibration
M. F. Hamilton, Nonlinear Acoustics
M. S. Howe, Atmospheric Acoustics and Aeroacoustics
M. Kleiner, Architectural Acoustics
J. G. McDaniel, Structural Acoustics and Vibration
D. D. O'Shaughnessy, Speech Processing and Communication Systems
A. J. Oxenham, Psychological Acoustics
R. Raspet, Ultrasonics, Physical Effects of Sound
W. L. Siegmann, Underwater Acoustics
J. A. Simmons, Bioacoustics
L. C. Sutherland, Atmospheric Acoustics and Aeroacoustics
L. L. Thompson, General Linear Acoustics
R. L. Weaver, Structural Acoustics and Vibration
E. G. Williams, Structural Acoustics and Vibration
N. Xiang, Architectural Acoustics

Associate Editors of JASA Express Letters (JASA-EL)

Term to 30 June 2009
D. S. Burnett, Computational Acoustics
J. V. Candy, Acoustic Signal Processing
C. C. Church, Bioacoustics
M. F. Hamilton, Nonlinear Acoustics
J. M. Hillenbrand, Speech Perception
A. Lofqvist, Speech Production
J. F. Lynch, Underwater Acoustics
B. L. Lonsbury-Martin, Physiological Acoustics
T. J. Matula, Ultrasonics, Quantum Acoustics and Physical Effects of Sound
J. G. McDaniel, Structural Acoustics and Vibration
A. N. Norris, General Linear Acoustics
D. D. O'Shaughnessy, Speech Processing and Communication Systems and Speech Perception
T. D. Rossing, Music and Musical Instruments
M. R. Stinson, Noise
R. M. Waxler, General Linear Acoustics
N. Xiang, Architectural Acoustics

Term to 30 June 2008
I. J. Busch-Vishniac, Transduction
D. Deutsch, Musical Acoustics
Q.-J. Fu, Psychological Acoustics
V. E. Ostashev, Aeroacoustics and Atmospheric Acoustics

Term to 30 June 2007
G. B. Deane, Underwater Sound
S. G. Kargl, Nonlinear Acoustics
C. F. Moss, Bioacoustics
J. Mobley, Ultrasonics, Quantum Acoustics and Physical Effects of Sound

Ad-Hoc Committees 2006–2007

JASA–EL Editorial Advisory

D. Keith Wilson, Chair
E. Carr Everbach
Anders Lofqvist
Philip L. Marston
Andrew N. Norris
Allan D. Pierce
Michael R. Stinson

Strategic Use of Investment Income

David Feit, Chair
Anthony A. Atchley
Charles E. Schmid

Education Outreach

Uwe J. Hansen
Jules S. Jaffe
Matthew E. Poese

ASA Webpage

Stephen C. Thompson, Chair
Susan Blaeser
Andrew Ganse
Katherine Kim
Elaine Moran

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

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

- ***Tuesday, 28 November 2006, 7:00 p.m.***

ASACOS Steering Committee

- ***Wednesday, 29 November 2006, 8:30 a.m.***

ASA Committee on Standards (ASACOS). Meeting of the Committee that directs the Standards Program of the Acoustical Society.

Photos from the Providence Meetings

The four Accredited Standards Committees and nine U.S. Technical Advisory Groups held their annual administrative meetings in Providence, RI in June, along with 15 standards working groups.

There are many opportunities to participate in the development of ANSI Standards. Working group membership is open to individuals with expertise in the subject and willingness to volunteer their time and knowledge. There is no cost to participate in standards working groups. Contact the Secretariat for details about the active working groups.

Rhona Hellman's plaque was for her work as Chair of S3/WG51, which completed ANSI S3.4-2005 American National Standard Procedure for the Computation of Loudness of Steady Sounds. Paul Schomer presented the plaque at the S3 meeting.

Matthew Nobile received a plaque from S1 Chair John Seiler for Chairing S1/WG 4 and completing ANSI S1.13-2005 American National Standard Measurement of Sound Pressure Levels in Air.

Martin Alexander was presented with a plaque for chairing S12/WG 43 which completed ANSI S12.652005 American National Standard for Rating Noise with Respect to Speech Interference. William Murphy, Vice Chair of S12, presented the plaque.

Changes in S2 Leadership

The Providence meeting of S2 represented the final meeting chaired by Richard J. Peppin. Both Mr. Peppin and S2 Vice Chair David J. Evans have completed their terms of office. ASACOS Chair Paul Schomer presented them each with a plaque and sincere thanks from S2 and from ASACOS.

Paul Schomer presented a plaque to Richard Peppin in appreciation of his six years of service as Chair of Accredited Standards Committee S2.

Paul Schomer presented a plaque to David J. Evans in appreciation of his six years of service as Vice Chair of Accredited Standards Committee S2, preceded by six years as S2 Chair.

The new Chair of S2 is Ronald L. Eshleman, Director of the Vibration Institute.

The new Vice Chair of S2 is Ali T. Herfat, Director of Global Technical Services for the Copeland Corporation.

Standards News from the United States

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

American National Standards Call for Comment on Proposals Listed

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

ASA (ASC S3)

New Standards

BSR S3.47-200x, Specification of Hearing Assistance Devices/Systems (new standard)

This standard describes definitions and measurements suitable for the specification and evaluation of Hearing Assistance Devices/Systems (HADS). HADS are a group of physically configured instruments that are intended to facilitate hearing by providing amplification of an acoustic signal and/or improving the signal-to-noise ratio by means of a nonacoustic signal transmission method. These include personal assistive listening devices, auditory trainers, hearing aids, large area assistive listening systems, telephone amplifiers, alerting devices, and similar devices. (24 July 2006).

ASABE (American Society of Agricultural and Biological Engineers)

New National Adoptions

ANSI/ASABE/ISO 5007-2003, Agricultural wheeled tractors—Operator's seat—Laboratory measurement of transmitted vibration (identical national adoption): 5/25/2006.

ANSI/ASABE/ISO 5008–2002W/Cor.1-2006, Agricultural wheeled tractors and field machinery—Measurement of whole-body vibration of the operator (identical national adoption) (25 May 2006).

Project Initiation Notification System (PINS)

ANSI Procedures require notification of ANSI by ANSI-accredited standards developers of the initiation and scope of activities expected to result in new or revised American National Standards. This information is a key element in planning and coordinating American National Standards. The following is a list of proposed new American National Standards or revisions to existing American National Standards that have been received from ANSI-accredited standards developers that utilize the periodic maintenance option in connection with their standards. Directly and materially affected interests wishing to receive more information should contact the standards developer directly.

ASA (ASC S2)

BSR S2.71-200x, Guide to the Evaluation of Human Exposure to Vibration in Buildings [revision of ANSI S2.71-1983 (R2006)]

Reactions of humans to vibrations of 1 to 80 Hz inside buildings are assessed in this standard by use of degrees of perception and associated vibration levels and durations. Accelerations or velocities inside buildings may be measured to assess perceptibility and possible adverse reactions from those inside. A variety of building types and situations are covered by the use of multiplying factors applied to the basic curves. Responses are related to the event durations, frequencies of vibration, and body orientation with respect to the vibration. Project Need: To put this standard

into alignment with related ISO standards, this version updates several outdated normative references, frequency weightings, and exposure limits.

ASA (ASC S12)

BSR S12.9-Part 6-200x, Quantities and Procedures for Description and Measurement of Environmental Sound-Part 6: Methods for Estimation of Awakenings Associated with Aircraft Noise Events Heard in Homes [revision of ANSI S12.9-Part 6-2000 (R2005)]

This Standard defines noise levels that are associated with sleep disturbance in home settings in which people are familiar with the neighborhood noise environment. Project Need: This type of measurement of environmental sound is critical to the assessment of airport impacts at night.

BSR/IEEE 1329-200x, Standard Method for Measuring Transmission Performance of Handsfree Telephone Sets (revision of ANSI/IEEE 1329-1999)

This standard provides the techniques for objective measurement of electroacoustic and voice switching characteristics of analog and digital handsfree telephones (HFTs). Due to the various characteristics of HFTs and the environments in which they operate, not all of the test procedures in this standard are applicable to all HFTs. The application of the test procedures to atypical HFTs should be determined on an individual basis.

NEMA (ASC C136) (National Electrical Manufacturers Association)

BSR C136.30-200x, Roadway and Area Lighting Equipment—Pole Vibration Testing Procedure (new standard)

This standard covers the minimum vibration withstand requirements and testing procedures for poles used in roadway and area lighting applications. Project Need: This standard would assure minimum product integrity for vibration under normal use.

ASTM (ASTM International)

BSR/ASTM WK8183/F2544-200x, Standard Test Method for Determining A-Weighted Sound Power Level of Central Vacuum Power Units (new standard)

This test method calculates the overall A-weighted sound power level emitted by central vacuum power units, intended for operation in domestic applications. This standard applies to the power unit only, at the power unit location. To test the sound power level of a central vacuum at the user's location, refer to ASTM test standard F1334. Project Need: This test method describes a procedure for determining the A-weighted sound power level of small noise sources. This test method uses a nonspecial semireverberant room.

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 S2)

Reaffirmations

ANSI S2.2-1959 (R2006), Standard Methods for the Calibration of Shock and Vibration Pickups [reaffirmation of ANSI S2.2-1959 (R2001)] (19 May 2006)

ANSI S2.16-1997 (R2006), Vibratory Noise Measurements and Acceptance Requirements for Shipboard Equipment [reaffirmation of ANSI S2.16-1997 (R2001)] (22 May 2006)

ANSI S2.26-2001 (R2006), Vibration Testing Requirements and Acceptance Criteria for Shipboard Equipment (reaffirmation of ANSI S2.26-2001) (22 May 2006)

ANSI S2.48-1993 (R2006), Servo-Hydraulic Test Equipment for Generating Vibration—Methods of Describing Characteristics [reaffirmation of ANSI S2.48-1993 (R2001)] (22 May 2006)

ANSI S2.71-1983 (R2006), Guide to the Evaluation of Human Exposure to Vibration in Buildings [reaffirmation and redesignation of ANSI S3.29-1983 (R2001)] (19 May 2006)

Revisions

ANSI S2.70-2006, Guide for the Measurement and Evaluation of Human Exposure to Vibration Transmitted to the Hand [revision of ANSI S3.34-1986 (R1997)] (19 May 2006)

Withdrawals

ANSI S2.47-1990, Vibration of Buildings—Guidelines for the Measurement of Vibrations and Evaluation of Their Effects on Buildings [withdrawal of ANSI S2.47-1990 (R2001)] (19 May 2006)

ASA (ASC S3)

Reaffirmations

ANSI S3.36-1985 (R2006), Manikin for Simulated in-situ Airborne Acoustic Measurements [reaffirmation of ANSI S3.36-1985 (R2001)] (27 April 2006)

ANSI S3.44-1996 (R2006), Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment [reaffirmation of ANSI S3.44-1996 (R2001)] (27 April 2006)

ASA (ASC S12)

Reaffirmations

ANSI S12.1-1983 (R2006), Guidelines for the Preparation of Standard Procedures to Determine the Noise Emission from Sources [reaffirmation of ANSI S12.1-1983 (R2001)] (3 May 2006)

ANSI S12.3-1985 (R2006), Statistical Methods for Determining and Verifying Stated Noise Emission Values of Machinery and Equipment [reaffirmation of ANSI S12.3-1985 (R2001)] (3 May 2006)

ANSI S12.17-1996 (R2006), Impulse Sound Propagation for Environmental Noise Assessment [reaffirmation of ANSI S12.17-1996 (R2001)] (3 May 2006)

ANSI S12.19-1996 (R2006), Measurement of Occupational Noise Exposure [reaffirmation of ANSI S12.19-1996 (R2001)] (23 May 2006)

ANSI S12.23-1989 (R2006), Method for the Designation of Sound Power Emitted by Machinery and Equipment [reaffirmation of ANSI S12.23-1989 (R2001)] (3 May 2006)

New National Adoptions

ANSI S12.55-2006/ISO 3745:2003, Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision methods for anechoic and hemi-anechoic rooms [identical national adoption and revision of ANSI S12.35-1990 (R2001)] (27 April 2006)

ATIS (Alliance for Telecommunications Industry Solutions)

New Standards

ANSI ATIS 0600005-200x, Acoustic Measurement (new standard) (16 May 2006)

U.S. Technical Advisory Groups Revised Operating Procedures American Society of Mechanical Engineers (ASME International)

The American Society of Mechanical Engineers (ASME International) has submitted a single set of revised operating procedures for the U.S. Technical Advisory Groups to the following ISO Technical Committees and Subcommittees that it currently administers:

TC 39: Machine tools (includes SC 2, Test conditions for metal cutting machine tools; SC 6, Noise of machine tools; SC 8, Work holding spindles and chucks) (10 July 2006)

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

TRACTORS AND MACHINERY FOR AGRICULTURE AND FORESTRY (TC 23)

ISO 22867/Cor1:2006, Forestry machinery—Vibration test code for portable hand-held machines with internal combustion engine—Vibration at the handles

IEC Standards

WIND TURBINE GENERATOR SYSTEMS (TC 88)

IEC 61400-11 Amd.1 Ed. 2.0 en: 2006, Amendment 1—Wind turbine generator systems—Part 11: Acoustic noise measurement techniques

ISO Draft Standard

ACOUSTICS (TC 43)

ISO 3822-1/DAmd1, Acoustics—Laboratory tests on noise emission from appliances and equipment used in water supply installations—Part 1: Method of measurement—Amendment 1: Measurement uncertainty—(5 August 2006)

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO/DIS 18436-4, Condition monitoring and diagnostics of machines—Requirements for training and certification of personnel—Part 4: Field lubricant analysis (23 July 2006)

ISO/DIS 18436-8, Condition monitoring and diagnostics of machines—Requirements for training and certification of personnel—Part 8: Thermography (23 July 2006)

SMALL CRAFT (TC 188)

ISO/DIS 14509-1, Small craft—Airborne sound emitted by powered recreational craft—Part 1: Pass-by measurement procedures (26 August 2006)

IEC Draft Standard

29/600/FDIS, IEC 60318-5 Ed.1: Electroacoustics—Simulators of human head and ear—Part 5: 2 cm 3 coupler for the measurement of hearing aids and earphones coupled to the ear by means of ear inserts (30 June 2006)

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

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

U.S. TAG	ISO and IEC documents
S1	**IEC/FDIS 60318-5 (29/600/FDIS)** "Electroacoustics—Simulators of human head and ear—Part 5: 2 cm^3 coupler for the measurement of hearing aids and earphones coupled to the ear by means of ear inserts" (18 June 2006)
	IEC 29/599/NP "Audio-frequency induction loop systems for assisted hearing—Method of measuring and specifying the performance of system components" (9 June 2006)
	IEC 29/598/NP "Audio-frequency induction loop systems for assisted hearing—Method of measuring the low frequency magnetic field emissions from the loop of assessing conformity with guidelines on limits for human exposure" (9 June 2006)
S12	**ISO/FDIS 16832** "Acoustics—Loudness scaling by means of categories" (12 June 2006)

BOOK REVIEWS

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

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

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

Vertebrate Hair Cells

Edited byRuth Anne Eatock, Richard R. Fay, and Arthur N. Popper

Springer, New York, 2006, 454 pp. Price: $129.00 (hardcover). ISBN: 0387952020

This book is the 27th of a series known as the "Springer Handbook of Auditory Research," which now constitutes the most comprehensive library of information on the auditory system. Although the sensory receptor cells of the inner ear—the hair cells—have been dealt with in other volumes of this series, this is the first to concentrate exclusively on the hair cells themselves. Rather than present a dozen shorter chapters covering all aspects, the editors chose to present 8 chapters, some of them quite extensive, on important features, most of which have been the subject of recent intense research. The authors are among the best-known workers in the respective research areas.

Put into perspective, hair cells are quite remarkable. Whereas our eyes work with more than 100 million receptor cells each, each of our ears has only about 30,000 hair cells at birth and, of these, only about one-quarter actually transmit information to the brain—that is four orders of magnitude fewer receptors than in the eye! Obviously, these are rather special cells that are worthy of a lot of attention.

Most of the book concentrates on the cochlea, but a final chapter deals with mammalian vestibular hair cells. The review of cochlear hair cells includes their development, their structure, and their function, including the physiology of the innervating afferent and efferent nerve fibers. Each of the comprehensive chapters can be read as a unit, and the number of very recent references demonstrates that they are up to date. Although the emphasis is on mammalian hair cells, many studies have been carried out using nonmammals, such as chickens, turtles and frogs, and these receive appropriate attention.

Recent advances in molecular biology have made it possible to gain much more insight into the processes guiding development, and a substantial chapter by Goodyear, Kros, and Richardson not only covers the genetics and the roles of many proteins in development but integrates it nicely into structural and electrophysiological development. This is accompanied by a number of new and very useful illustrations. It is refreshing to see authors take the time to avoid recycling old figures!

The structural basis of mechanoelectrical transduction by the hair-cell bundle is examined by Furness and Hackney. This chapter is mostly about the nature and the spatial localization of a large number of proteins thought to be involved in connecting bundle stereovilli, in the transduction channels and the cytoskeleton. Recent findings indicate that the tip link itself is not the so-called gating spring, but rather the ankyrin repeats that connect the cytoplasmatic side of the channel to the cytoskeleton. Transduction examination using physiological techniques is discussed by Fettiplace and Ricci. Here, the extensive work carried out on turtle auditory and frog vestibular hair cells—using sophisticated micromechanical measurements, electrophysiology, and chemical manipulation of channels and ionic concentrations—is thoroughly discussed. There is a useful discussion of the various aspects of channel adaptation, and this is particularly interesting with respect to the apparent linkage between adaptational phenomena and active processes involving hair-cell channels. Fewer data are available from mammals, but those that are indicate a fundamentally similar situation but with faster kinetics. Unfortunately, two recent important pieces of work on active movements in mammalian hair-cell bundles were obviously published at a time permitting only a cursory mention in the very last sentence.

Electrical frequency tuning of hair cells has not been demonstrated in mammals, but in many nonmammalian auditory and in vestibular hair cells. Although much of the data discussed in a chapter by Art and Fettiplace is not new, electrical tuning does present new challenges and opportunities for the molecular study of hearing processes. Tuning to specific frequencies in these cells is achieved by a particular combination of channel numbers, kinetics, and accessory channel subunits, and some of the proteins involved are known to have a number of regions of alternative splicing. This combination of factors apparently makes it possible to create thousands of different combinations that permit a tonotopic arrangement of hair cells over a wide frequency range. The challenge for the future will be to show how a hair cell's frequency response is determined, and how neighboring cells can have just very slightly differing response ranges.

Fuchs and Parsons' chapter on synaptic physiology emphasizes that the afferent synapse of auditory hair cells is very interesting. Not only are some capable of sustaining spontaneous firing rates of over 100/s in the nerve fibers, many can sustain even higher response rates over long periods of time, and some phase lock to auditory stimuli of nearly 10 kHz (in the barn owl). These amazing feats require mechanisms that can recycle and replenish synaptic vesicles extremely rapidly, and it appears that the enigmatic and little-understood synaptic bodies and ribbons play an important role here. This is an active field that will produce many interesting results in the future. Curiously, the section on efferent synapses concentrates exclusively on the effects acetylcholine receptors, and does not even mention that other transmitters can be found, at least as co-transmitters.

I found the chapter by Brownell on piezoelectric effects of hair-cell membranes difficult to follow, but this reflects my own lack of training in physics. Brownell acknowledges the roles of bundle- and cell-membrane prestin motors in cochlear amplification, but emphasizes that piezoelectric effects are also necessary, especially to cope with extremely high frequencies. However, I did understand some claims that Brownell makes regarding the evolution of this system, and find them unacceptable. The fact that "the spiral-shaped cochlea appeared with the OHC" (outer hair cells) is manifestly untrue, since OHC are found in egg-laying mammals that do not have a coiled cochlea. In addition, the first mammals were not "rodentlike," and almost certainly were not capable of ultrasonic hearing. These minor details do not, however, detract from the chapter's thrust.

The longest chapter, by Eatock and Lysokowski on vestibular hair cells, very nicely covers in the one chapter what the rest of the book covers for auditory hair cells. It reviews all aspects of the structures and functions of different types of vestibular hair cells, and is a joy to read. It includes a number of very useful new illustrations, for example, summarizing the complex molecular mechanisms of afferent and efferent synapses in type I and type II vestibular hair cells.

This book will be a very useful reference work for all who work with vertebrate hair cells, and for those who are looking for comprehensive reviews of the most important aspects of the structure and function of these amazing sensory cells.

GEOFFREY A. MANLEY
Technische Universitaet Muenchen
Lehrstuhl Fuer Zoologie
Lichtenbergstr. 4
85747 Garching, Germany

Collected Works of Rohn Truell

Edited byC. F. Ying

Chinese Academy of Sciences, Beijing

This book has an interesting history. In 1948, Rohn Truell established the Metals Research Laboratory in the Applied Mathematics Department at Brown University. Under Truell's direction, researchers in this laboratory performed pioneering ultrasonic studies of solids in many areas. The effects of point defects and dislocations on ultrasonic attenuation and velocity were investigated. In 1956, the dislocation studies led to the development, by A.V. Granato and K. Lucke, of the very successful Granato-Lucke string model of dislocation damping. The theory of the scattering of ultrasonic waves by point defects and by macroscopic obstacles was considered in a series of papers by Ying, Einspruch, Waterman, and Truell. Later work investigated nonlinear sound propagation in solids, and the anharmonic interaction between a sound wave and thermal phonons. A very important and lasting contribution was made in the area of electronic instrumentation. B. B. Chick, working together with G. P. Anderson, Truell, and others, developed electronics that could be used to make accurate attenuation and velocity measurements over the frequency range from 1 MHz up to several GHz. This instrumentation was later made commercially available, and has since been used in a large number of ultrasonics laboratories around the world.

Much of this work was summarized in the well-known text "Ultrasonic Methods in Solid State Physics" by R. Truell, C. Elbaum, and B. B. Chick, published in 1969 by Academic Press. This book appeared shortly after the death of Truell in 1968, and has since continued to be a valuable resource for workers in the field.

C. F. Ying came from China in 1948 to work in the Metals Research Laboratory soon after it was founded. During the Korean war, the U.S. government did not allow him to return to China, since it was considered that Chinese scientists might provide help to the North Korean war effort. He finally returned to China in 1955. There, he founded a research group in the Institute of Applied Sciences of the Chinese Academy of Sciences to study ultrasonics, and concentrated on introducing ultrasonic techniques to solve many practical problems in both industry and in medicine. Ying has always maintained contact with Brown University, and has very good memories of the time that he spent working with Truell. Last year, he decided to organize a collection of Truell's papers to distribute to interested acousticians. Two hundred copies have been printed, and some of these have already been delivered to former members of the Metals Research Laboratory, to acoustics institutes, and to other interested acousticians. Currently, there remain a number of copies at Brown University and the author of this review will be happy to send a copy to any reader who requests one.

HUMPHREY J. MARIS
Department of Physics,
Brown University, Providence,
Rhode island 02912

 0001-4966/2006/120(3)/1160/1/$22.50

OBITUARY

John W. Kopec●1936–2004

John W. Kopec, a long-time member of the Acoustical Society of America, died in Chicago, Illinois, on October 14, 2004 at the age of 67. Kopec passed away after a vigorous recurrence of leukemia, from which he had enjoyed a lengthy sabbatical after his original 1998 diagnosis.

John Kopec was born in Chicago, on 5 November 1936, the first of two children of the late John Sr. and Marie Kopec. After graduation from high school in Chicago, he served in the U. S. Air Force where he received training in radio, navigation, radar, and teletype maintenance and repair, as well as in electronic detonation for demolition. After 4 years of military service, Kopec studied electrical engineering and computer science, and earned an Associates degree from the Illinois Institute of Technology (IIT).

Kopec began working for IIT Research Institute (IITRI) (now Alion Science and Technology) in 1961 as an Assistant Experimentalist, and became a major contributor to the fire and explosives research being conducted there. He was assigned to Riverbank Acoustical Laboratories in 1976 where he served in a number of positions, finally being named laboratory manager in 1993. He was the first three-time recipient of IITRI's *Commitment to Excellence Award*.

Kopec directed the execution of the various standard tests performed at the Riverbank Acoustical Laboratories. His supervision covered client contact, installation of materials, test procedures, and evaluation of results through narrative reports that he augmented with design drawings and schematics. He participated in numerous community noise impact studies involving highways, airports, railroads, automobiles, trucks, and various industrial complexes. His role as project engineer on noise control programs centered on portable air compressors, heating, ventilation, and air conditioning equipment, diesel engines, generators, and line pull transformers. He documented various performance specifications and calibration of specific project-oriented acoustical equipment.

Kopec was the author or coauthor of more than 25 papers on acoustics. Serving as Laboratory Supervisor and Laboratory Manager of the Riverbank Acoustical Laboratories (RAL) for 26 years (1972–1998), no one knew the history and capabilities of the unique laboratory better than John Kopec. In 1997, he published *The Sabines at Riverbank* in which he described the history of RAL from the early 1900s beginning with its founder, Colonel George Fabyan, and Professor Wallace C. Sabine, the father of the science of architectural acoustics. Kopec was recognized worldwide for his understanding of architectural acoustics and its history. His contacts in the acoustics community have been very instrumental in building the business base at RAL.

Kopec joined the Acoustical Society of America in 1977, was elected to full Membership in 1979, and to Fellowship in 1994. He served on the Committee on Architectural Acoustics (1982–2000), the Committee on Archives and History (1988–2000, Chair 1991–1994), and the Committee on Regional Chapters (1984–2004). He was the Secretary and Chapter Representative for the Chicago Regional Chapter (1983–2004), and served as Technical Program Cochair for the Spring 2001 meeting of the Acoustical Society. Kopec was a member of the Canadian Acoustical Association, the American Society for Testing and Materials, and the New York Academy of Sciences.

Kopec is survived by a son, Brian, daughter, Vaune Martens, and four grandchildren, Melissa, Candice, Matthew, and Corey.

DAVID L. MOYER
WILLIAM J. CAVANAUGH

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
JOHN M. EARGLE, *JME Consulting Corporation, 7034 Macapa Drive, Los Angeles, California 90068*
JOHN ERDREICH, *Ostergaard Acoustical Associates, 200 Executive Drive, West Orange, New Jersey 07052*
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*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
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, *University of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

7,021,560

43.10.Pr SYSTEM AND METHOD FOR AEROSOL DELIVERY

Larry B. Gray and Richard J. Lanigan, assignors to DEKA Products Limited Partnership
4 April 2006 (Class 239/303); filed 25 September 2003

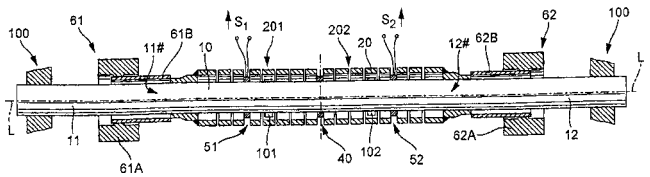
Aerosolized delivery of drugs is often advantageous over other types of drug delivery, but aerosol systems need to be able to deliver predetermined doses reliably. The delivery system described in this patent employs a variable acoustic source and a microphone, both acoustically coupled to a volume consisting of a fluid region and an air region. A fluid valve allows an amount of fluid to exit the fluid region, depending on the volume of the air region, and aersolizes a portion of this amount of fluid. A processor determines the volume of the air region on the basis of a signal from the microphone and controls the opening and closing of the fluid valve. A second processor may be provided to determine the volume of the aerosolized fluid.—EEU

7,017,424

43.35.Zc VIBRATORY TRANSDUCER

Alfred Rieder and Wolfgang Drahm, assignors to Endress Endress+Hauser Flowtec AG
28 March 2006 (Class 73/861.355); filed in Germany 8 May 2002

The patent covers a torsionally vibrating tube **10** to measure the viscosity of a fluid **12**, surrounded by tuned damper frame **20**. A transducer **40** excites the tube's torsional eigenmode of vibration, which is diminished by



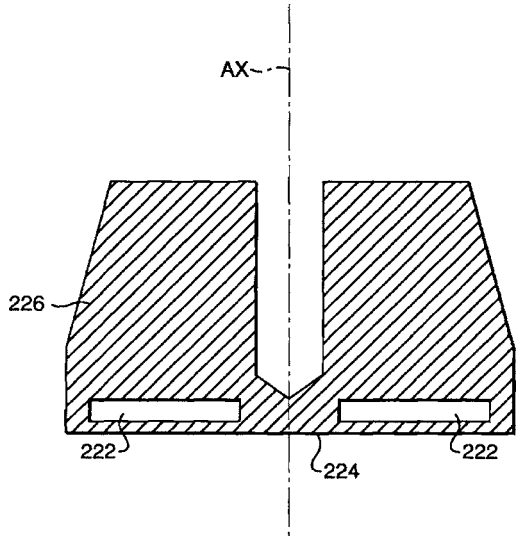
viscous fluid **12**. In addition, a tube-bending eigenmode can be excited at another frequency. The resulting Coriolis force on fluid **12** is proportional to fluid density. This result can also be detected by sensors **61**, **62**. These sensors may be any of a number of types.—AJC

7,019,439

43.35.Zc HIGH POWER ULTRASONIC TRANSDUCER WITH BROADBAND FREQUENCY CHARACTERISTICS AT ALL OVERTONES AND HARMONICS

Eugene A. DeCastro *et al.*, assignors to Blackstone-Ney Ultrasonics, Incorporated
28 March 2006 (Class 310/325); filed 30 July 2002

Longevan transducers for cleaning are described that operate at two or more frequencies among 40, 80, 120, 170, 220, and 270 kHz when driven with FM currents swept over a 10% bandwidth. An attached back mass (not shown), a mirror image of the front mass **226**, provides longitudinal



resonance. Longitudinal and radial modes are excited. One embodiment employs a flapping mode where a thin face **224** over a void **222** in the front

 0001-4966/2006/120(3)/1162/19/$22.50

mass **226** gives rise to a high-power ultrasonic resonance that directs acoustic streaming for particle removal.—AJC

7,010,963

43.38.Ar LIQUID DETECTING APPARATUS

Shinsuke Miura, assignor to CBC Materials Company, Limited
14 March 2006 (Class 73/54.24); filed in Japan 20 November 2002

This patent describes a liquid density and viscosity sensor that works on the torsional oscillator principle. A small cylindrical mass with vanes on a thin shaft is immersed in the liquid under experiment and caused to twist by the use of piezoelectric actuators disposed on a diameter of the suspension of the shaft. The frequency and damping of the torsional oscillation are used to calculate the density and viscosity of the liquid. What is new about this? It is difficult to understand due to the fact that the writing and/or translation is terrible, in two cases resulting in sentences that are 330 words (one paragraph) long!—JAH

7,019,605

43.38.Ar STACKED BULK ACOUSTIC RESONATOR BAND-PASS FILTER WITH CONTROLLABLE PASS BANDWIDTH

John D. Larson III, Palo Alto, California
28 March 2006 (Class 333/187); filed 30 October 2003

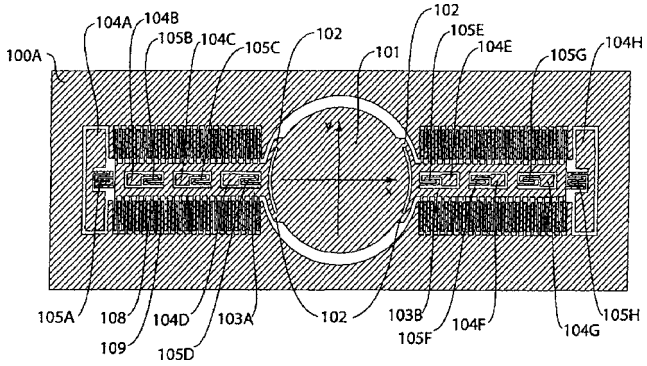
This patent describes the use of plastic as a weak acoustic coupling material used between stacked-film bulk acoustic resonators. Nothing is presented that would not have been evident to a practitioner of the art.—JAH

7,014,115

43.38.Bs MEMS SCANNING MIRROR WITH DISTRIBUTED HINGES AND MULTIPLE SUPPORT ATTACHMENTS

Yee-Chung Fu, assignor to Advanced Nano Systems, Incorporated
21 March 2006 (Class 235/454); filed 10 October 2003

One of the interesting things about MEMS sensors and actuators is that they can be integrated so tightly with the device itself. In this patent, we see how the drive actuator is most of the device, distributed all along the torsional spring elements **105A–105H**. The device is a scanning mirror device, apparently intended to operate at its resonant frequency in constant torsional



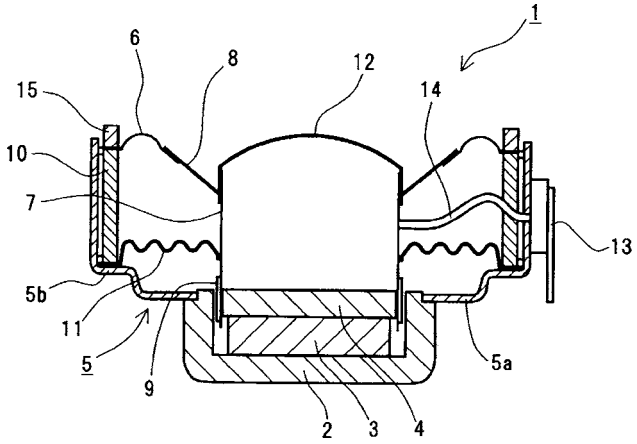
oscillation. The use of an extended set of springs is risky in that it would usually allow a lot of parasitic modes of oscillation, but this is mitigated somewhat by the distributed drive elements. The primary novelty of this device seems to be in the spring supports, which allow easy fabrication using surface micromachining techniques. The patent is brief and devoid of any dimensional or construction details.—JAH

6,993,146

43.38.Dv SPEAKER HAVING SPACER RING INSIDE FRAME

Eiji Sato, assignor to Mineba Company, Limited
31 January 2006 (Class 381/398); filed in Japan 15 February 2002

"A speaker comprises: a magnetic circuit composed of a pot yoke, a magnet, and a pole piece; a diaphragm shaped conical, defining an outer periphery and an inner periphery, and having a surround half-rolled and fixedly disposed at the outer periphery and a voice coil bobbin with a voice coil wound therein fixedly disposed at the inner periphery; a spider having



an outer diameter equal to or larger than the outer diameter of the surround; a frame having the magnetic circuit fixedly disposed at its central opening; and a spacer ring fixedly disposed inside the frame, and having a surround edge receiving portion adapted to fixedly receive the entire outer edge of the surround thereon."—NAS

7,017,419

43.38.Dv MICRO-MECHANICAL CAPACITIVE INDUCTIVE SENSOR FOR WIRELESS DETECTION OF RELATIVE OR ABSOLUTE PRESSURE

Michael Pedersen *et al.*, assignors to Corporation for National Research Initiatives
28 March 2006 (Class 73/718); filed 31 August 2004

What do you get when you combine a capacitive pressure gauge and a coil on different levels of the same circuit board? A wireless sensor! This patent describes the way in which the inventors think this should be done, on a low-temperature cofired ceramic (LTCC) circuit board. The whole concept has been done before, just the materials are different.—JAH

7,019,437

43.38.Fx HIGH-EFFICIENCY PIEZOELECTRIC SINGLE-PHASE UNI-POLAR ULTRASONIC ACTUATORS WITH A NOTCHED PZT BACK DISC

Min-Shen Ouyang and Swe-Kai Chen, both of Hsinchu, Taiwan, Province of China
28 March 2006 (Class 310/323.08); filed 4 September 2003

This patent covers a funny little motor that spins due to rotational asymmetries in its construction. The authors argue that asymmetry in a stacked-disk piezoelectric actuator can cause thickness-mode vibrations to be coupled efficiently to rotational motion of a backing plate. The concept of failing to utilize pre-existing mode structures of circular symmetry seems odd, but the authors maintain that radial slots and screws placed slightly off equi-angular divisions work well, allowing "the maximum speed of the

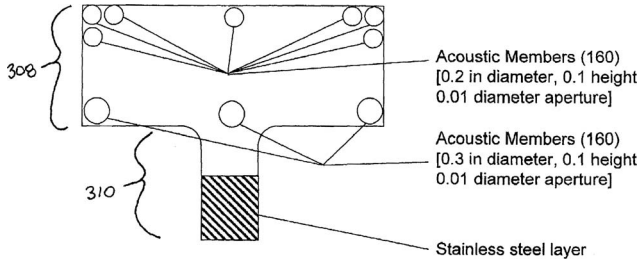
rotor to reach up to above 3000 rpm.'' Well, one should hope so. The reader is advised to look elsewhere for insights into ultrasonic motors.—JAH

7,019,621

43.38.Fx METHODS AND APPARATUS TO INCREASE SOUND QUALITY OF PIEZOELECTRIC DEVICES

Stanley E. Woodard, Hampton, Virginia *et al.*
28 March 2006 (Class 340/384.6); filed 2 January 2001

The authors describe how one can make a broadband piezoelectric transducer that can produce vibrations (buzzer) as well as produce good sound. They propose ''T''-shaped unimorph and bimorph actuators having various holes through them similar to those shown in the figure. They stress



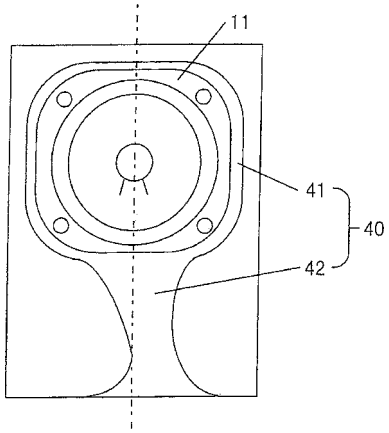
the ability of this transducer to generate broadband vibration, but who needs such a thing? A device of this general structure would also seem to have an irregular acoustic response with general lack of bass. No design or performance information is given.—JAH

7,021,419

43.38.Ja SPEAKER SYSTEM

Koichi Sadaie and Kenichiro Toyofuku, assignors to Onkyo Corporation
4 April 2006 (Class 181/156); filed in Japan 3 March 1999

This is a rewrite of United States Patent 6,798,891, originally filed in Japan in 1999, reviewed in J. Acoust. Soc. Am. **117**(4), 1692 (2005). That patent described a small, sealed-box system in which sound from one or two loudspeakers **11** emerged from a side-firing flared vent **42**. Test results of ten exemplary systems have now been included, apparently to demonstrate that the basic idea really does work. Well, yes—in the sense that an additional air mass coupled to the cone of a loudspeaker lowers the system resonance. The



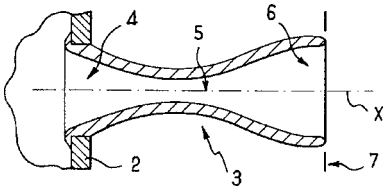
technique was patented more than 50 years ago and is used in a number of commercial subwoofers. This new patent also includes data demonstrating that a flared vent produces less turbulence than a cylindrical vent (true) and that lining the vent with absorptive material attenuates high-frequency noise (true again). However, the implication that this geometry can magically ''widen the bass reproduction band'' for a given size and efficiency is not true.—GLA

7,011,178

43.38.Ja SOUND TRANSMITTER AND SPEAKER

Jean-Pierre Morkerken, Paris, France
14 March 2006 (Class 181/151); filed in France 15 May 2001

At high sound levels, vented loudspeaker systems get into trouble. Turbulence in the vent introduces unwanted damping and partial rectification. During the past 5 years, we have seen considerable research into vent



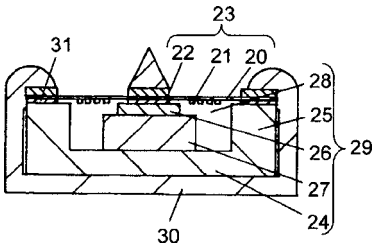
turbulence and practical ways to minimize it. A number of patents have already been granted, however, the application for this patent dates back to 2001, so it may predate some of them. Like other designs, a flared vent is described but this one includes a bell-shaped section **6** at one end.—GLA

7,020,301

43.38.Ja LOUDSPEAKER

Satoshi Koura *et al.*, assignors to Matsushita Electric Industrial Company, Limited
28 March 2006 (Class 381/421); filed in Japan 5 November 2001

Commercial ribbon tweeters have been available since the 1930s. A number of variants have been developed, including what this patent refers to as a ''leaf tweeter,'' in which a flat, multiple-turn voice coil **21** is bonded to a nonmetallic diaphragm **20**. The overall shape can be circular or rectangular. In this new version, magnet **27** overlaps pole piece **26**, permitting the use of a larger and/or shallower magnet. Also, the stepped gap tends to



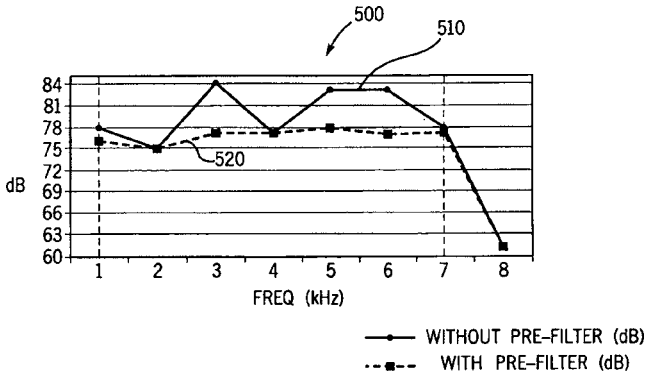
concentrate a greater portion of the magnetic field in the voice coil region.—GLA

7,013,281

43.38.Lc SOUND GENERATOR CIRCUIT PRE-FILTER SYSTEM AND METHOD

James B. Henrie and Wayne B. Hile, assignors to Palm, Incorporated
14 March 2006 (Class 704/270.1); filed 12 March 2001

At one time, GenRad produced a multitone calibrator for sound level meters. The transducer response was far from flat, but internal circuitry adjusted the amplifier gain for each frequency. This patent describes a similar concept. Many electronic devices include a digital multitone generator that can play tunes. By adding a lookup table matched to the particular



transducer used, the levels of individual frequencies can be held more or less constant.—GLA

6,992,570

43.38.Md REST ROOM SOUND PRODUCING DEVICE

Linda A. Weinstein Ide, Bronx, New York
31 January 2006 (Class 340/384.1); filed 22 May 2003

Sometimes just the abstract says it best: ''The device is located with the (rest)room and can be activated to produce a sound upon a person entering the cubicle such that the sound so produced covers up the normal bodily noises of a person using the rest room.''—MK

6,998,961

43.38.Md ALARM CLOCK WITH VOICE MESSAGE INPUT

Shirley Ann Stihler, Del Rey Oaks, California
14 February 2006 (Class 340/309.9); filed 28 January 2004

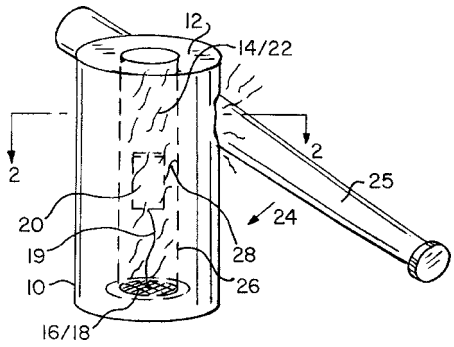
Once again, the ubiquitous sound chip makes an appearance. The basic idea is to permit an alarm clock to play a wake-up call. Additionally, the alarm clock can be connected to a telephone, thereby also serving as an answering machine. Imagine the possibilities: parents could yell at sleeping teenagers, reveille for boy scouts, the possibilities are endless.—MK

7,006,001

43.38.Md SPEECH-EMITTING CELEBRATION DEVICE

Mark A. Estrada, El Paso, Texas *et al.*
28 February 2006 (Class 340/692); filed 4 June 2002

Take the ubiquitous sound chip **20**—activate it with a shock sensor **26**



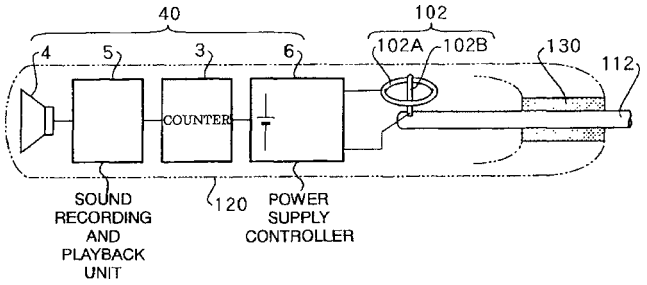
and you have a talking (or singing) piñata.—MK

7,013,522

43.38.Md TOOTHBRUSH ASSEMBLY WITH SOUND GENERATING FUNCTION

Toshinori Kumagai, assignor to Ablecorporation, Limited
21 March 2006 (Class 15/105); filed in Japan 1 July 2002

Ostensibly for children, this automated toothbrush has the ubiquitous sound chip **5**. By monitoring the number of rotations of the brush, differing



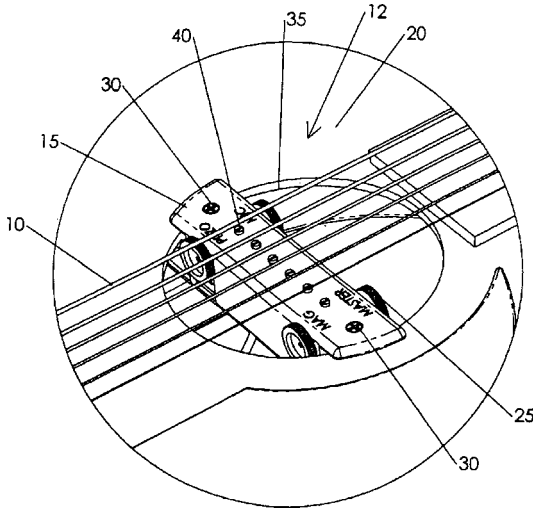
encouraging messages can be given. However, without sensing the torque, how will it know that it's actually in the mouth? Next step: talking floss.—MK

7,015,390

43.38.Md TRIAD PICKUP

Wayne A. Rogers, Titusville, Florida
21 March 2006 (Class 84/723); filed 15 January 2004

Simply put, the inventor wants to combine the outputs of a magnetic



guitar pickup, a microphone, and a piezoelectric pickup, all in one removable assembly. A mixer schematic is provided.—MK

7,015,848

43.38.Md RECORDING/REPRODUCING APPARATUS USING REFERENCE OSCILLATOR FOR DIGITAL AUDIO

Masamichi Ohashi and Masamichi Tsuchiya, assignors to Nippon Sogo Seisaku Company Limited
21 March 2006 (Class 341/155); filed in Japan 12 February 2002

Much is made of the importance of using a low-jitter reference clock in the recording chain. Wouldn't it make sense to use an atomic clock with the ion of your choice? Wouldn't this be obvious? Apparently not.—MK

7,023,770

43.38.Md ENCODING RATE-BASED, PROGRAMMABLE, LINEAR VELOCITY, COMPRESSED AUDIO COMPACT DISC PLAYER

Stephen J. Fedigan and Thomas N. Millikan, assignors to Texas Instruments Incorporated
4 April 2006 (Class 369/47.33); filed 12 September 2002

A simple way to change different playback rates in a compact disc is to change the rotation speed of the motor depending on the encoding. That's it.—MK

6,987,992

43.38.Si MULTIPLE WIRELESS MICROPHONE SPEAKERPHONE SYSTEM AND METHOD

Sukhdeep S. Hundal and Supajet Guy Pothiboon, assignors to VTech Telecommunications, Limited
17 January 2006 (Class 455/569.1); filed 8 January 2003

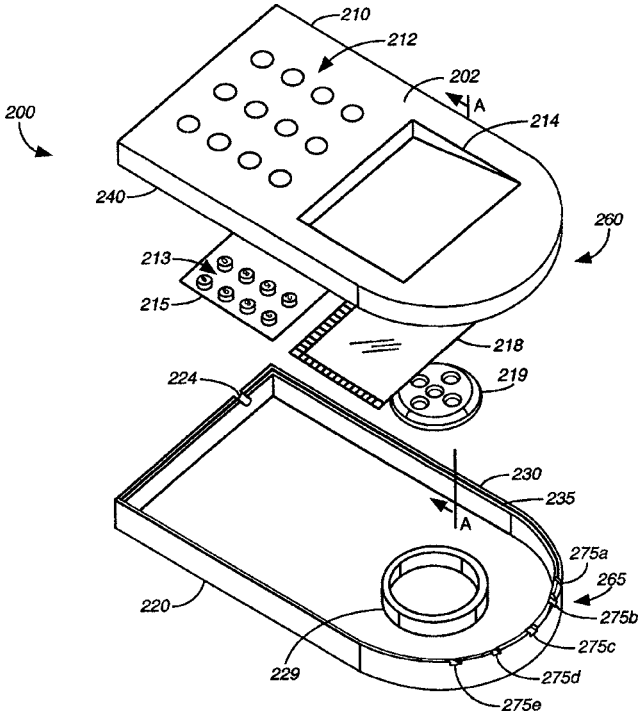
A system is described that appears to embody multiple instances of prior art fashioned in a system that does not appear at all novel—with 23 claims—wireless microphones and receivers, automatic mixing, adaptive echo cancelling, among others.—NAS

6,990,197

43.38.Si HANDS-FREE SPEAKER TELEPHONE

John Patrick Wong *et al.*, assignors to Nokia Corporation
24 January 2006 (Class 379/433.02); filed 8 November 2001

Here is a simpler and, as claimed in the patent, better way to build the



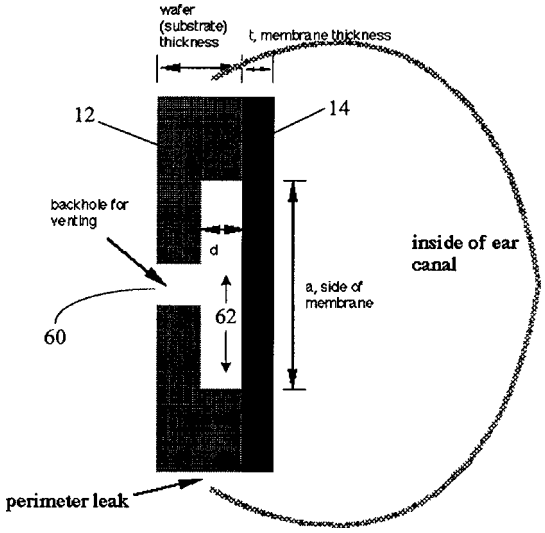
device described in the patent title. The patent principally appears to cover a means of assembly for the case of the device.—NAS

7,019,955

43.38.Si MEMS DIGITAL-TO-ACOUSTIC TRANSDUCER WITH ERROR CANCELLATION

Wayne A. Loeb *et al.*, assignors to Carnegie Mellon University
28 March 2006 (Class 361/230); filed 18 February 2004

A low-cost MEMS (micro-electrical-mechanical system) microspeaker is described comprising a silicon chip CMOS (complementary metal oxide semiconductor) diaphragm **14** and DSP (digital signal processing) drive circuits **12** that fits into an ear canal. The electrostatically biased MEMS diaphragm **14** is driven by pulses of charge from DSP circuitry within the



MEMS structure **12**, fed from an external digital-audio signal (not shown here). Diaphragm **14** can also serve as a microphone sensor, with appropriate DSP circuits added in **12**, to provide, among other things, variable capacitance feedback and air impedance-matching signals for sound-reproduction linearity and frequency-response compensation.—AJC

7,023,326

43.38.Si VIBRATION APPARATUS FOR A MOBILE TELECOMMUNICATION TERMINAL AND METHOD FOR CONTROLLING THE SAME

Seon-Woong Hwang, assignor to Samsung Electronics Company, Limited
4 April 2006 (Class 340/384.71); filed in the Republic of Korea 26 May 2001

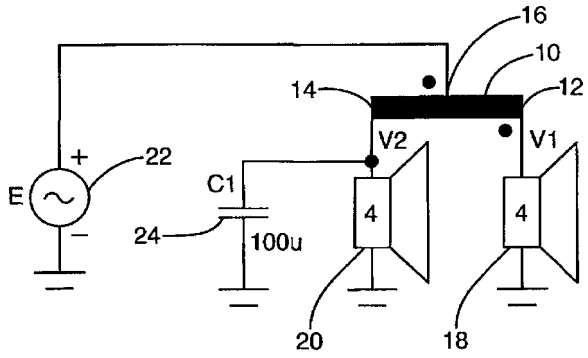
The vibrator used in a cellular phone to signal an incoming call is driven at its resonance by a digital frequency generator. However, as this patent warns, the resonant frequency can be changed by physical abuse or by adjacent objects, thus reducing the intensity of the signal. The solution suggested here is to sequentially drive the vibrator at two or more frequencies. A sensor can be added to identify and maintain the frequency of maximum vibration.—GLA

6,993,141

43.38.Tj SYSTEM FOR DISTRIBUTING A SIGNAL BETWEEN LOUDSPEAKER DRIVERS

Stefan R. Hlibowicki, assignor to Audio Products International Corporation
31 January 2006 (Class 381/111); filed 30 August 2002

The patent discloses that the sum of the voltages **V1** and **V2** remain the same regardless of the impedances of the loudspeakers **18** and **20** when



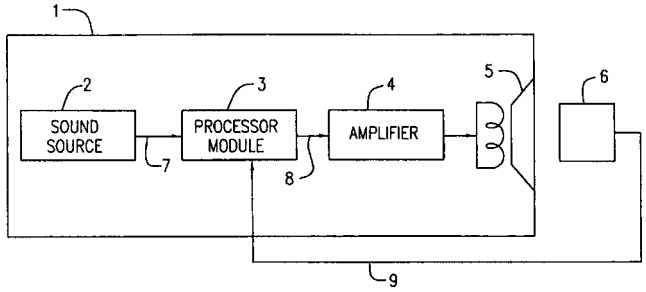
connected to a center tapped autotransformer. Ohm and Kirchoff were right.—NAS

6,996,240

43.38.Tj LOUDSPEAKER UNIT ADAPTED TO ENVIRONMENT

Atsushi Sasaki, assignor to NEC Corporation
7 February 2006 (Class 381/58); filed in Japan 21 March 1997

A means is described of correcting the sound emanating from a loudspeaker **5**, where the output of the speaker is picked up by a microphone **6**, digitally processed **3**, and compared with the original electrical signal from



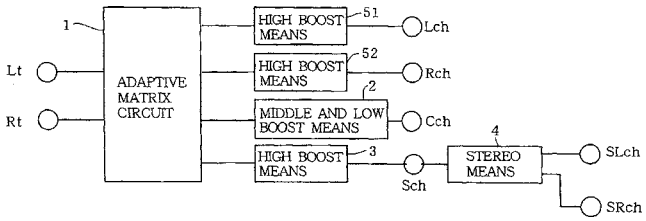
sound source **2**. Equalization is applied in the processor module **3** to adjust the electrical signal to compensate for the environment in which the loudspeaker **5** is placed. Sounds familiar.—NAS

7,013,013

43.38.Vk SURROUND DEVICE

Yoshinori Takei, assignor to Pioneer Electronic Corporation
14 March 2006 (Class 381/307); filed in Japan 20 March 1998

"Dolby Stereo" refers to a two-channel matrix coding scheme in which front-center is represented by an in-phase signal in both channels and the surround effect is represented by an anti-phase relationship. It was



formerly used as the principal carrier of motion picture surround program as well as in VHS duplicated tapes. This patent relates to improvements in the decoding parameters, via equalization, to improve the subjective aspects of localization.—JME

7,024,002

43.38.Vk METHOD AND APPARATUS FOR SPATIALLY ENHANCING THE STEREO IMAGE IN SOUND REPRODUCTION AND REINFORCEMENT SYSTEMS

Baron C. Dickey, Mercer Island, Washington
4 April 2006 (Class 381/77); filed 14 October 2004

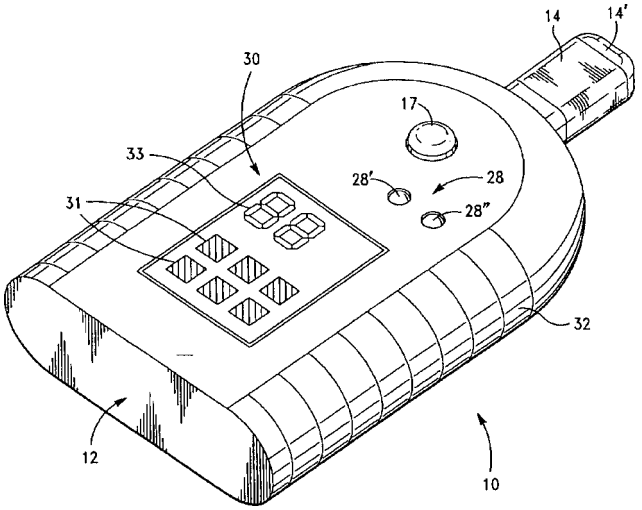
This patent proposes to accomplish something long ago abandoned by practitioners of large-scale concert sound reinforcement—the literal creation over a large listening area of stereo sound coverage as it is normally enjoyed in the home environment. In short, the approach is to divide the listening area into major zones from front to back, with a separate set of loudspeakers covering each zone. Left-right tapering of radiation angles in each zone is presumably relied on to produce accurate left-to-right imaging. How well can such an approach work? In the home environment, cross-firing of loudspeakers can increase the stereo "sweet spot" by subtly trading off delay and amplitude cues [Blauert, *Spatial Hearing* (MIT Press, 1999), p. 209]. However, in making the move from the living room to the arena, relative amplitude cues scale with size, but delay cues do not. Conventional stereo does not work in large spaces, and this is one of the reasons that the motion picture art, as well as concert sound reinforcement, have long made use of a center channel. Regarding fore-aft coverage uniformity, both motion pictures and concert sound have gravitated to asymmetric radiation patterns in order to maintain reasonably consistent response.—JME

7,011,638

43.40.Ng DEVICE AND PROCEDURE TO TREAT CARDIAC ATRIAL ARRHYTHMIAS

Eleanor L. Schuler and Claude K. Lee, assignors to Science Medicus, Incorporated
14 March 2006 (Class 601/46); filed 14 November 2001

The goal of this device and method is to noninvasively control human and animal hearts in the process of treating emergency arrhythmias of the cardiac atrium. The device consists of a body containing a vibration member that creates stimulation at a rate adjustable from nonvibrating mode to a



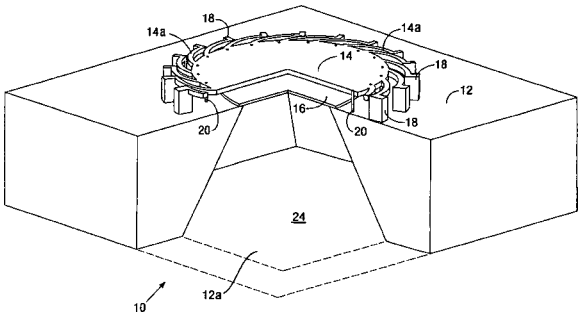
preferred operating range. The vibration member incorporates a vibrating tip, which is used to contact the body. The process consists of positioning the device in the vicinity of the cardoid artery bifurcation and sinus afferent nerve sensors that carry coded signals to the medulla oblongata.—DRR

7,023,066

43.40.Ph SILICON MICROPHONE

Sung Bok Lee and Peter V. Loeppert, assignors to Knowles Electronics, LLC
4 April 2006 (Class 257/415); filed 20 November 2001

MEMS (micro-electromechanical-systems) technology is used here to produce a ''silicon'' microphone. The figure is self-evident, showing the



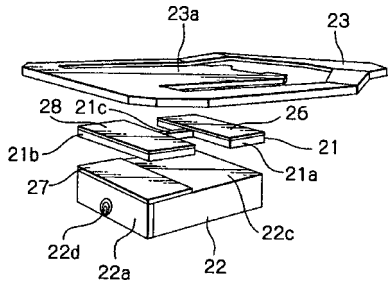
exposed diaphragm **14** and the set of compliant tangential arms **14a** that support the diaphragm and relieve stresses in it.—JME

7,016,155

43.40.Tm VIBRATION-CANCELING MECHANISM AND HEAD GIMBAL ASSEMBLY WITH THE VIBRATION-CANCELING MECHANISM

Tamon Kasajima and Masashi Shiraishi, assignors to SAE Magnetics (H.K.) Limited
21 March 2006 (Class 360/234.6); filed in Japan 4 July 2001

The author claims tuned dynamic damping of a gyro-stabilized disk read head **22d**. The damper **28-21-26-23** consists of flat material for ease of precise manufacturing. Lateral vibration of the gimbaled arm supporting beam **23** will cause torsion bending of beam **21c**, resulting in tilting of mass



22 that maintains the position of read head **22d** on its track. The resonant frequency of mass **22** on spring **21c** is tuned by micromachining of **21c** to be the same as the resonant frequency of the arm (not shown) supporting **23**. Material **28** is soft and also provides damping.—AJC

7,013,206

43.40.Vn ELECTRICALLY ADJUSTABLE SEMIACTIVE DAMPER CONTROL

Alexander Stiller *et al.*, assignors to Continental Aktiengesellschaft
14 March 2006 (Class 701/37); filed in Germany 27 April 2001

A shock absorber for automotive vehicles according to this patent is actuated via a control system that makes use of wheel- and body-motion signals. The range of the absorber force is adjusted to take road and driving conditions into account, making use of signals that may represent such parameters as the vehicle's longitudinal deceleration, changes in the brake pressure, engine torque, or accelerator pedal position.—EEU

7,013,832

43.40.Vn ARRANGEMENT FOR DAMPING OF STRUCTURAL RESONANCE

David C. Sexton *et al.*, assignors to Bae Systems plc
21 March 2006 (Class 114/382); filed in the United Kingdom 28 April 2000

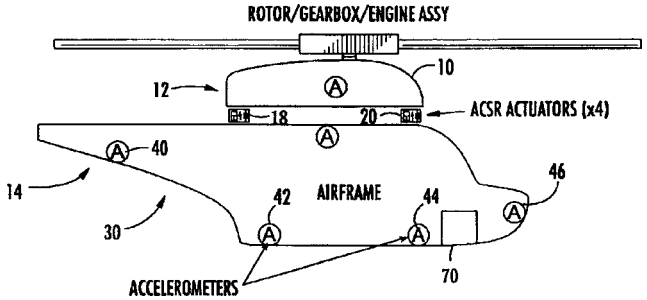
In order to suppress the vibrations induced in the hull of a marine vessel by its propulsion system, signals from a number of accelerometers that are mounted on the hull are fed to a detector/controller that determines the onset of resonant excitation of particular modes. Signals produced by this controller are employed to actuate inertial masses (essentially, electrodynamic shakers) attached to the vessel's thrust block or other suitable component so as to generate forces that attenuate the resonant mode.—EEU

7,017,857

43.40.Vn ACTIVE VIBRATION CONTROL SYSTEM

Wayne Hill and Lev S. Tsimring, assignors to Foster-Miller, Incorporated
28 March 2006 (Class 244/17.13); filed 16 September 2002

A 20-Hz, helicopter-blade, passage-frequency, vibration-reduction system is claimed where an accelerometer **10** provides a proxy signal that is conditioned by a computer to drive actuators **18** and **20** located between the



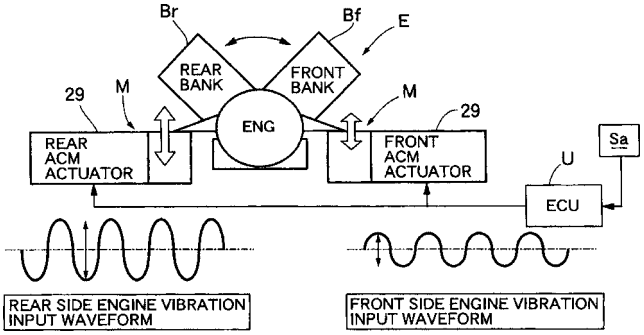
airframe and the rotor/gearbox/engine assembly. Conditioning uses the state of vibration to predict amplitude to improve cancellation accuracy. A 48% reduction in vibration amplitude was observed.—AJC

7,017,889

43.40.Vn ACTUATOR DRIVE CONTROL DEVICE FOR ACTIVE VIBRATION ISOLATION SUPPORT SYSTEM, AND METHODS OF USING SAME

Atsushi Abe, assignor to Honda Motor Company, Limited
28 March 2006 (Class 267/140.15); filed in Japan 13 June 2003

An engine firing-frequency vibration-reduction system is claimed where crank-pulse sensor **Sa** provides a reference-time proxy signal that is



conditioned by a computer to valves in actuators **29** placed between the frame and the engine. Differences in firing state between the front and rear

cylinder banks, which are computed based on previous engine-cylinder firing history, are used to compute the time when the actuator valves should be operated.—AJC

7,015,812

43.50.Hg SOUND PROTECTING HEADSET WITH PROXIMITY COLLISION AVOIDANCE PROTECTION

Marc S. Lemchen, New York, New York
21 March 2006 (Class 340/541); filed 4 October 1995

Earmuff-type hearing protectors incorporate an audible proximity detector to warn of otherwise inaudible moving objects.—JE

7,024,010

43.50.Hg ELECTRONIC EARPLUG FOR MONITORING AND REDUCING WIDEBAND NOISE AT THE TYMPANIC MEMBRANE

William R. Saunders *et al.*, assignors to Adaptive Technologies, Incorporated
4 April 2006 (Class 381/317); filed 19 May 2003

An active noise-reduction earplug is described that provides for determining the complex sound pressure at the tympanic membrane and at the active transducer. This permits location of the microphone and the transducer at various locations within the earplug. Effective noise reduction up to 5 kHz is achieved.—JE

7,024,013

43.50.Hg SOUND REDUCTION/ELIMINATION DEVICE

Gregory A. Van Dam, Delray Beach and James Padula, Coconut Creek, both of Florida
4 April 2006 (Class 381/376); filed 12 January 2004

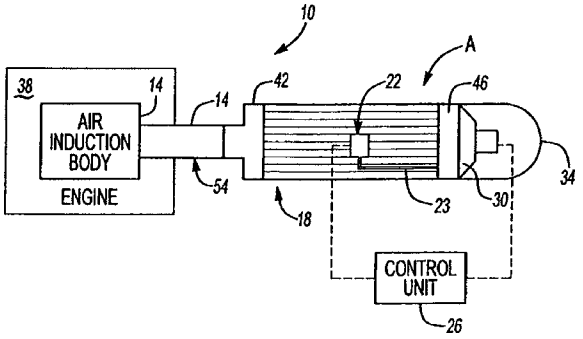
This is an earmuff with a cloth covering designed to be worn by the sleeping partner of a snoring person. The device may ''improve longevity of relationship.''—JE

7,016,506

43.50.Ki MODULAR ACTIVE NOISE AIR FILTER SPEAKER AND MICROPHONE ASSEMBLY

Brian Chiara, assignor to Siemens VDO Automotive Incorporated
21 March 2006 (Class 381/71.4); filed 25 September 2002

Automotive engines emit substantial noise from the air intake as well as the exhaust. The intake would seem to be an attractive candidate for active noise suppression because the system components do not have to withstand corrosive gases at high temperatures. Even so, the sensing



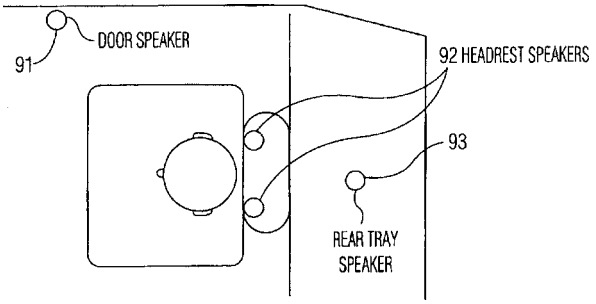
microphone may be exposed to debris and foul weather. This very short patent describes a method of burying the microphone inside the intake air filter. In a preferred embodiment, the microphone, speaker, and air filter are combined into a single unit that can be easily installed.—GLA

7,020,288

43.50.Ki NOISE REDUCTION APPARATUS

Toshihiko Ohashi, assignor to Matsushita Electric Industrial Company, Limited
28 March 2006 (Class 381/71.4); filed in Japan 20 August 1999

This is an active noise cancellation (ANC) system for the interior of moving vehicles in which multiple interior microphones, engine-speed-, and running-speed-sensor signals provide information to an ANC unit containing signal processing and amplification to feed interior speakers **91** and **93** and



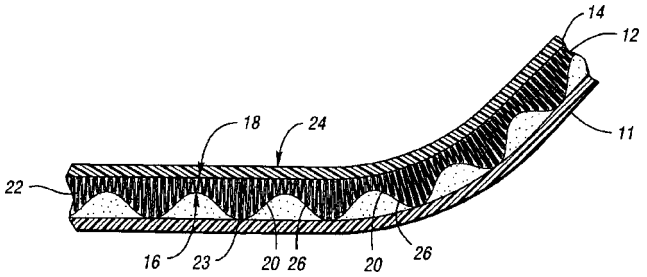
passenger headrest speakers **92**. Bone-conduction actuators, not shown, placed near **93**, are an alternate. The ANC unit includes adaptive-filter signal processing for voice-sound-frequency recognition and loudness enhancement according to individual passenger seat locations.—AJC

7,011,181

43.55.Ev SOUND INSULATION SYSTEM

Donald C. Albin, Jr., assignor to Lear Corporation
14 March 2006 (Class 181/290); filed 8 July 2003

This is a vehicle floor-pan acoustical absorber comprising pleated material **22** with fibers oriented vertically and corrugated surface **12–16**



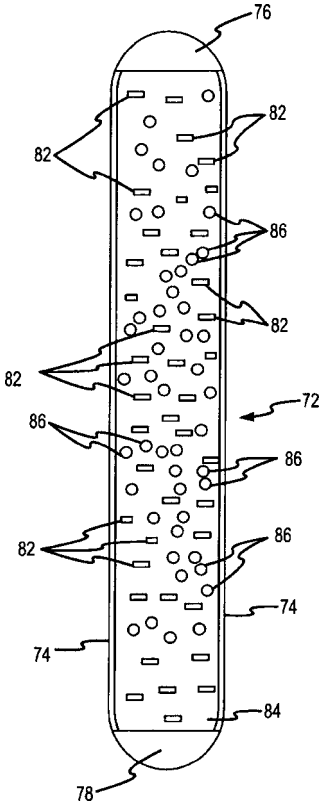
forming a remaining air gap. Pleating **22** is formed by a Strutto machine. Claims include mold method, shape **24**, heating, and sequence. Some improved absorption above 1 kHz is evident from the author's test data.—AJC

6,991,549

43.58.Wc SOUND PRODUCING PLAY APPARATUS

Grant Ballin, assignor to Wonderworks LLC
31 January 2006 (Class 472/106); filed 24 July 2003

When tilted, a flat platform with pegs **82** will induce balls **86** to roll and hit the pegs, thereby producing sound. Of course the pegs can be of any



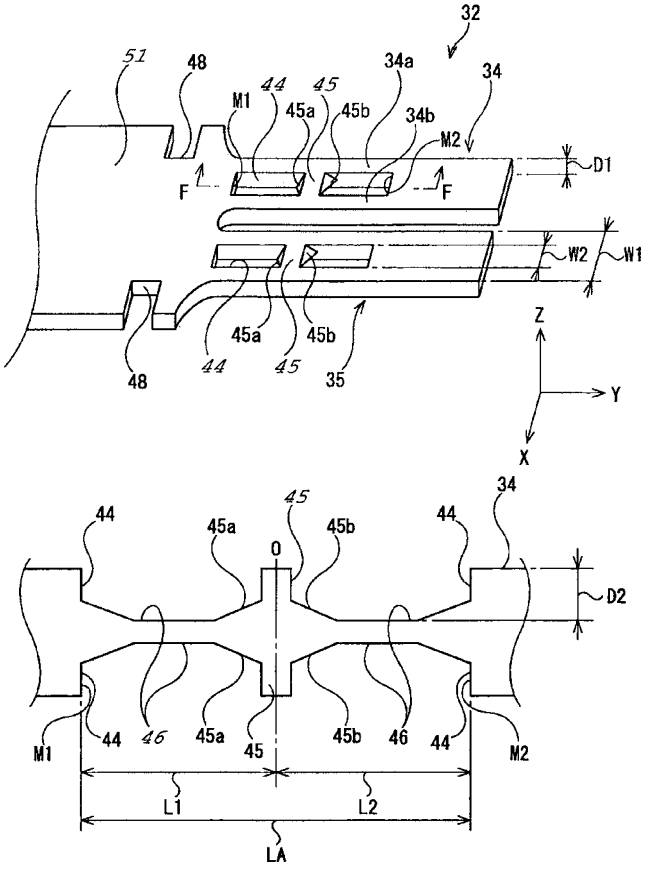
shape and could be tuned to any scale. Think of a pinball machine with more than one ball and tuned bumpers.—MK

7,015,630

43.58.Wc PIEZOELECTRIC VIBRATION PIECE, PIEZOELECTRIC DEVICE USING PIEZOELECTRIC VIBRATION PIECE, PORTABLE PHONE UNIT USING PIEZOELECTRIC DEVICE, AND ELECTRONIC EQUIPMENT USING PIEZOELECTRIC DEVICE

Hideo Tanaya, assignor to Seiko Epson Corporation
21 March 2006 (Class 310/370); filed in Japan 17 December 2002

This patent seems oddly focused on high-stability piezoelectric resonators of the tuning-fork-type for use as buzzers. The device pictured is said to be useful for all sorts of sound and vibration actuators, but nothing is mentioned pertaining to use in a clock. The tuning-fork bars are recessed in a peculiar way that is supposed to make them less temperature sensitive; the inventors argue that this is due to the bar shape. One might conjecture that



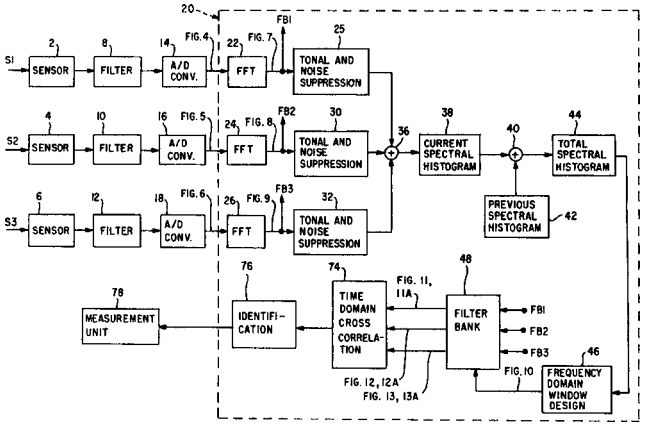
this is due to reduced coupling to out-of-plane vibrations, as the authors did, but this is not obvious from the design.—JAH

7,012,854

43.60.Dh METHOD FOR DETECTING EMITTED ACOUSTIC SIGNALS INCLUDING SIGNAL TO NOISE RATIO ENHANCEMENT

Pei-hwa Lo, assignor to Honeywell International Incorporated
14 March 2006 (Class 367/135); filed 21 June 1990

This method detects emitted acoustic signals and improves signal-to-noise ratios, wherein the emitted signals are distinguished from self-noise transient signals. Because the emitted signals are unsteady, the process uses multiple sensors and an adaptive-filter technique having the capability to



track the emitted signal and to enhance the signal-to-noise ratio. By monitoring noise statistics of the sensor signals, the digital adaptive-tuning filter bank automatically adjusts its upper threshold (to eliminate strong tonals)

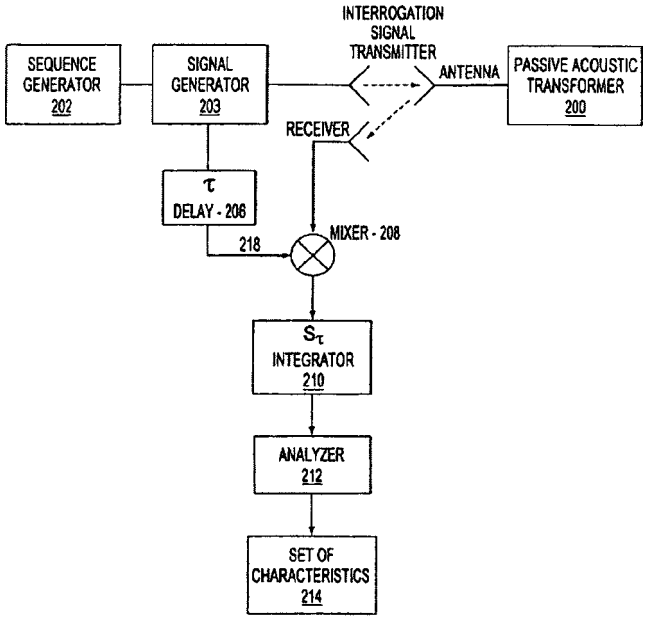
and its lower threshold (to eliminate background noise) in order to obtain a discovered frequency band.—DRR

7,023,323

43.60.Dh FREQUENCY HOPPING SPREAD SPECTRUM PASSIVE ACOUSTIC WAVE IDENTIFICATION DEVICE

Paul A. Nysen, assignor to X-Cyte, Incorporated
4 April 2006 (Class 340/10.1); filed 14 October 2003

This is a transponding identification system relying on acoustic delays encoded in SAW tag **200** (''passive acoustic transformer'') affixed to another object which may be in motion. A combination (**202–214**) of matching encoding and decoding SAW, delay line, and acoustical and frequency-hopping techniques, where the ''chirp'' method is replaced with a discontinuous serial-hopping sequence (66 kHz rate), provide encoding and decoding of tag information with up to 16 degrees of freedom. Modulated rf



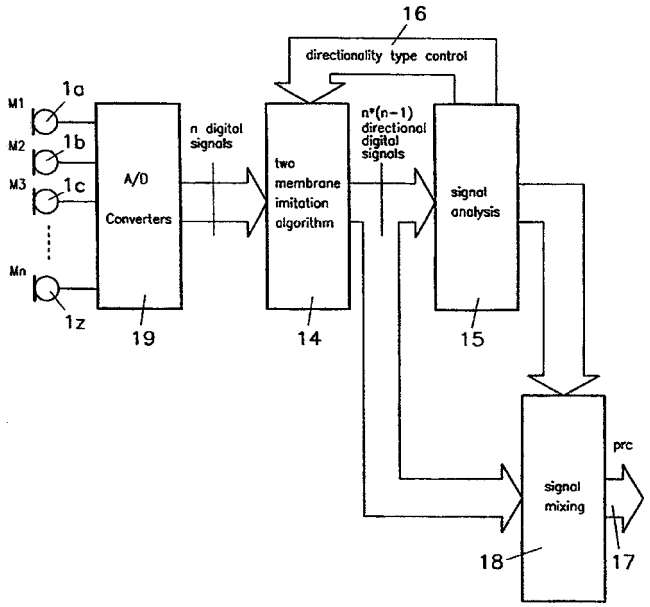
signals are transmitted to, and received from, tag **200** by microwave rf carriers. The tagged object may be in motion. The interrogating-receiver homodyne demodulator **206-218-208-210** produces a base-band signal that may need Doppler compensation by a DSP in analyzer **214**. The frequency range from demodulator **210** is below 3000 Hz and is amenable to long distance transmission from mixer **208** to **210–212** over conventional twisted-pair wires. Characteristics stored in **214** provide decoding of the information received from tag **200**.—AJC

7,020,290

43.60.Fg METHOD AND APPARATUS FOR PICKING UP SOUND

Zlatan Ribic, Vienna, Austria
28 March 2006 (Class 381/92); filed in the European Patent Office 7 October 1999

The patent deals with the adaptive synthesis of directional pickup patterns in hearing aid applications to improve reception by the user. Higher-



order frontal directivity can be attained if an in-line microphone array can be accommodated.—JME

7,020,291

43.60.Fg NOISE REDUCTION METHOD WITH SELF-CONTROLLING INTERFERENCE FREQUENCY

Markus Buck *et al.*, assignors to Harman Becker Automotive Systems GmbH
28 March 2006 (Class 381/92); filed in Germany 14 April 2001

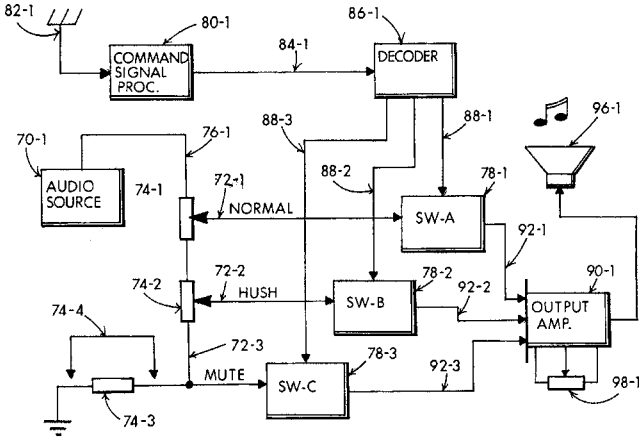
Quoting from the abstract, the patent ''relates to a method with which speech is captured in a noisy environment with as high a speech quality as possible. To this end, a compact array of, for example, two single microphones is combined to form one system through signal processing methods consisting of adaptive beam formation and spectral subtraction. Through the combination with a spectral subtraction, the reference signal of the beam former is freed from speech signal components to the extent that a reference signal of the interference is formed and the beam former produces high gains.''—JME

7,012,652

43.60.Qv AUDIO HUSH FOR ENTERTAINMENT EQUIPMENT AND PARTICULARLY TELEVISION RECEIVERS

Harold J. Weber, assignor to Sam Stuff Property Trust
14 March 2006 (Class 348/632); filed 9 January 2003

A remote control is designed to provide television-sound quieting as a practical alternative to the usual mute control. This device provides a method of quickly reducing the audio volume of a television or radio program from an overbearing sound level to a more tolerable level and then returning the volume to the original level that was set prior to the reduction.



As far as this reviewer is concerned, he'd just as soon zip past the commercials on his Tivo.—DRR

7,002,069

43.66.Lj BALANCING MID INSTRUMENT VOLUME LEVELS

Adarsh S. Desai *et al.*, assignors to Motorola, Incorporated
21 February 2006 (Class 84/645); filed 9 March 2004

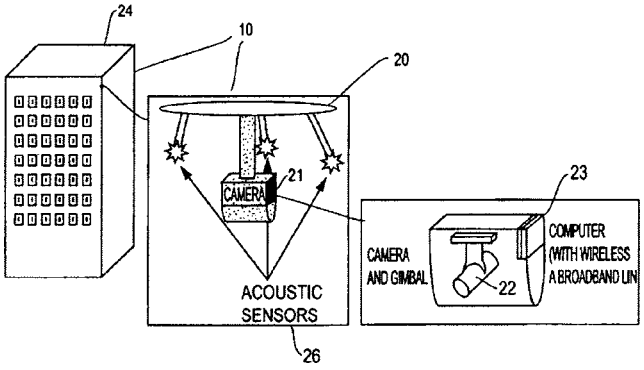
Consider the mobile telephone: when playing a MIDI file on the phone, the levels will vary from file to file. So, use a perceptual mapping of the MIDI notes and the tiny loudspeaker on the phone. This perceptual map adjusts the gain for each MIDI channel.—MK

6,965,541

43.66.Qp GUN SHOT DIGITAL IMAGING SYSTEM

Brett D. Lapin and Nicholas D. Beser, assignors to The Johns Hopkins University
15 November 2005 (Class 367/118); filed 7 November 2003

Several recent patents have covered methods for detecting gunshots and locating the source of the sound, usually in an open, outdoor setting. This system would not only do that, and produce recordings of the detected



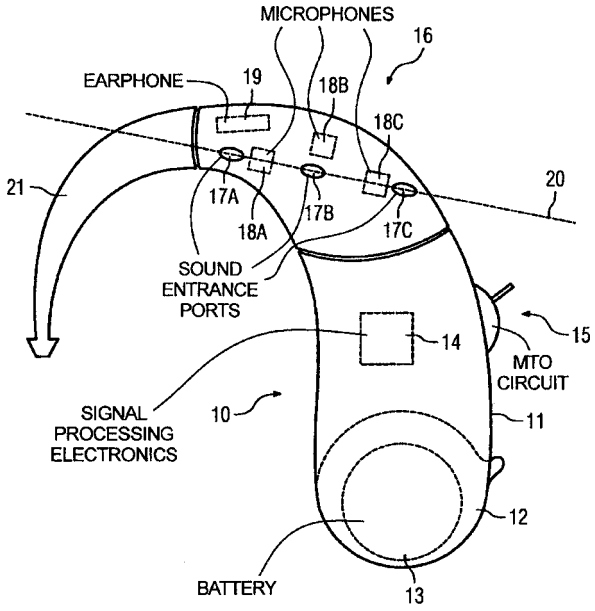
sound, but would also aim cameras at the computed source locations in an attempt to produce photographs of the shooter. The patent text, as well as the claims, are more concerned with the optical aspects of camera control than with the acoustical details of the event detection.—DLR

7,013,014

43.66.Ts HEARING DEVICE SYSTEM WITH BEHIND-THE-EAR HEARING AID DEVICES FASHIONED SIDE-SPECIFIC

Gerhard Ach-Kowalewski and Hartmut Ritter, assignors to Siemens Audiologische Technik GmbH
14 March 2006 (Class 381/313); filed in Germany 20 December 2002

In order to improve the directivity of behind-the-ear hearing aids utilizing directional microphones, at least three openings for microphone inlets



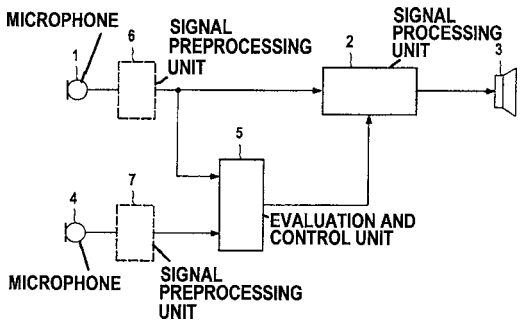
are placed on either the left or the right side of the hearing aid housing. The hearing aids are positioned on each ear so that the microphone openings will always face outward from the wearer's head.—DAP

7,013,015

43.66.Ts METHOD FOR THE OPERATION OF A HEARING AID DEVICE OR HEARING DEVICE SYSTEM AS WELL AS HEARING AID DEVICE OR HEARING DEVICE SYSTEM

Volker Hohmann *et al.*, assignors to Siemens Audiologische Technik GmbH
14 March 2006 (Class 381/318); filed in Germany 2 March 2001

To reduce acoustic-feedback oscillation in hearing aid fittings, signals are compared from two microphones spaced apart, for example, on each ear



of a binaural fitting. If oscillations are present in only one of the microphone signals, acoustic feedback is assumed and gain is reduced at that frequency or correlated parts of the signal are removed.—DAP

7,014,657

43.66.Ts MIDDLE-EAR IMPLANT

Saumil N. Merchant and Joseph B. Nadol, Jr., assignors to Massachusetts Eye & Ear Infirmary
21 March 2006 (Class 623/10); filed 26 July 2000

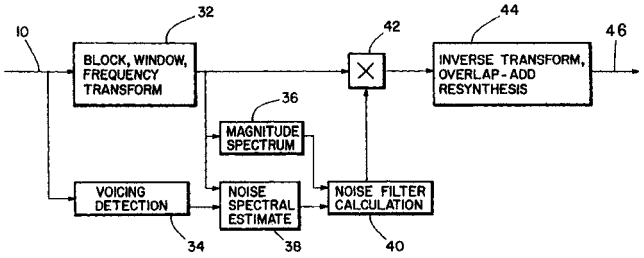
Diseases such as otitis media can produce fluid or fibrous tissue accumulation in the middle-ear chamber that might interfere with sound propagation to the inner ear. A gas-filled balloon is surgically placed in the middle ear in partial contact with the eardrum to displace the fluid and tissue in order to allow near-normal functionality of the eardrum, the ossicles in the middle ear, and the round window of the inner ear.—DAP

7,016,507

43.66.Ts METHOD AND APPARATUS FOR NOISE REDUCTION PARTICULARLY IN HEARING AIDS

Robert Brennan, assignor to AMI Semiconductor Incorporated
21 March 2006 (Class 381/71.6); filed 16 April 1998

In the presence of speech, the magnitude spectrum of the input signal is compared to the noise-magnitude spectral estimate that has been



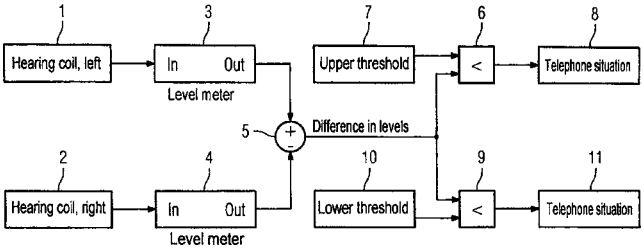
computed in the absence of speech. An attenuation function modifies the input signal depending on signal-to-noise ratio: no input-signal modification is made for very high and very low signal-to-noise ratios.—DAP

7,016,510

43.66.Ts HEARING AID AND OPERATING METHOD FOR AUTOMATICALLY SWITCHING TO A TELEPHONE MODE

Georg-Erwin Arndt *et al.*, assignors to Siemens Audiologische Technik GmbH
21 March 2006 (Class 381/312); filed in Germany 10 October 2003

The object is to detect that the wearer of a binaural hearing aid fitting is making a telephone call and to automatically switch from microphone mode to induction-coil mode. The method uses the difference detected in the



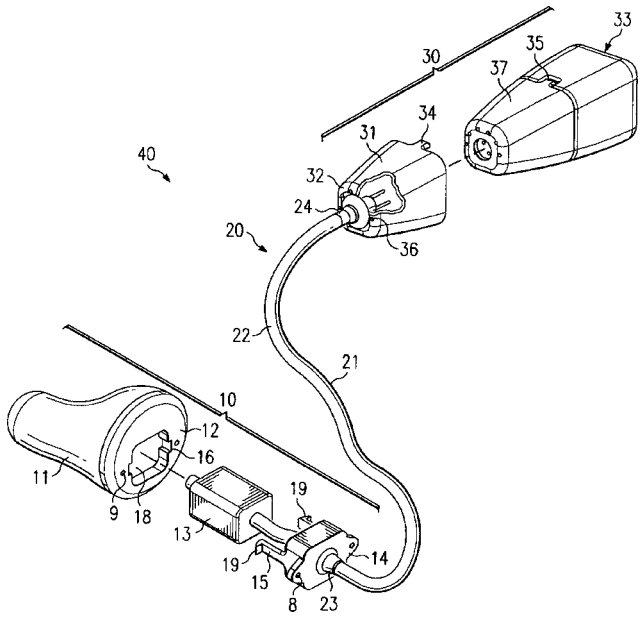
induction-coil outputs of the left and right hearing aids caused by the magnetic input signal from the telephone, which may include interference signals.—DAP

7,016,512

43.66.Ts BTE/CIC AUDITORY DEVICE AND MODULAR CONNECTOR SYSTEM THEREFOR

Jim Feeley and Mike Feeley, assignors to Hear-Wear Technologies, LLC
21 March 2006 (Class 381/324); filed 29 August 2003

An elastic-deformable cushion tip forms the acoustic termination of a combination over-the-ear/in-the canal hearing aid. The speaker module, which is located in the in-the-canal portion of the device, is removably



connected via wires to an over-the-ear assembly in which audio processing is performed. The speaker module also has a raised ridge to prevent its removal from the cushion tip.—DAP

7,017,276

43.66.Ts HEARING AID DRYER

Bernard Jay Greenspan, San Diego, California *et al.*
28 March 2006 (Class 34/80); filed 3 June 2004

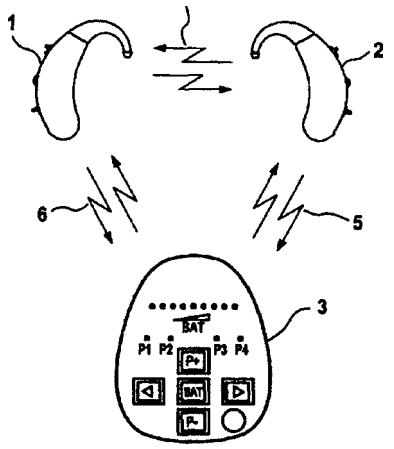
The components and interconnections within hearing aids are sensitive to moisture ingress, which is sometimes alleviated by placing the hearing aids into dryers to remove the moisture. In time, the dryer itself becomes saturated with moisture, which is typically remedied by removal of the desiccant and application of high temperatures. To more safely and conveniently regenerate the desiccant material, a dryer is constructed with a separate compartment for the desiccant with perforated cover and a heating cycle for desiccant regeneration.—DAP

7,020,296

43.66.Ts METHOD FOR OPERATING A HEARING AID SYSTEM AND HEARING AID SYSTEM

Torsten Niederdränk, assignor to Siemens Audiologische Technik GmbH
28 March 2006 (Class 381/315); filed in Germany 29 September 2000

Characteristics of the acoustic fields are detected and analyzed by each hearing aid in a binaural fitting and shared via bidirectional wireless communication across the wearer's head or to/from an external remote device. The signal processing of each hearing aid is adapted based on parameters



such as frequency spectra, spatial characteristics, modulation frequency and depth of the desired signal, and noise signals incident at each hearing aid.—DAP

7,020,297

43.66.Ts SUBBAND ACOUSTIC FEEDBACK CANCELLATION IN HEARING AIDS

Xiaoling Fang *et al.*, assignors to Sonic Innovations, Incorporated
28 March 2006 (Class 381/317); filed 15 December 2003

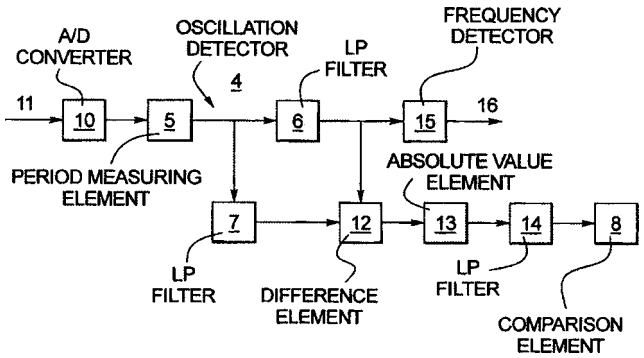
An audio signal is divided by a filterbank into several digital subband signals. Two cascaded narrow-band filters are combined with a fixed delay to represent the feedback path in each subband. The first filter is used to model the static portion of the feedback path in each subband. The second filter is typically a FIR type and is used for tracking variations in the feedback path in each subband.—DAP

7,024,011

43.66.Ts HEARING AID WITH AN OSCILLATION DETECTOR, AND METHOD FOR DETECTING FEEDBACK IN A HEARING AID

Volkmar Hamacher *et al.*, assignors to Siemens Audiologische Technik GmbH
4 April 2006 (Class 381/320); filed in Germany 12 May 1999

If a sinusoidal signal exists over a number of sample periods, there is a possibility that the hearing aid is oscillating. A long-term average value is compared to a short-term average value of the digitized input signal. If the



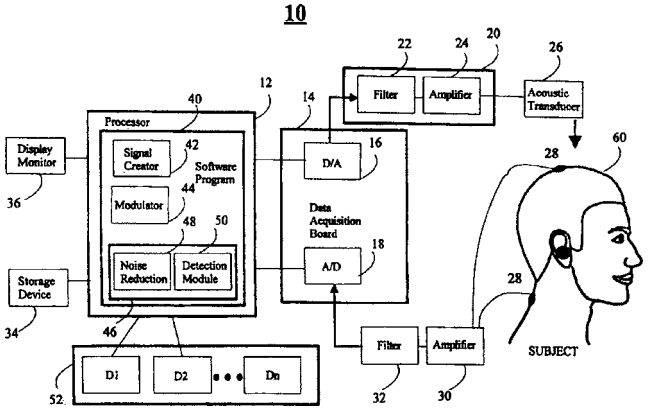
two values are essentially the same, it is assumed that oscillation is occurring, and a filter may be adapted to attenuate energy at that frequency.—DAP

7,014,613

43.66.Yw SYSTEM AND METHOD FOR OBJECTIVE EVALUATION OF HEARING USING AUDITORY STEADY-STATE RESPONSES

Michael S. John and Terence W. Picton, both of Toronto, Ontario, Canada
21 March 2006 (Class 600/559); filed 29 April 2004

This patent describes an apparatus and method for evaluating a subject's hearing by recording steady-state auditory evoked responses. The equipment generates a steady-state auditory-response-potential stimulus



containing multiple stimulus components, relays the stimulus to the subject, senses response potentials while simultaneously presenting the stimulus, and determines whether the sensed potentials contain multiple responses to the stimulus.—DRR

7,016,504

43.66.Yw PERSONAL HEARING EVALUATOR

Adnan Shennib, assignor to Insonus Medical, Incorporated
21 March 2006 (Class 381/60); filed 21 September 1999

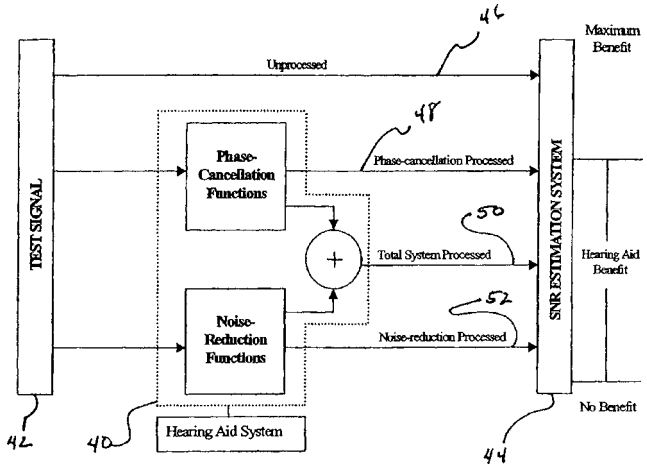
This is a portable, handheld, hearing evaluator consisting of an audio transducer (speaker) and auxiliary circuitry for delivering acoustic test stimuli to a test subject within the direct sound field of the device. The device delivers multilevel and multifrequency test signals for a subjective response by the test subject holding the device. An ultrasonic position sensor determines the position of the device with respect to the subject's ear or a portion of interest of the head of the test subject. The acoustic test stimuli are controlled and regulated according to the position of the device with respect to the test subject so that appropriate levels of test stimuli are presented only when the device is within a proper range and regardless of its exact position on the test subject's head. In one embodiment, the handheld device can be connected to an auxiliary instrument, e.g., a computer or a microprocessor-based audiometer for remote control of the device and for registering audible responses via a response switch on the device.—DRR

7,020,581

43.66.Yw MEDICAL HEARING AID ANALYSIS SYSTEM

Andrew B. Dittberner, assignor to Medacoustics Research & Technology
28 March 2006 (Class 702/183); filed 20 October 2003

This is a hearing-aid analysis system for objectively evaluating the effectiveness of advanced hearing aid technologies by comparing the results of computer speech recognition obtained from enhanced and unenhanced speech. The system first presents an original unprocessed speech signal to the speech recognition software as a control measure. A speech signal is then processed through the hearing aid and through hearing-loss filtering to simu-



late as closely as possible the effect of the hearing aid as worn by the patient and the system performs recognition on this speech signal. The software then compares the recognition rates of the two different signals. On this basis, the system creates an objective indication of the benefit to be gained from the hearing aid under test, relative to the control measure.—DRR

6,963,841

43.70.Dn SPEECH TRAINING METHOD WITH ALTERNATIVE PROPER PRONUNCIATION DATABASE

Anthony H. Handal *et al.*, assignors to Lessac Technology, Incorporated
8 November 2005 (Class 704/270); filed 9 January 2003

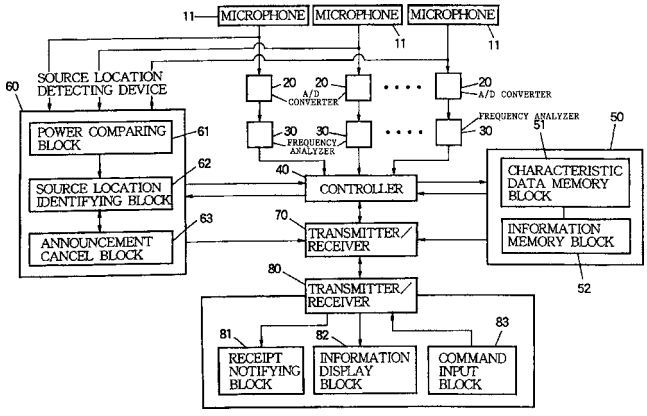
This patent is a riot. It begins with a couple of pages of pathetically naive discussion of phonetics and speech recognition technology. A computer system is then described which would analyze your spoken inputs and then teach you how to speak correctly. This would be done using something called the ''Lessac System,'' which seems to be some sort of theory of the phonetic structure of speech. The system, as provided, would include various dictionaries of correct pronunciations, such as general American English and southern American dialect. The system would allegedly be able to alert the user to substandard pronunciations, such as ''nuculer'' for ''nuclear.''—DLR

7,020,610

43.72.Fx APPARATUS FOR RESPONDING TO SOUND AT A REMOTE TERMINAL

Hiromi Okitsu, assignor to Yamaha Corporation
28 March 2006 (Class 704/270.1); filed in Japan 3 August 1998

This omnibus patent deals with adaptive analysis and identification of speech and other signals in terms of source location and spectral/temporal characteristics. Applications range from remote sensing in security systems to providing information for the hearing impaired.—JME



7,016,746

43.72.Gy DIGITAL AUDIO SIGNAL FILTERING MECHANISM AND METHOD

Philip R. Wiser *et al.*, assignors to Microsoft Corporation
21 March 2006 (Class 700/94); filed 16 May 2002

The audio processing profiles for a particular audio source signal are organized and stored according to the program content and the specific bandwidths to be delivered, e.g., 14.4 kbps or 28.8 kbps. From these categorizations, sound engineers can compare and select from different combinations of preprocessing, filtering, and encoding that optimize sound quality for the bandwidth required.—DAP

6,965,862

43.72.Ja READING MACHINE

Carroll King Schuller, Deptford, New Jersey
15 November 2005 (Class 704/258); filed 11 April 2002

The author apparently never heard of Kurzweil's patents on this technology, issued in the 1980s. There are no references here to that work, but neither does there seem to be anything new here that was not then covered. The speech synthesis subsystem is represented in a couple of the claims simply as the ''sound-producing element.''—DLR

6,963,839

43.72.Ne SYSTEM AND METHOD OF CONTROLLING SOUND IN A MULTI-MEDIA COMMUNICATION APPLICATION

Joern Ostermann *et al.*, assignors to AT&T Corporation
8 November 2005 (Class 704/260); filed 2 November 2001

This idea seems fairly obvious, but that doesn't mean the patent office would cast it out. When a text email message is to be synthesized and delivered in the form of speech, the synthesis strategy would include scanning for text emoticons, the little character patterns which typically denote moods, such as surprise, irritation, etc. The synthesized voice output would be modified by adjusting vocal intensity or pauses. There is no mention here of adjusting pitch, although several other patents, mentioned and incorporated by reference, may include that nicety.—DLR

6,963,840

43.72.Ne METHOD FOR INCORPORATING MULTIPLE CURSORS IN A SPEECH RECOGNITION SYSTEM

Charles Sumner, assignor to International Business Machines Corporation
8 November 2005 (Class 704/270); filed 12 January 2001

This patent deals with techniques for managing the edit cursor during the operation of a speech-recognition dictation system. The central idea seems to be the use of multiple cursors to minimize the confusions inherent in trying to move the cursor from one edit position to another. A typical situation might be the switch from entry at the end of the body of text to a temporary location where editing is to be done on previously entered text. After the editing, the most natural operation, it is argued here, would be to simply resume the entry of new text without having to reposition the cursor to the end.—DLR

6,965,863

43.72.Ne SPEECH RECOGNITION USER INTERFACE

Sarah E. Zuberec *et al.*, assignors to Microsoft Corporation
15 November 2005 (Class 704/270); filed 12 November 1998

This patent deals with a protocol for notifying the user when the speech recognition function is operating and when it has been shut down to save power or to switch to an alternate mode. In this case, the application is an automobile navigation system. A row of dots on a display would notify the passage of the active time interval. A pattern of distinctive, and perhaps programmable, sounds would indicate the changes of state, which could involve switching vocabularies, or other such system changes, as well as power control.—DLR

6,965,812

43.72.Ne SPEECH INTERFACE FOR AN AUTOMATED ENDOSCOPIC SYSTEM

Yulun Wang and Darrin Uecker, assignors to Computer Motion, Incorporated
15 November 2005 (Class 700/258); filed 11 March 2002

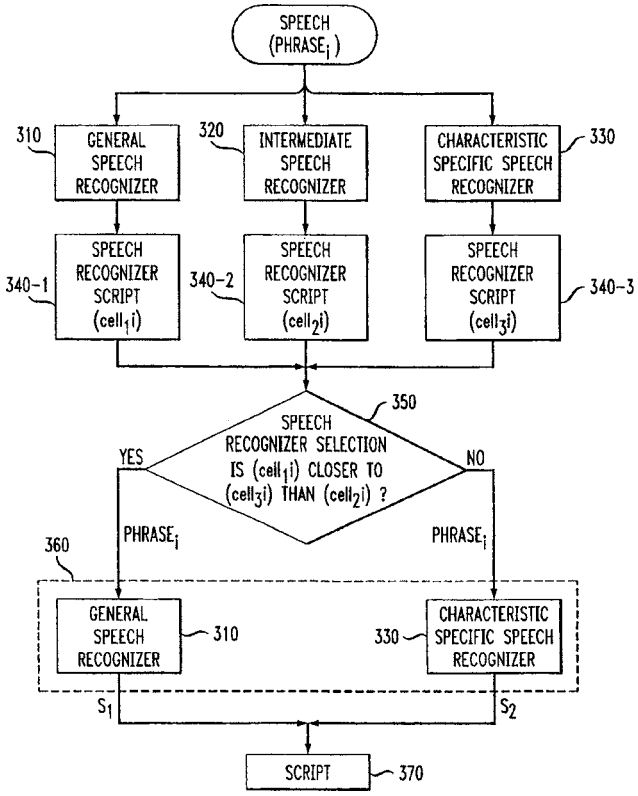
United States patent law allows the granting of a patent for any trivially novel application of well-known technology. It's not bad enough that a patent may be granted for the idea of using well-known speech-recognition technology in a generally new area, such as system control, but that each and every such system may be granted its own patent. The requirement is merely that the new combination not be obvious. Here, the application is a medical endoscope. While watching the progress on a monitor, the doctor would speak commands, such as "move," "step," "left," "right," "slower," "faster," and so on. A standard, commercial speech recognizer would do the analysis.—DLR

7,016,835

43.72.Ne SPEECH AND SIGNAL DIGITIZATION BY USING RECOGNITION METRICS TO SELECT FROM MULTIPLE TECHNIQUES

Ellen Marie Eide *et al.*, assignors to International Business Machines Corporation
21 March 2006 (Class 704/231); filed 19 December 2002

If the input information has certain physical parameters that affect recognition accuracy, such as fast speech, speech with background noise or music, or accented speech, a characteristic-specific recognizer is used that



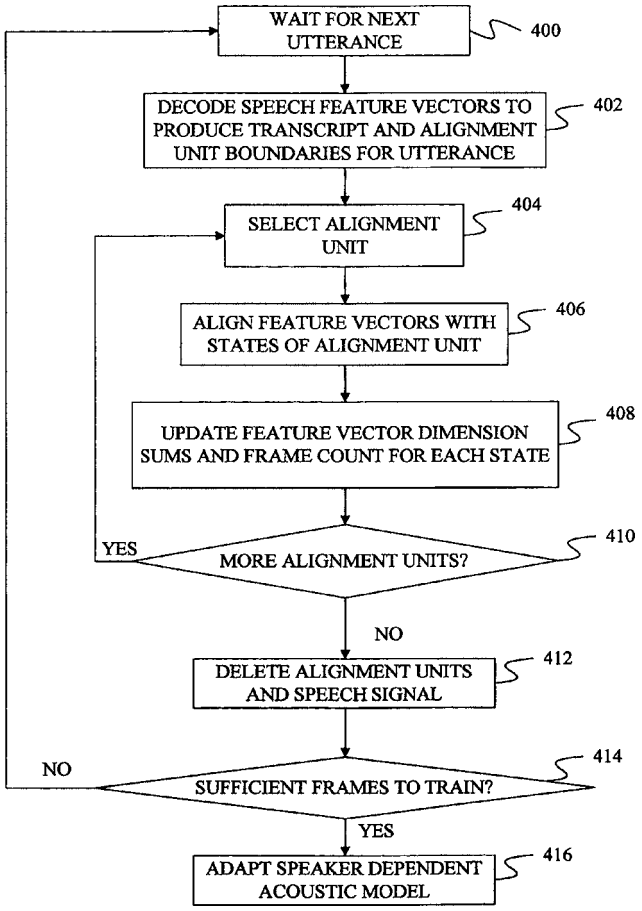
demonstrates improved performance for the specific physical parameter or subsets of the input information. Otherwise, a general recognizer is utilized.—DAP

7,016,838

43.72.Ne METHOD AND SYSTEM FOR FRAME ALIGNMENT AND UNSUPERVISED ADAPTATION OF ACOUSTIC MODELS

William H. Rockenbeck *et al.*, assignors to Microsoft Corporation
21 March 2006 (Class 704/244); filed 12 November 2004

Alignment unit boundaries are formulated for individual utterances to reduce processing time and the amount of data storage reserved for decoding the transcript in unsupervised adaptation. Alignment time is reduced for



aligning individual frames of speech with particular acoustic units in the transcript by using alignment-unit boundaries with a reduced set of feature vectors for each utterance.—DAP

7,019,202

43.75.Gh ACOUSTIC GUITAR WITH REVERBERATING BRIDGE ASSEMBLY

Robert Hetzel, Staten Island, New York
28 March 2006 (Class 84/299); filed 14 June 2004

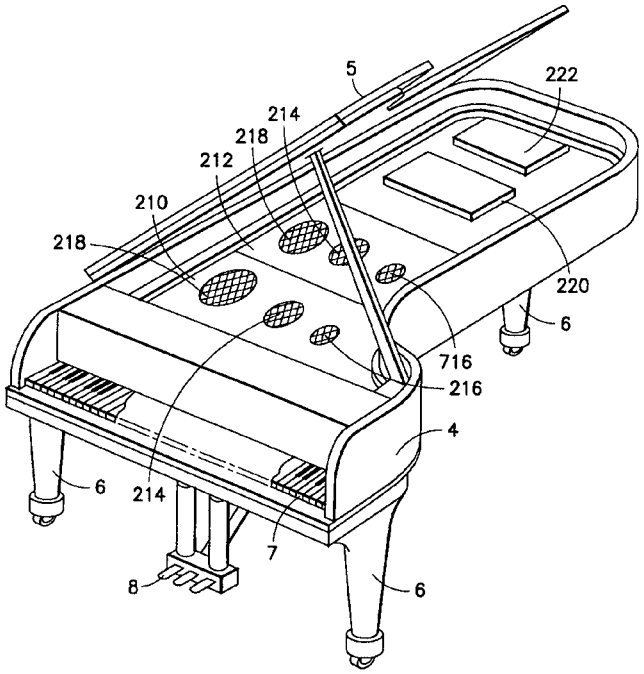
The inventor believes that inserting a hollow metal tube underneath the strings at the bridge will produce ''reverberation.'' A more likely explanation for sound effects produced by the tube would include induction of vibration in other strings as well as the sliding of the string on the tube.—MK

7,002,070

43.75.Mn ELECTRONIC PIANO

Shelley Katz, East Sussex, the United Kingdom
21 February 2006 (Class 84/718); filed 24 June 2002

The inventor proposes adding distributed mode loudspeakers **220** and **222** (see the numerous NXT patents for more details) to an electric piano as



shown in the figure. These loudspeakers provide the more reverberant sound of the soundboard. Other cone loudspeakers are also used to provide the string sounds (**218**, **214**, **216**).—MK

7,005,571

43.75.Wx MIDI CONTROLLER PEDALBOARD

Warren R. Groff, Los Osos, California
28 February 2006 (Class 84/645); filed 11 October 2002

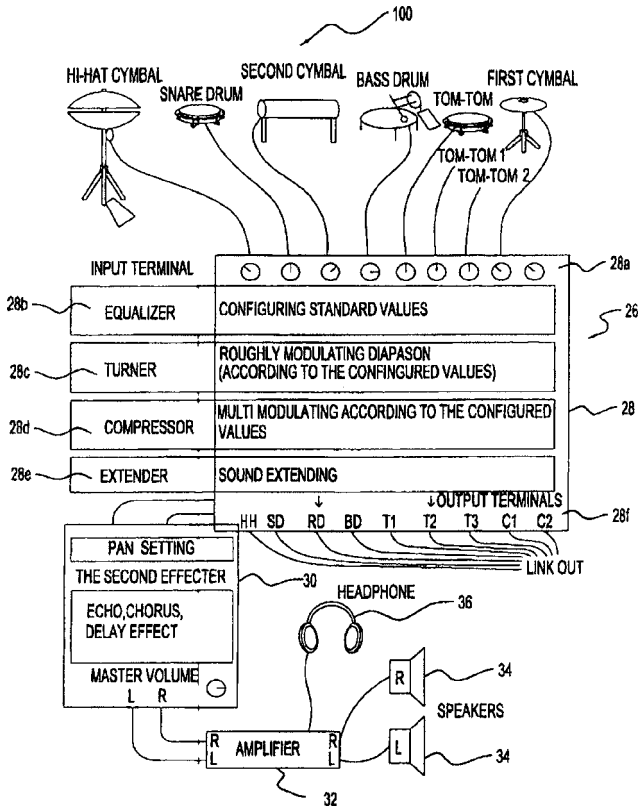
It is right that organ players with happy feet should be left out of the MIDI world? This inventor thinks not and proposes the obvious way to construct such a device.—MK

7,015,391

43.75.Wx ANALOG ELECTRONIC DRUM SET, PARTS FOR DRUM STICK, ANALOG ELECTRONIC DRUM SET AND FOOT-PEDAL UNIT

Shingo Tomoda, Hachiouji-shi, Tokyo, Japan
21 March 2006 (Class 84/726); filed in Japan 7 September 2000

Rather than depend on MIDI effects for a drum set (MIDI has time



delay), the inventor wants to use analog effects. Besides being obvious, the patent is full of poor translations and spellings.—MK

7,019,205

43.75.Wx ENTERTAINMENT SYSTEM, ENTERTAINMENT APPARATUS, RECORDING MEDIUM, AND PROGRAM

Takafumi Fujisawa *et al.*, assignors to Sony Computer Entertainment Incorporated
28 March 2006 (Class 84/609); filed in Japan 14 October 1999

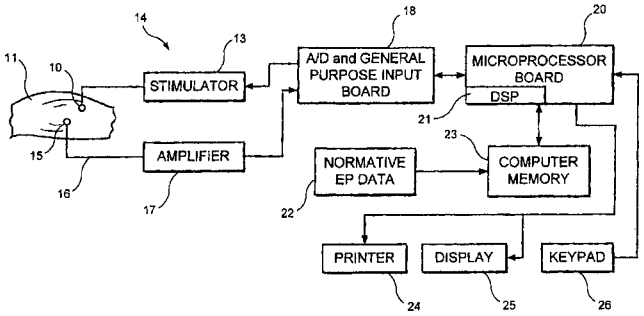
Another in the long line of Sony patents covering the PlayStation, this disclosure covers the user interface of the music editor. While specific to the manual controller of the PlayStation, the inventors do describe the interface in a fair amount of detail.—MK

7,016,722

43.80.Qf SYSTEM AND METHOD FOR FETAL BRAIN MONITORING

Leslie S. Prichep, assignor to New York University
21 March 2006 (Class 600/544); filed 28 February 2003

The device and methodology are directed to a noninvasive method of monitoring of brain wave response of a fetus *in utero*. The method involves connecting an auditory transducer, including at least one biosensor electrode, to a mother's abdomen. The transducer is pulsed with rectangular waves and tone pips of a selected frequency to emit pulsed audible sounds at predetermined times in combination with the steps of detecting, for each



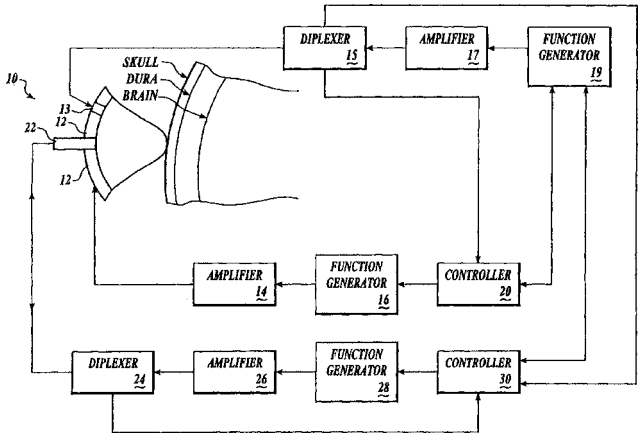
pulsed audible sound, a series of voltage oscillations corresponding to brainstem auditory evoked responses (BAER) of the fetus. The oscillations are time-locked to the corresponding audible sound. A computer-based quantitative EEG system improves the signal-to-noise ratio of the BAER digital data using an adaptive-optimum filtering algorithm and analyzes the BAER digital data relative to a reference database.—DRR

7,022,077

43.80.Qf SYSTEMS AND METHODS FOR MAKING NONINVASIVE ASSESSMENTS OF CARDIAC TISSUE AND PARAMETERS

Pierre D. Mourad *et al.*, assignors to Allez Physionix Limited
4 April 2006 (Class 600/449); filed 1 July 2003

A number of methods and systems are disclosed in this patent for noninvasive assessment of cardiac-tissue properties and cardiac properties through the use of ultrasound. Determination of myocardial tissue stiffness, tension, strain, strain rate, and other mechanical properties may be used to



assess myocardial contractility, myocardial ischemia and infarction, ventricular filling, atrial pressures, and diastolic functions. Noninvasive systems are also described in which ultrasound is used to acoustically stimulate or palpate target cardiac tissue or to induce a response at a cardiac tissue site that relates to cardiac tissue properties and/or cardiac parameters.—DRR

7,022,076

43.80.Qf BONE VELOCITY DETERMINATION

Edward Kantorovich and Yehuda Niv, assignors to Sunlight Medical Limited
4 April 2006 (Class 600/449); filed in Israel 8 November 1998

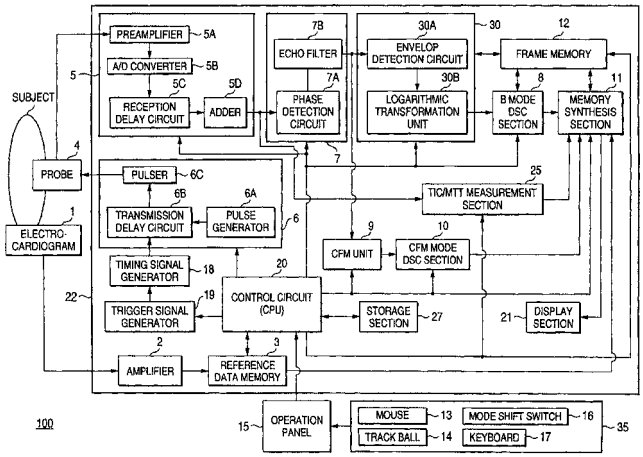
The object of this device is to noninvasively measure the mechanical properties of bone through determination of the sound velocity measured along the main axis of the bone. In some preferred embodiments, an ultrasonic wave is transmitted through a joint between two bones. In one of these preferred embodiments, the wave is transmitted from the pelvis to the knee, thereby passing through a pelvic bone and a hip bone. A moderately high ultrasonic frequency is used so that the wavelength is smaller than the cross-section diameter of the bone, or at most, two or three times its size, for example, about 160 kHz for a femoral neck. Theoretically, if the cross-sectional diameter is greater than 0.7 times the wavelength, the velocity of the wave is about the same as if the cross section were infinite.—DRR

7,022,079

43.80.Qf ULTRASONIC DIAGNOSTIC APPARATUS

Yoichi Ogasawara, assignor to Kabushiki Kaisha Toshiba
4 April 2006 (Class 600/458); filed in Japan 22 August 2001

This ultrasonic diagnostic device uses multiple ultrasonic pulses at sufficiently high intensities to cause the collapse of a contrasting agent injected into the patient. An ultrasonic reception section receives an echo-signal cluster from the irradiated subject as the result of these electronic



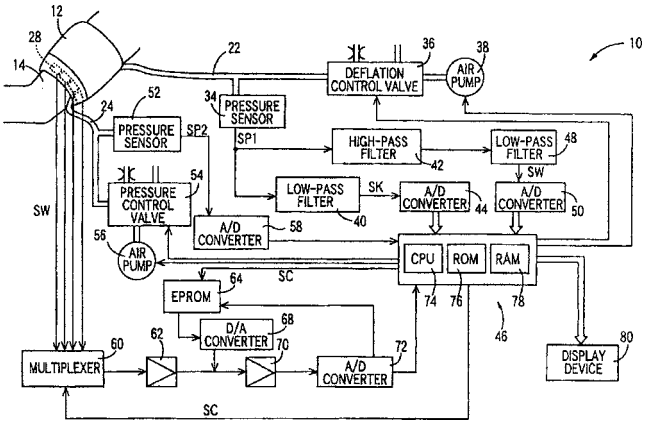
pulses and combines the echo with an additive signal to generate multiple rf data signals. A TIC/MTT section measures and displays the time intensity curve (TIC) and measures the mean transit time (MTT) of the blood flow based on the TIC.—DRR

7,022,084

43.80.Qf HEART-SOUND DETECTING APPARATUS AND HEART-SOUND DETECTING METHOD

Toshihiko Ogura, assignor to Colin Medical Technology Corporation
4 April 2006 (Class 600/528); filed in Japan 7 February 2001

This is an apparatus for detecting a heart sound that uses a pressure-pulse-wave sensor designed to be worn on a body part of the patient that is distant from the chest. The pressure-pulse wave is generated by an artery at



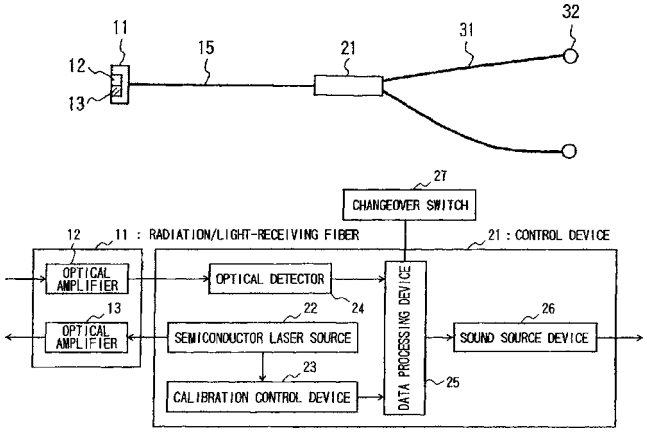
the location where the sensor is placed. The sensor generates a pressure-pulse-wave signal that is processed by a heart-sound extracting device.—DRR

7,024,001

43.80.Qf STETHOSCOPE

Tsutomu Nakada, assignor to Japan Science and Technology Corporation
4 April 2006 (Class 381/67); filed in Japan 30 September 1999

This is not the usual stethoscope for listening to heart beats. Rather, it is a portable device for acquiring data regarding hemoglobin of an organism by the use of near-infrared light. A radiation/light-receiving fiber serves as a probe for noninvasively irradiating a diseased part with near-infrared light. The probe is applied to the diseased part in order to measure, for example, a change of the cerebral circulation blood flow. The change is heard as



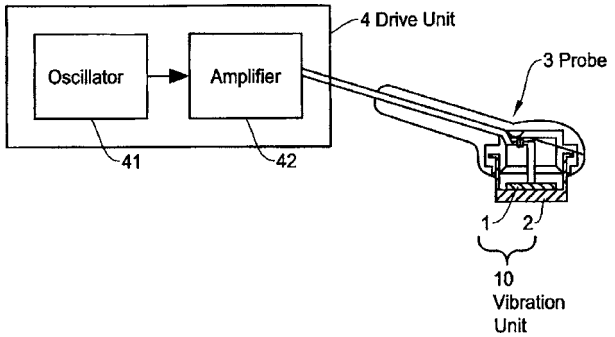
sound pulse modulation. More specifically, a light beam of three wavelengths, $\lambda = 760$, 800, and 830 nm, from a semiconductor laser light source is applied to the diseased part, the reflection data from the diseased part are processed by a control device, and the change in reflection data is converted to a change in pulse frequency of the sound of constant pitch and volume, thereby enabling a physician to carry out auscultation with his/her own ears.—DRR

7,022,089

43.80.Sh ULTRASONIC WAVE COSMETIC DEVICE

Takafumi Ooba *et al.*, assignors to Matsushita Electric Works, Limited
4 April 2006 (Class 601/2); filed in Japan 28 May 2001

This device is a possible tool for health spas which would produce ultrasonic wave stimulation of facial skin by an ultrasonic-wave vibration unit. The device includes a probe provided with a vibration unit in which one surface of the vibration unit contacts the skin and another surface is



structured as a horn connected to the ultrasound transducer. A transducer drive unit produces a fundamental frequency which results in a half wavelength that matches the thickness of the vibration unit (equal to the sum of the transducer and the horn). Alternatively, the drive can produce a harmonic frequency that is an integer multiple of the fundamental frequency.—DRR

7,022,075

43.80.Vj USER INTERFACE FOR HANDHELD IMAGING DEVICES

Sorin Grunwald *et al.*, assignors to Zonare Medical Systems, Incorporated
4 April 2006 (Class 600/446); filed 29 October 2003

This interface includes icons, tabs, and menu items that vary with context and can be voice activated or selected by touch.—RCW

7,022,073

43.80.Vj BORDER DETECTION FOR MEDICAL IMAGING

Liexiang Fan *et al.*, assignors to Siemens Medical Solutions USA, Incorporated
4 April 2006 (Class 600/437); filed 2 April 2003

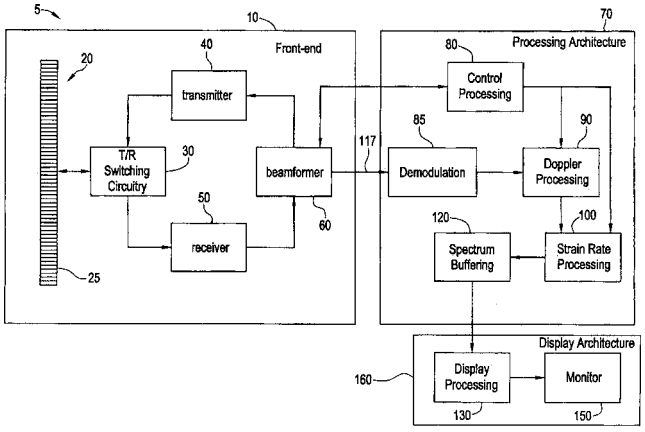
A boundary in an image is defined by a curve with a tangent that is nominally perpendicular to the gradient of the image amplitude along the curve. Starting from the initial boundary, a boundary associated with a minimum cost function or associated with the closest boundary where the tangent and the gradient are orthogonal is identified. The initial boundary is refined to minimize the departure of the tangent from being normal to the gradient or by identifying a boundary based on the orthogonality relationship.—RCW

7,022,078

43.80.Vj METHOD AND APPARATUS FOR SPECTRAL STRAIN RATE VISUALIZATION

Andreas Heimdal and Hans Garmann Torp, assignors to GE Medical Systems Global Technology Company, LLC
4 April 2006 (Class 600/453); filed 28 February 2002

A strain rate corresponding to the deformation of tissue in a region



defined by a time gate is computed using Doppler echo signals that are processed to obtain spectra.—RCW

7,011,632

43.80.Vj METHODS AND APPARATUS FOR ULTRASONIC COMPOUND IMAGING

Franz Steinbacher and Josef Steininger, assignors to Kretztechnik AG
14 March 2006 (Class 600/437); filed 18 September 2001

Two or more ultrasound transmit beams are directed to the same scan position. The beams differ in focus depth, aperture size, center frequency, or

pulse length. Echos from the beams are received with bandwidths or center frequencies that are varied. Beams formed from the echos can be weighted before or after compounding.—RCW

7,022,074

43.80.Vj METHOD AND APPARATUS FOR GENERATING A MULTI-LEVEL ULTRASOUND PULSE

Kjell Kristoffersen and Hans Garmann Torp, assignors to GE Medical Systems Global Technology Company, LLC
4 April 2006 (Class 600/437); filed 12 June 2003

A sequence of three pulses is produced, each having different amplitudes. One of the amplitudes is positive, another of the amplitudes is negative, and the third amplitude is an intermediate level. The pulse sequence is applied to a transducer to produce desired excitations for transmit beamforming.—RCW

7,022,081

43.80.Vj HEAT SEALABLE TRANSDUCER SHIELD AND METHOD OF APPLICATION

Nelson H. Oliver, assignor to Siemens Medical Solutions USA, Incorporated
4 April 2006 (Class 600/459); filed 20 August 2003

Diced ultrasonic-transducer stacks are covered with a film that is bonded to the stacks without using heat. The layer forms an insulating barrier to the environment. Since adhesive is not used for bonding, kerfs in the stacks are air filled.—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.

The Pekeris waveguide revisited (L)

Ivan Tolstoy[a)]
Knockvennie, Castle Douglas, Kirkcudbrightshire, United Kingdom, DG73PA

(Received 12 June 2006; accepted 14 June 2006)

The traditional Pekeris treatment of the shallow water acoustic waveguide is marred by a seeming paradox: The so-called nonorthogonality of the propagation modes. This anomaly is due to the use of mistaken orthogonality criteria, and is easily exorcised by the use of normal coordinate theory (which leads to the correct orthonormality conditions for any linear conservative system).
 [DOI: 10.1121/1.2221536]

PACS number(s): 43.30.Bp, 43.20.Bi, 43.20.Mv [ADP] Pages: 1183–1185

I. INTRODUCTION

The classic Pekeris[1] waveguide model for shallow water sound propagation consists of a water layer of thickness h, sound velocity c_1, density ρ_1, overlying a homogeneous fluid half-space of sound velocity c_2 and density ρ_2. If $c_1 < c_2$, part of a source's energy may be trapped in the surface layer by total reflection. Given a harmonic point source of frequency ω higher than a cutoff value determined by the parameters h, c_1, c_2, ρ_1, and ρ_2, this type of model will in general display two possible mode families:

1) Modes with a discrete spectrum of wave numbers k_m (eigenvalues), corresponding to guided waves whose energy is concentrated in the surface layer by total reflection, and whose amplitude in the underlying half-space falls off exponentially with distance from the interface.
2) Modes with a continuous wave number spectrum corresponding to waves partially reflected at the 1-2 boundary, whose energy extends through the whole half-space, i.e., their amplitudes, are sinusoidal functions of the depth.

Use of inappropriate orthogonality criteria led Pekeris to conclude that the guided modes of this model were not orthogonal.

II. OUTLINE OF THEORY

Acoustic mode properties are determined by second-order differential equations for the field in a fluid in terms of displacements, velocities pressures or, when feasible, potentials. Thus, using $\mathbf{a}$ for the fluid displacement perturbation, p for the corresponding pressure perturbation, ρ for the bulk fluid density, and c for the sound velocity, Euler's classic equations of fluid dynamics are, for small motions:

$$\rho\ddot{\mathbf{a}} = -\nabla p, \tag{1}$$

where dots represent differentiation with respect to time, and p is proportional to the volume strain ε:

$$\varepsilon = \nabla\mathbf{a}, \tag{2}$$

$$p = -\lambda\varepsilon, \tag{3}$$

where λ is the bulk modulus:

$$\lambda = \rho c^2. \tag{4}$$

If ρ is not a function of the coordinates then, by Eq. (1), curl $\mathbf{a}=0$, i.e., the field is irrotational and may be derived from a potential at any point in space:

$$\mathbf{a} = \nabla\varphi, \tag{5}$$

and it is seen that φ obeys the standard wave equation:

$$\ddot{\varphi} = c^2\nabla^2\varphi, \tag{6}$$

where c can be any function of the space coordinates. Standard procedure is to assume simple harmonic solutions, satisfy a set of linear homogeneous boundary conditions, and solve the resulting characteristic equation by a spectrum of wave numbers. In the case of a finite system, the spectrum is discrete, giving wave numbers, i.e., eigenvalues $\mathbf{k}_m$ and eigenfunctions φ_m. If the system is separable, e.g., if it is stratified in the z (vertical) direction, the latter will satisfy a simple orthonormality condition:

$$\int_0^l \varphi_n\varphi_m dz = \mu_m\delta_{nm}, \tag{7}$$

where μ_m is the normalization constant and δ_{nm} is the Kronecker delta.

If ρ is a function of the coordinates, curl $\mathbf{a}\neq 0$, and it is not permissible to invoke a potential function everywhere,

[a)]Electronic mail: adp@bu.edu

and reduce the problem to the solution of a wave equation, such as Eq. (6).

In the Pekeris model, $\rho(z)$ has a step behavior at the interface and is therefore a function of the coordinates, albeit its variation is confined to the plane $z=h$. As Pekeris pointed out, the horizontal displacement is discontinuous at this boundary, i.e., $z=h$ is a vortex sheet on which curl $\mathbf{a}\neq 0$. Thus, whereas the displacement fields $\mathbf{a}_1$ and $\mathbf{a}_2$ are irrotational in the regions $z<h$, and $z>h$ and may be described there by potential functions obeying Eqs. (5) and (6), this representation leads to interpretational difficulties at the boundary $z=h$ and it is not surprising that potential eigenfunctions [solutions of Eq. (6)] do not appear to satisfy the standard orthogonality condition Eq. (7).

Pekeris states:[1] "…*even in the case of a continuous sound velocity* $c(z)$, *the peculiar boundary conditions imposed by the sound problem make the normal functions nonorthogonal in the presence of a density discontinuity*." Pekeris' classic monograph was published in 1948 but the myth of nonorthogonality has largely been treated as gospel and persists to this day.[2] Classic Sturm-Liouville theory suggests what is wrong here.

Let us consider the general stratified case in which ρ is a function of z. Let primes indicate differentiation with respect to z, and v represents the z component of $\mathbf{a}$. In the harmonic case, $\ddot{v}=-\omega^2 v$ and Eqs. (1)–(3) yield:

$$(\rho c^2 v')' + \omega^2 \rho v = 0. \tag{8}$$

We recognize the Liouville equation for the inhomogeneous string. Courant-Hilbert[3] and other textbooks give the correct form for the orthonormality condition for a string of length l and eigenfunctions v_n and v_m as

$$\int_0^l \rho v_n v_m = \mu_m \delta_{nm}. \tag{9}$$

One might, at this point, simply observe that Pekeris forgot his Sturm-Liouville theory, and leave it at that. But standard treatments assume that that the coefficients of Eq. (8) are well behaved, and the existence of discontinuities must be addressed separately before one can claim to have exorcized the nonorthogonality myth. While it is possible to do this, it is simpler and more informative to appeal to the displacement field equations and the technique of normal coordinates.[4,5]

III. NORMAL COORDINATES

The technique of normal coordinates is a time-honored method first developed in the study of small vibrations of discrete systems. It has been dealt with in all classical mechanics and acoustic textbooks since Rayleigh's Theory of Sound (1877) (Ref. 6) and as such needs no detailed explanation here, beyond a few remarks which underline how the extension to continuous media is made.

One starts with a system of discrete masses and springs. Assuming small displacements from equilibrium η, the kinetic and potential energies are quadratic functions of the components η_i, η_j and η_i, η_j defining the symmetric matrices T_{ij}, V_{ij}.

For a given (characteristic) frequency ω_k, the normal modes $\mathbf{a}_k$ and normal coordinates q_k are

$$\boldsymbol{\eta} = \sum_k \boldsymbol{\eta}_k, \tag{10}$$

$$\boldsymbol{\eta}_k = q_k \mathbf{a}_k. \tag{11}$$

This choice of η_k leads to the diagonalization of T_{ij}, V_{ij}, and Lagrange's equations show that q_k obeys the simple oscillator equation:

$$\ddot{q}_k + \omega_k^2 q_k = Q_k \mu_k^{-1}, \tag{12}$$

where ω_k is the characteristic frequency, Q_k is the generalized driving force (obtained from the actual forces by the principle of virtual work), and μ_k is a measure of the mode's kinetic energy[4,5] given by diagonalization of the kinetic energy matrix, i.e.,

$$\sum_{i,j} a_{ik} T_{ij} a_{jl} = \mu_k \delta_{kl}. \tag{13}$$

This is the orthonormality condition for the discrete system, and expresses the fact that the vectors $\mathbf{a}_k$ are orthogonal in a space of metric T_{ij}.

Biot and Tolstoy extended the technique to unbounded continuous conservative media of any kind.[4,5] The key then is the calculation of μ_k, which now requires an integral and defines *the orthonormality condition in its most general form*:

$$\int_\tau \rho \mathbf{a}_k \mathbf{a}_l d\tau = \mu_k \delta_{kl}, \tag{14}$$

where the integration extends over the entire space τ occupied by the system.[4,5] This is the generalization to continuous media of the kinetic energy matrix diagonalization condition for discrete systems and is valid for any type of medium—regardless of the order of the (conservative) field equations, of the anisotropy of the medium, or the shape of its boundaries, as long as the relevant boundary conditions are also conservative. It is, in other words, a very general and powerful result.

Equation (14) resolves the nonorthogonality paradox: Pekeris was applying the wrong orthonormality condition to the layered waveguide. This equation shows how, in this simple model, orthogonality is restored by multiplying the integrand in Eq. (7) by the appropriate $\rho(z)$—or, equivalently, multiplying the eigenfunctions by $\rho^{1/2}$.

Apart from resolving the orthogonality paradox, the use of normal coordinates and displacement fields allows one to bypass other complexities of the traditional Pekeris approach. For instance, Eq. (12) is directly soluble for a realistic transient force $Q(t)$ without going through the intermediate step of finding a solution for the simple harmonic source—which, as is well known, requires manipulations in the complex plane and appeals to an *ad hoc* radiation condition to get rid of spurious arrivals from infinity. Equations (10)–(12) lead directly to progressive solutions for physically realizable transient sources with beginnings at $t=t_o$, and bypass these complications.

Finally, the need–and nuisance—of a *branch-line integral* of the kind appearing in Pekeris' solution is eliminated. For example, an integral representing the contribution of the continuous spectrum modes for a $-1(-t)$ source derived by the normal coordinate method is given explicitly in Eq. (3.97) of Tolstoy and Clay.[5] This is easily transcribed into a solution for an $e^{-i\omega}o^t$ source. A straightforward change of contour in the complex plane then explicitly *recovers Pekeris' branch-line integral*—as in his Eq. (A62)—which is therefore seen to be simply the contribution of the continuous spectrum modes, i.e., of that fraction of source energy which is not channeled in the low velocity surface layer. Whereas the branch-line integral gives a correct result, Pekeris was not happy with it, suggesting that it was a "contradiction to standard theory of normal modes" and somehow related to the "question of orthogonality." The normal coordinate theory and, in particular the Tolstoy and Clay result, just alluded to shows that this is not the case.

IV. CONCLUDING REMARKS

The paradox of nonorthogonality of the waveguide modes in the classic Pekeris model arises through the assumption of *mistaken orthogonality conditions*. However, in deriving his solutions for the relevant acoustic fields, Pekeris never appeals to orthogonality; he only needs to assume that his solutions may be expressed as integrals with respect to **k**, and sees to it that these satisfy the relevant boundary conditions of continuity of pressure and normal particle velocities at $z=h$. His description of the acoustic field is therefore unaffected by his erroneous views on the orthogonality question. Nevertheless, this fundamental flaw should be excised from future treatments of this and similar problems, not only as a matter of principle but also because it may continue to prove confusing or even misleading in more detailed studies.

[1]C. L. Pekeris, "Theory of propagation of explosive sound in shallow water," Mem.-Geol. Soc. Am. **27**, 1–117 (1948).

[2]M. J. Buckingham and E. M. Giddens, "On the acoustic field in a Pekeris waveguide with attenuation in the bottom half-space," J. Acoust. Soc. Am. **119**, 123–142 (2006).

[3]Courant-Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. **1**.

[4]M. A. Biot and I. Tolstoy, "Formulation of wave propagation in infinite media by normal coordinates with an application to diffraction," J. Acoust. Soc. Am. **29**, 381–391 (1957).

[5]Tolstoy and Clay, *Ocean Acoustics: Theory and Experiment*, 2nd Ed. (Acoustical Society of America and American Institute of Physics, New York, 1987).

[6]Lord Rayleigh, *Theory of Sound* (Macmillan and Co., Ltd., London, 1877), Vol. **1**.

Utilizing a modified impedance analogy on sound propagation past a hard, curved, rough surface (L)

James P. Chambers[a)]
National Center for Physical Acoustics, University of Mississippi, University, Mississippi 38677

Yves H. Berthelot
Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

(Received 10 November 2005; revised 23 February 2006; accepted 17 March 2006)

The prediction of sound propagation past a curved, rigid, rough surface can be improved by utilizing a modified surface admittance that is calculated by replacing the standard roughness parameter, ϵ, with a complex roughness parameter. The inclusion of a complex roughness, ϵ, accounts for the attenuation of the boundary wave with distance by incoherent scattering. This analogy has been successfully used to predict the sound field behind a rough, curved, rigid surface at several frequencies, by computing the creeping wave series with a best-fit surface parameter, $q=i(kl)\beta=(kl)(k\epsilon)$. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2216767]

PACS number(s): 43.28.En, 43.20.El, 43.20.Fn, 43.50.Vt [KA] Pages: 1186–1189

I. INTRODUCTION

A classical result concerning the reflection of sound from randomly rough surfaces is that incoherent scattering by the surface irregularities causes the reflection coefficient to decrease by a factor of the order $e^{-2k_z^2d^2}$, where d is a characteristic length scale of the roughness and k_z is the wave number component normal to the surface.[1] This incoherent scattering is a second-order effect in kd. However, Biot and Tolstoy have shown that the coherent part of the scattered field is a first-order perturbation in kd which leads to more important effects, especially near the rough boundary.[2–6] Specifically, the coherently scattered field appears in the form of a boundary wave localized near the rough surface. This boundary wave can be important at large distances in outdoor sound propagation near a rough ground. Methods based on perturbation techniques, for instance, cannot predict this first-order effect because these methods are intrinsically second-order methods. The strength of the Biot-Tolstoy approach is that it is a first-order theory, it is valid for steep-slope, densely packed scatterers, it includes *de facto* multiple scattering and backscattering, and it is valid at grazing incidence. In addition, it is also valid, in principle, for nonisotropic roughness, and even for space-dependent surface statistics. All these elements are common occurrences in long-range sound propagation.

In a nutshell, the Biot-Tolstoy approach consists of replacing the rough boundary by a distribution of monopoles and dipoles whose scattering coefficients depend on the geometry of the boundary. If the scatterers are smaller than a wavelength, their strength can be calculated from Rayleigh scattering theory. In the case of isotropic roughness, the Biot-Tolstoy boundary condition is written as[3]

$$\phi_z + \epsilon(k^2\phi + \sigma\phi_{zz}) = 0, \tag{1}$$

where ϕ is the acoustic velocity potential, the z axis is normal to the surface baseline, k is the acoustic wave number in the upper fluid, and ϵ and σ are two parameters directly related to the geometry of the rough surface.[3,7] For the purpose of this paper, and for the discussion below, the scatterers are assumed to be perfectly rigid, although this is not a limitation of the Biot-Tolstoy approach.[4,8]

When $|\sigma\phi_{zz}| \ll |k^2\phi|$, i.e., when $\sigma\sin^2\theta \ll 1$, where θ is the grazing angle, the boundary condition (1) is approximated by

$$\phi_z + \epsilon k^2\phi = 0. \tag{2}$$

This approximate boundary condition is therefore appropriate at near-grazing sound propagation and/or for sparse scatterers. This boundary condition at a rough interface is in the form of a standard impedance boundary condition at a flat plane (with the $e^{-i\omega t}$ convention),

$$\phi_z + \eta\phi = 0, \tag{3}$$

where $\eta = ik\beta$ and $\beta = \rho c/Z_s$ is the normalized surface admittance, ρc being the characteristic impedance of the upper fluid. By comparing boundary conditions (2) and (3), it appears that there is a formal analogy between sound propagation near a smooth impedance plane and near a randomly rough, rigid plane, if $\eta = k^2\epsilon$, i.e., if

$$\beta = -ik\epsilon. \tag{4}$$

This analogy has been suggested by several authors, and it has also been discussed in the context of nonrigid rough surfaces.[4,8,9] With this analogy, the Biot-Tolstoy boundary wave that propagates along a rough surface is formally equivalent to Wenzel's trapped surface wave[10] near an impedance plane with large reactance. Indeed, Wenzel's Eq. (20) in Ref. 10 is exactly the same as Tolstoy's Eq. (13) in Ref. 6 if $\eta = k^2\epsilon$, and κ, the wave number of the boundary mode, is $\sqrt{k^2+\eta^2}$. Both boundary modes (surface wave or

a)Author to whom correspondence should be addressed. Electronic mail: chambers@olemiss.edu

boundary wave) are confined to an exponential layer near the boundary with $1/e$ height $(\epsilon k^2)^{-1}$ (or $1/\eta$).

II. COMPLEX IMPEDANCE OR ADMITTANCE

The analogy represented by Eq. (4) is appealing but it must be taken cautiously. In fact, it is easy to show that, because ϵ is a real number, the equivalent surface impedance is purely imaginary and there is no corresponding attenuation of the boundary wave, an unphysical attribute. Previous work has shown indeed that propagation past a rough surface must incorporate incoherent scattering which manifests itself through a real part of the admittance.[11–13] This real part of the admittance will produce attenuation $e^{-\delta r}$ of the boundary wave with distance r along the boundary. One potential method to introduce a complex admittance is to let $\epsilon = |\epsilon| e^{im}$, where $|\epsilon|$ is the usual roughness parameter discussed extensively by Tolstoy, Medwin, and, recently, Attenborough and Chambers *et al.*,[3,7,8,14] which incorporates the mean height, rms height, and correlation length for random 3D elements. While it may seem somewhat unphysical to have a complex length, it should be viewed as a mathematical construct that, if successful, will provide a fairly simple and straightforward method to incorporate 3D random roughness into propagation routines. It has been shown previously that for propagation over flat rough surfaces, the roughness to impedance analogy works quite well to describe the sound field for near-grazing propagation, with the predominant feature being a shift in the ground effect dip.[7,14] The analogy breaks down at grazing incidence and for propagation into the shadow zone of curved rough surfaces, which is inherently a grazing incidence propagation.[14,15] In the rest of the paper, the analogy between the rough plane boundary condition and the impedance plane boundary condition is tested with the proposed complex roughness parameter ϵ in an attempt to predict the data reported by Chambers and Berthelot[16]in the case of the sound field in the shadow zone of a curved, rigid, rough surface.

III. DIFFRACTION OF SOUND BY A CURVED, RIGID, ROUGH SURFACE

The sound field behind a curved, smooth, boundary of finite normalized admittance, β, is accurately modeled by a creeping wave series (CWS) which depends on a parameter q that includes-both the effect of curvature and the finite impedance of the surface.[17] In fact, q is defined by

$$q = i(kl)\beta, \tag{5}$$

where l is the Fock diffraction length defined by $l = (R/2k^2)^{1/3}$, k being the acoustic wave number and R the radius of curvature of the hill ($kR \gg 1$). Following the analogy $\beta = -ik\epsilon$ [see Eq. (4)], one can model the sound field behind a curved, rough, rigid surface with the creeping wave series in which the surface parameter q is replaced by $(kl) \times (k\epsilon)$, where ϵ must be complex to account for the attenuation of the boundary wave through incoherent scattering.

This model was tested against the experimental data reported in Ref. 16. Because the attenuation coefficient for a boundary wave along a curved surface is not known, the problem was reformulated as follows: is there any complex coefficient q (i.e., any complex admittance, or in our methodology, any complex roughness parameter ϵ) that can be used in the creeping wave series to model accurately the data reported in Ref. 16?

TABLE I. Optimum values of q to fit data.

f(kHz)	$\|q\|$	Phase q	rms error (dB)
5	0.48	0.25	0.25
6	0.61	0.22	0.35
8	0.81	0.23	0.6
10	1.1	0.27	1
12	1.49	0.29	1.7

A. Optimum parameter *q*

The optimum value of q was found by a least-square minimization of the difference between the model (CWS) and the data, successively at 5, 6, 8, 10, and 12 kHz. For each frequency, all three geometries reported in Ref. 16 were used: on the surface, along the line of sight, and along a vertical axis starting from the surface in the shadow zone. The radius of curvature was approximately 2.5 m, the data on the surface were taken from −0.2 to +1.0 m, the data along the line of sight were taken from 0 to 1.4 m, and the data along the vertical axis, at an arc radius of 1.00 m behind the apex, ranged from the surface ($z=0$ m) to 0.40 m, and crossed the line of sight at 0.20 m.

The quantity that was minimized is the standard deviation, $\bar{\sigma}$, between the theoretical and experimental values of the sound-pressure levels relative to free field,

$$\bar{\sigma} = \left[\frac{1}{N} \sum_{i=1}^{N} (L_{\text{theo}} - L_{\text{data}})^2 \right]^{1/2}, \tag{6}$$

where the summation is over the N data points at that frequency for the three geometries discussed above. The values for L_{theo} are obtained by use of the creeping wave series[17] and are not repeated here for brevity. The results are given in Table I.

For the geometry given in Fig. 1, the results are shown in Figs. 2–4 at 6 and 12 kHz, respectively. Each figure contains two plots of the insertion loss level for each configuration. The abscissa is either the arclength s along the surface with $s=0$ being the apex (Fig. 2), the horizontal distance x along the line of sight with $x=0$ being the apex (Fig. 3), or the vertical axis that starts at $s=1$ m in the deep shadow

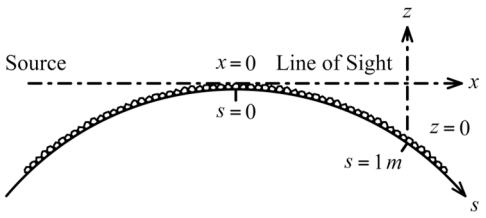


FIG. 1. Geometry of the Problem.

FIG. 1. Geometry for the problem at hand.

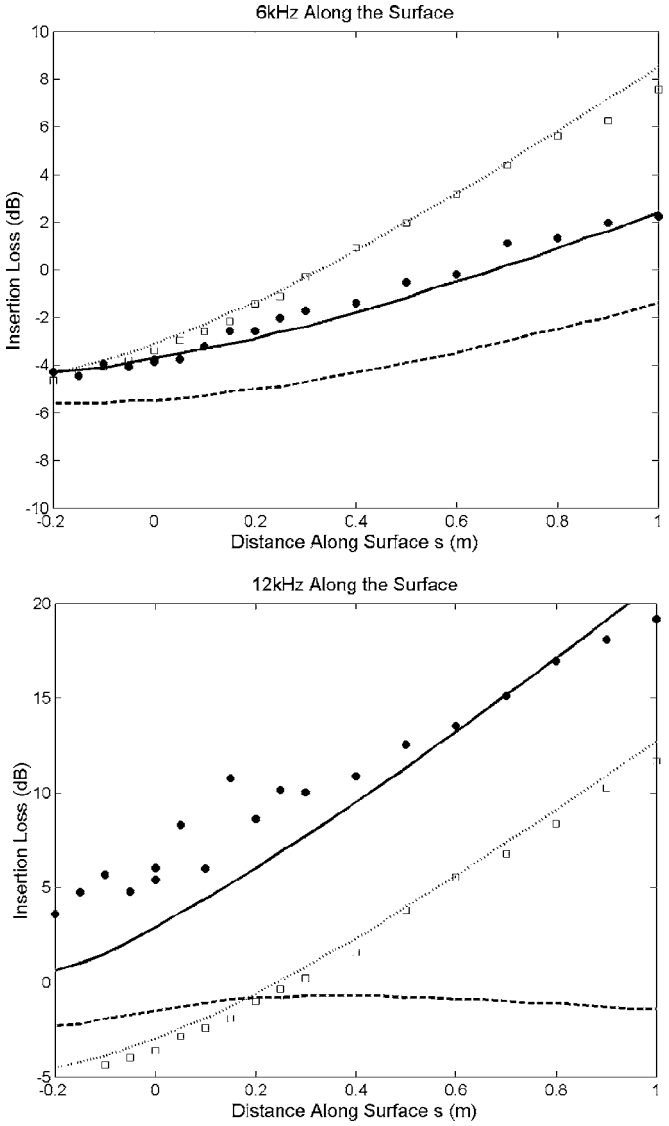


FIG. 2. Insertion loss along the surface at (a) 6 kHz and (b) 12 kHz. Data over the rough surface are presented as solid circles and $s=0$ m is the top of the surface. The theory for the smooth surface is the dotted line while the squares are the measured data. The theory utilizing the standard rough-impedance analogy is the dashed line. The newly proposed model using a modified admittance of roughness parameter (best fit for each frequency) is the solid line.

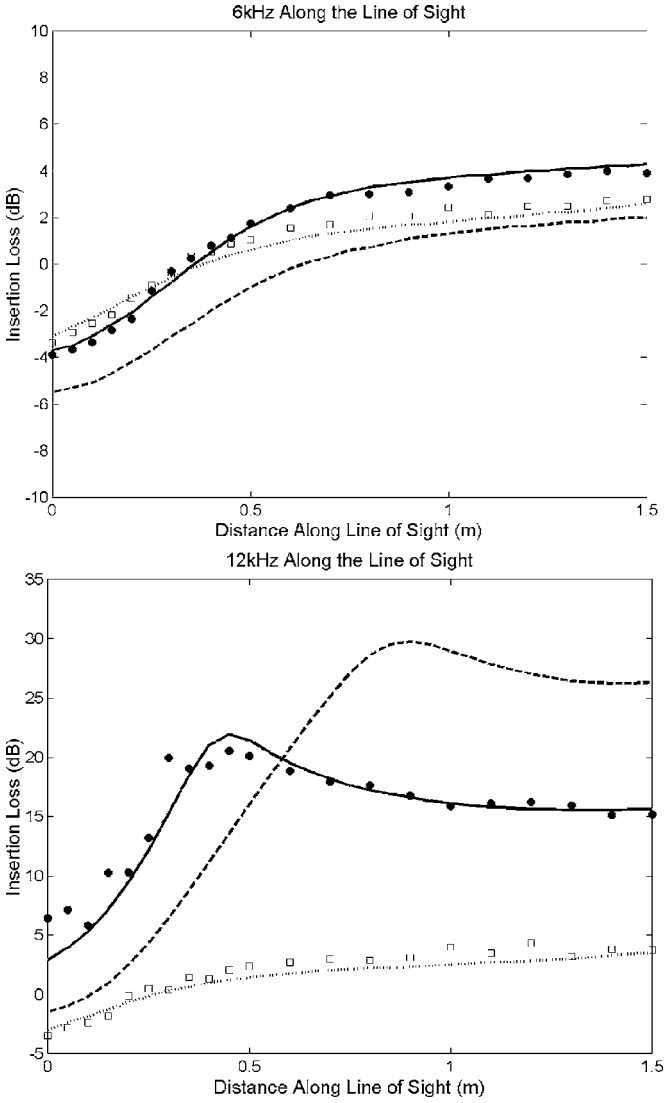


FIG. 3. Insertion loss along the line of sight at (a) 6 kHz and (b) 12 kHz. Data over the rough surface are presented as solid circles and 0 m is the top of the surface. The theory for the smooth surface is the dotted line while the squares are the measured data. The theory utilizing the standard rough-impedance analogy is the dashed line. The newly proposed model using a modified admittance or roughness parameter (best fit for each frequency) is the solid line.

behind the apex (Fig. 4). In Figs. 2–4, the circles are the data points (curved, rough, rigid surface) and the solid line represents the creeping wave series with the optimum value of q (see Table I). To illustrate the importance of the roughness effects, the theory for $q=0$, i.e., for a rigid, smooth surface is plotted with a dotted line and the experimental data with squares. The dashed line corresponds to the analogy [Eq. (4)] in which ϵ is taken to be a real quantity which will be referred to as the "standard analogy." It is worth noting that the effects of the roughness can indeed be dramatic with +6 to −20-dB variations in insertion loss observed for rough surfaces as compared to smooth surfaces.[16]

Clearly, at each frequency, there is a complex q, or alternatively, a complex admittance or a complex ϵ, that can be used to model accurately the sound field behind the curved, rigid, rough surface. The analogy presented in Eq. (4) appears to be indeed a powerful modeling tool, provided that one allows for a real portion of the admittance, The analogy fails when the admittance is purely imaginary. If the analogy described by Eq. (4) holds, the ratio $|q|/kl$ should be equal to $k|\epsilon|$. The results are shown in Table II, where the roughness parameter $|\epsilon|$ was taken to be about 1 mm.[14] Indeed, the ratio of $|q|/kl$ to $k|\epsilon|$ is nearly equal to unity at all frequencies, thus confirming the validity of the analogy, at least with respect to the modulus of the roughness parameter.

IV. SUMMARY AND CONCLUSIONS

As has been reported for a flat rough surface,[7,12,13] the propagation of sound over a curved, rigid, rough surface can be accurately modeled by introducing a real portion to the surface admittance. Here, this has been achieved by utilizing a complex roughness parameter ϵ whose magnitude is the

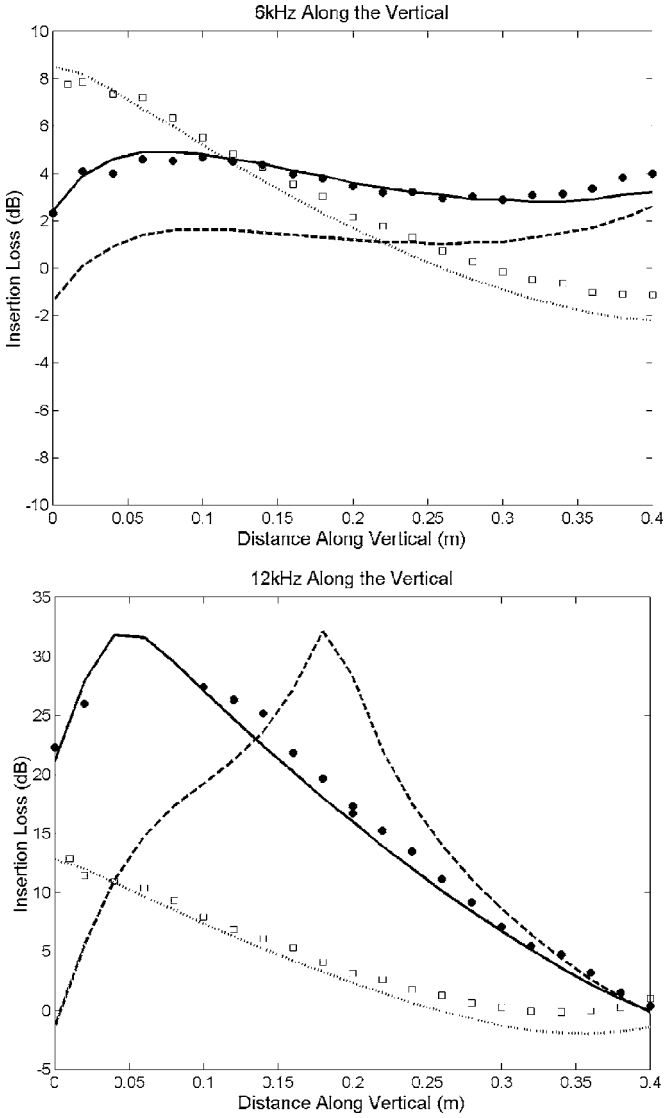


FIG. 4. Insertion loss along vertical at $s=1.0$ m in the shadow zone at (a) 6 kHz and (b) 12 kHz. Data over the rough surface are presented as solid circles and 0 m is on the surface while 0.2 m is the shadow-bright boundary. The theory for the smooth surface is the dotted line while the squares are the measured data. The theory utilizing the standard rough-impedance analogy is the dashed line. The newly proposed model using a modified admittance or roughness parameter (best fit for each frequency) is the solid line.

standard roughness parameter discussed by Tolstoy and Medwin, and whose phase accounts for the attenuation of the boundary wave with distance. With this complex roughness ϵ, the analogy between a flat impedance boundary condition and a rough, rigid boundary condition is $\beta=-ik\epsilon$, where β is the normalized surface admittance and k the wave number.

TABLE II. Dependence of $|q|/(kl)$ on $k|\epsilon|$.

f(kHz)	$\lvert q\rvert$	kl	$\lvert q\rvert/(kl)$	$k\lvert\epsilon\rvert$	$\lvert q\rvert/(kl)/k\lvert\epsilon\rvert$
5	0.48	4.85	0.099	0.091	1.08
6	0.61	5.15	0.118	0.109	1.08
8	0.81	5.67	0.143	0.145	0.98
10	1.1	6.11	0.180	0.183	0.99
12	1.49	6.49	0.230	0.219	1.05

This method has been successfully used to predict the sound field behind a curved, rigid surface at several frequencies, successively on the surface, along the line of sight, and along a vertical axis starting in the deep shadow, by computing the creeping wave series with a modified surface parameter $q=i(kl)\beta=(kl)(k\epsilon)$, ϵ being now complex. The form for the complex ϵ has been determined empirically by a best fit of the parameter q to measured data. A more extensive data collection, incorporating a larger range of $k\epsilon$ and kR, is underway to determine an appropriate form for the phase of the complex roughness length scale, ϵ, envisioned here.

Clearly, surface roughness should be taken into account for prediction purposes in long-range sound propagation, either over natural hills covered with individual scatterers (assumed to be smaller than the wavelength), or in the case where the shadow zone is caused by sound refraction in the atmosphere.

ACKNOWLEDGMENTS

The work and manuscript preparation were supported in part by a grant from the Fannie and John Hertz Foundation and the U.S. Army Research Laboratory.

[1]J. A. Ogilvy, *Theory of Wave Scattering from Random Surfaces* (Adam Hilger, New York, 1991).
[2]M. A. Biot, "Generalized boundary condition for multiple scatter in acoustic reflection," J. Acoust. Soc. Am. **44**, 1616–1622 (1968).
[3]I. Tolstoy, "Scattering of spherical pulses by rough surfaces," J. Acoust. Soc. Am. **66** (4), 1135–1145 (1979).
[4]I. Tolstoy, "Energy transmission into shadow zone by rough surface boundary wave," J. Acoust. Soc. Am. **69** (5), 1290–1298 (1981).
[5]I. Tolstoy, "Smoothed boundary conditions, coherent low frequency scatter, and boundary modes," J. Acoust. Soc. Am. **75** (5), 1–22 (1984).
[6]I. Tolstoy, "Rough surface boundary wave attenuation do to incoherent scatter," J. Acoust. Soc. Am. **77** (2), 482–491 (1985).
[7]K. Attenborough and S. Taherzadeh, "Propagation from a point source over a rough finite impedance boundary," J. Acoust. Soc. Am. **98** (3), 1717–1722 (1995).
[8]H. Medwin and G. D'Spain, "Near-grazing, low frequency propagation over randomly rough rigid surfaces," J. Acoust. Soc. Am. **79** (3), 657–665 (1986).
[9]M. S. Howe, "On the long range propagation of sound over irregular terrain," J. Sound Vib. **98**, 83–94 (1985).
[10]A. R. Wenzel, "Propagation of waves along an impedance boundary," J. Acoust. Soc. Am. **55** (5), 956–963 (1974).
[11]R. J. Lucas and V. Twersky, "Coherent response to a point source irradiating a rough plane," J. Acoust. Soc. Am. **76**, 1847–1863 (1984).
[12]P. Boulanger, K. Attenborough, S. Taherzadeh, T. Waters-Fuller, and K. M. Li, "Ground effect over hard rough surfaces," J. Acoust. Soc. Am. **104** (3), 1474–1482 (1998).
[13]P. Boulanger, K. Attenborough, S. Taherzadeh, and Q. Qin, "Effective impedance of surfaces with porous roughness: Models and data," J. Acoust. Soc. Am. **117** (3), 1146–1156 (2005).
[14]J. P. Chambers, R. Raspet, J. M. Sabatier, and Y. H. Berthelot, "Incorporating the effects of roughness in outdoor sound propagation models," Proceedings of Noise-Con 96, 905–910 (1996).
[15]Y. H. Berthelot and J. P. Chambers, "On the analogy between sound propagation over a rough surface and sound propagation over a smooth surface with modified surface impedance," Proceedings of the 6th Long Range Sound Propagation Symposia (1996).
[16]J. P. Chambers and Y. H. Berthelot, "An experimental investigation of the propagation of sound over a curved, rough, rigid surface," J. Acoust. Soc. Am. **102** (2), 707–714 (1997).
[17]A. Berry and G. A. Daigle, "Controlled experiments on the diffraction of sound by a curved surface," J. Acoust. Soc. Am. **83** (6), pp. 2047–2058 (1988).

Adjoint modeling with a split-step Fourier parabolic equation model (L)

Kevin B Smith[a)]
Department of Physics, Naval Postgraduate School, Monterey, CA 93943

(Received 9 December 2005; revised 16 June 2006; accepted 18 June 2006)

Previous work on adjoint modeling of underwater acoustic propagation for the purposes of environmental inversions developed the necessary numerical algorithms based on a finite difference implementation of the parabolic equation. In this letter, the corresponding formulation for the split-step Fourier implementation is presented. Although no calculations have yet been made with this approach, the method is completely consistent with previous work, and the final form of the tangent linear model is exactly the same. However, the formulation of the operators in the split-step Fourier approach provides a simple means of estimating both local and global effects on uncertainty. This letter is then provided for anyone who may wish to attempt adjoint modeling with a split-step Fourier model. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2221559]

PACS number(s): 43.30.Pc, 43.60.Pt [AIT] Pages: 1190–1191

I. INTRODUCTION

Recent interest in acoustic variability due to environmental uncertainty, and the means of estimating environmental parameters through model-data comparisons, has led to a variety of approaches for acoustic data inversion processing. Of particular interest here is the adjoint method of computing acoustic variability due to unknown environmental fluctuations, and the inversion of observed mismatch between measured and modeled acoustic data. In their work, Hursky *et al.*[1] developed the associated tangent linear model for acoustic variability in the context of a parabolic equation model based on a first-order finite difference computational algorithm. The goal of this letter is to develop the corresponding tangent linear model equations consistent with a split-step Fourier (SSF) implementation.

While the approach outlined in this letter is not expected to provide improvements in the final inversion results over the previously defined finite difference approaches[1,2] there are advantages to using a SSF approach in the forward problem. Other than the simple manner in which a tangent linear model can be coded, the split-step algorithm propagates the field, and its perturbation, forward by a simple well-defined propagation operator. This allows the first-order perturbation approach to produce simple expressions for the forward propagation of the acoustic variability/uncertainty. As will be shown, these expressions have physical significance that highlight the manner in which the acoustic variability/uncertainty evolves in the propagating field.

II. THEORETICAL DEVELOPMENT

The framework of the approach closely follows the work of Hursky *et al.*[1] We begin by defining the parabolic equation (PE) field function and the associated marching algorithm used to propagate the solution out in range. In terms of the acoustic pressure, $p(\mathbf{r})$, the PE field function, $\psi(\mathbf{r})$, can be defined simply by[3,4]

$$p(\mathbf{r}) = \frac{1}{\sqrt{r}}\psi(\mathbf{r})e^{ik_0 r}. \tag{1}$$

The evolution of the field then follows by employing the marching algorithm:

$$\boldsymbol{\psi}(r+\Delta r) = e^{-i\Delta r k_0 \mathbf{H}_{\mathrm{op}}(r)}\boldsymbol{\psi}(r), \tag{2}$$

where the implementation is performed with the SSF algorithm.[4,5] Note that we shall use a notational convention where bold, lower case variables are vectors in depth while bold, upper case variables are matrices. In Eq. (2), the Hamiltonian operator is defined by the Thomson and Chapman wide-angle approximation,[6] i.e.,

$$\mathbf{H}_{\mathrm{op}}(r) = \mathbf{T}_{\mathrm{op}} + \mathbf{U}_{\mathrm{op}}(r), \tag{3}$$

where $\mathbf{T}_{\mathrm{op}}$ is a range-independent operator (containing the differential operator $\partial^2/\partial z^2$), and $\mathbf{U}_{\mathrm{op}}(r)$ is a diagonal matrix with elements on the jth row/column (depth z_j) of $[1 - n_j(r)]$. The spatial variations of the environment are captured in the definition of the index of refraction, $\mathbf{n}(r)$. Note that this formulation does not yet include density variations, but this can be incorporated later. In the SSF algorithm, the operator $\mathbf{T}_{\mathrm{op}}$ is applied in the vertical wave number domain, where it also becomes a diagonal matrix.

Perturbations to the environment will then only produce variations in $\mathbf{U}_{\mathrm{op}}$ (and not in $\mathbf{T}_{\mathrm{op}}$). Performing a perturbation expansion of $\mathbf{U}_{\mathrm{op}}$ to first order then produces

$$\mathbf{U}_{\mathrm{op}}(r) \rightarrow \mathbf{U}_{\mathrm{op}}(r) + \varepsilon\tilde{\mathbf{U}}_{\mathrm{op}}(r), \tag{4}$$

which generates perturbations in the PE field function. To first order, Eq. (2) is then modified to

[a)]Currently on sabbatical at the Royal Netherlands Naval College, Den Helder, Netherlands; electronic mail: kbsmith@nps.edu

$$\begin{aligned}
&\boldsymbol{\psi}(r+\Delta r)+\varepsilon\tilde{\boldsymbol{\psi}}(r+\Delta r)\\
&\quad=e^{-i\Delta rk_0\mathbf{H}_{\mathrm{op}}(r)}e^{-i\Delta rk_0\varepsilon\tilde{\mathbf{U}}_{\mathrm{op}}(r)}[\boldsymbol{\psi}(r)+\varepsilon\tilde{\boldsymbol{\psi}}(r)]\\
&\quad\approx e^{-i\Delta rk_0\mathbf{H}_{\mathrm{op}}(r)}[1-i\Delta rk_0\varepsilon\tilde{\mathbf{U}}_{\mathrm{op}}(r)]\boldsymbol{\psi}(r)\\
&\quad+\varepsilon e^{-i\Delta rk_0\mathbf{H}_{\mathrm{op}}(r)}\tilde{\boldsymbol{\psi}}(r). \qquad (5)
\end{aligned}$$

Note that Eq. (5) assumes that $\mathbf{H}_{\mathrm{op}}$ and $\tilde{\mathbf{U}}_{\mathrm{op}}$ commute. As long as the range step is small, and/or the perturbations have a weak range dependence, this formulation is justified. The elements of the diagonal matrix $\tilde{\mathbf{U}}_{\mathrm{op}}(r)$ are now simply $\tilde{\mathbf{n}}_j(r)$, the perturbations of the index of refraction profile.

Equation (5) provides the expression for the tangent linear model of the PE/SSF algorithm. Specifically, the perturbed field is computed from

$$\tilde{\boldsymbol{\psi}}(r+\Delta r)=e^{-i\Delta rk_0\mathbf{H}_{\mathrm{op}}(r)}[\tilde{\boldsymbol{\psi}}(r)-i\Delta rk_0\tilde{\mathbf{U}}_{\mathrm{op}}(r)\boldsymbol{\psi}(r)]. \qquad (6)$$

Introducing the matrix notation $\mathbf{F}_n=e^{-i\Delta rk_0\mathbf{H}_{\mathrm{op}}(r_n)}$ and $\mathbf{D}_n=i\Delta rk_0\tilde{\mathbf{U}}_{\mathrm{op}}(r_n)$, Eq. (6) can be rewritten in the more compact form

$$\tilde{\boldsymbol{\psi}}_{n+1}=\mathbf{F}_n(\tilde{\boldsymbol{\psi}}_n-\mathbf{D}_n\boldsymbol{\psi}_n). \qquad (7)$$

Note that this shows how the perturbed field in range step $n+1$ is created by two physical mechanisms: (1) The perturbed field in step n (computed from the previous step) is propagated forward through the mean environment, and (2) the unperturbed field in step n is perturbed by local fluctuations in step n, *then* carried forward through the mean environment. The first mechanism may be considered a global effect; while the second is purely a local one. It is this expression, and the simple form of the associated operators, that provide some advantage in using the SSF approach. Each physical mechanism can easily be studied independently, for example, to ascertain whether initial short-range perturbations in the field will propagate throughout the region and dominate the signal uncertainty, or whether it is a long-range environmental fluctuation that rapidly creates a signal mismatch. Order of magnitude estimates of the impact of each type of mechanism can easily be constructed from Eq. (7).

Since $\mathbf{F}_n^{-1}=\mathbf{F}_n^*$, then

$$\mathbf{F}_n^*\tilde{\boldsymbol{\psi}}_{n+1}=\tilde{\boldsymbol{\psi}}_n-\mathbf{D}_n\boldsymbol{\psi}_n. \qquad (8)$$

Using the trick of Hursky *et al.*,[1] we swap elements of the diagonal matrix $\mathbf{D}_n$ with elements of vector $\boldsymbol{\psi}_n$ to produce

$$\mathbf{F}_n^*\tilde{\boldsymbol{\psi}}_{n+1}=\tilde{\boldsymbol{\psi}}_n-\boldsymbol{\Gamma}_n\mathbf{u}_n, \qquad (9)$$

where $\boldsymbol{\Gamma}_n$ has diagonal elements corresponding to $\boldsymbol{\psi}_n$, and $\mathbf{u}_n=i\Delta rk_0\tilde{U}_{\mathrm{op}}(z,r_n)$. Introducing $\mathbf{G}_n=\mathbf{F}_n\boldsymbol{\Gamma}_n$ (which should be diagonal with elements corresponding to $\boldsymbol{\psi}_{n+1}$), we obtain the tangent linear model in the same form as Hursky *et al.*,[1] i.e.,

$$\tilde{\boldsymbol{\psi}}_{n+1}=\mathbf{F}_n\tilde{\boldsymbol{\psi}}_n-\mathbf{G}_n\mathbf{u}_n. \qquad (10)$$

It is the tangent linear model which runs the forward model "engine" within the adjoint method of inversion. The remaining higher-level processing used for data inversion can follow the approach of Hursky *et al.*[1] or others (e.g., Ref. 2).

III. CONCLUSIONS

In this letter, a brief derivation has been presented to illustrate how one may define the tangent linear model for propagation of acoustic uncertainty utilizing a SSF computational algorithm. While this is not expected to produce any unexpected results or significant differences from the finite difference algorithms employed in previous work, it does provide the necessary expressions for use with a SSF/PE model. This letter will then serve as a reference for such work.

This author has neither yet implemented these expressions nor attempted to utilize any tangent linear model for environmental inversions, but expects to study phenomena associated with forward propagation of acoustic uncertainty in the future. As stated, the key advantage of this approach is the simple evaluation of local versus global effects on signal uncertainty/variability. Such knowledge could potentially be used in a tactical sense, whereby an analysis of the local propagation of uncertainty is used to affect operating parameters (source/receiver depth, frequency, geographical placement, etc.) in order to minimize the combined (local plus global) impact of long-range acoustic variability.

ACKNOWLEDGMENTS

The author wishes to thank the reviewers for their comments, and the many useful discussions with Dr. Paul Hursky (HLS Research, Inc.), Vincent van Leijen (RNLNC), and Professor Jean-Pierre Hermand and Matthias Meyer (ULB). This work was supported by the Office of Naval Research, Code 321OA, and by sabbatical support from the Naval Postgraduate School in Monterey, CA and the Royal Netherlands Naval College in Den Helder, Netherlands.

[1] P. Hursky, M. B. Porter, B. D. Cornuelle, W. S. Hodgkiss, and W. A. Kuperman, "Adjoint modeling for acoustic inversion," J. Acoust. Soc. Am. **115**, 607–619 (2004).
[2] M. Matthias and J.-P. Hermand, "Optimal nonlocal boundary control of the wide-angle parabolic equation for inversion of a waveguide acoustic field," J. Acoust. Soc. Am. **117**, 2937–2948 (2005).
[3] F. D. Tappert, "The parabolic approximation method," in *Lecture Notes in Physics, Wave Propagation and Underwater Acoustics* Vol. **70**, edited by J. B. Keller and J. S. Papadakis (Springer, New York, 1997), pp. 224–287.
[4] K. B. Smith, "Convergence, stability, and variability of shallow water acoustic predictions using a split-step Fourier parabolic equation model," J. Comput. Acoust. **9**, 243–285 (2001).
[5] R. H. Hardin and F. D. Tappert, "Applications of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations," SIAM Rev. **15**, 423 (1973).
[6] D. J. Thomson and N. R. Chapman, "A wide-angle split-step algorithm for the parabolic equation," J. Acoust. Soc. Am. **74**, 1848–1854 (1983).

Reply to comment on "Auditory-nerve first-spike latency and auditory absolute threshold: A computer model" (L)

Ray Meddis[a)]
Department of Psychology, Essex University, Colchester, CO4 3SQ, United Kingdom

(Received 21 March 2006; revised 5 June 2006; accepted 9 June 2006)

Krisha [J. Acoust. Soc. Am., in press (2006)] has commented that an explanation based on presynaptic calcium accumulation at the inner hair cell is an incorrect explanation for the success of a model of the auditory periphery [Meddis, R., J. Acoustic. Soc. Am. **119**, 406–417 (2006)] in explaining data on first-spike auditory nerve latency. This reply accepts the criticism and accepts the strength of an alternative explanation based on expected latencies in random sequences of low-probability events. This reply also goes on briefly to explore the application of this argument to other phenomena, including the dependence of absolute auditory threshold on the duration of the stimulus. This has wide-ranging implications for the concept of "temporal integration" in psychophysics. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2221413]

PACS number(s): 43.64.Bt [BLM] Pages: 1192–1193

In his comment, Krishna (2006) questions my interpretation of how a model of the auditory periphery (Meddis, 2006) was able to simulate the auditory-nerve (AN) first-spike latency data of Heil and Neubauer (2001). The model in question simulated a speculation of Heil and Neubauer (2001), where delays between the onset of a tone stimulus and the first driven AN spike might be explained by the time taken to accumulate presynaptic calcium in inner hair cells (IHCs). A computer model of the auditory periphery, which included a specific account of presynaptic calcium, was able to simulate first-spike latency data. In particular, the model was able to simulate latencies, as long as 100 ms, to stimuli at threshold intensities despite having calcium integration time constants in the submillisecond range. The conclusion was drawn that presynaptic calcium dynamics were a sufficient explanation of the result.

Krishna (2006) claims that this explanation is unsatisfactory, and that a simpler statistical explanation of the same phenomenon is to be preferred. He demonstrates that the integration of calcium in the Meddis' model (2006) reaches its highest steady-state level within a few milliseconds of the onset of a stimulus, even though the latency of response is expected to be much longer. This is a convincing objection. It is a reasonable conclusion, therefore, that a alternative explanation for the success of the model is required.

Krishna then draws attention to an alternative, and much simpler, explanation based on the statistics of low probability events. Put simply, this argument says that low probability events, by virtue of their scarcity, are necessarily characterized by long delays between each event. The expected delay to the first event in a randomly distributed series of events is the inverse of its probability. At threshold, it is assumed that the release of transmitter into the IHC/AN cleft is a rare event, and that the delay between the onset of the stimulus on the first event will necessarily be longer for lower probabilities. This simple and elegant explanation was implicit in his earlier publication (Krishna, 2002), and was missed by the present author. The same idea was proposed much earlier to account for visual thresholds as a function of duration (Watson, 1969). It is also similar in some respects to the theory of "temporal integration and multiple looks" by Viemeister and Wakefield (1991). Heil and Neubauer (2003), in the discussion section of a more recent paper, also propose a similar idea.

If, as seems reasonable, the statistical account of first-spike latencies is accepted, it has ramifications well beyond those indicated in Krishna's comment. For example, Heil and Neubauer (2001) discussed their first-spike latency data in the context of the auditory receptor as an "energy integrator." This is based on the many observations that sentient biological systems are likely to detect longer, rather than shorter, stimuli when presented at the same intensity in the threshold region. The standard explanation for this phenomenon is that the organism integrates the energy of the stimulus, in some way, until some critical level is reached and the detection is made. Heil and Neubauer (2001) originally presented their results as a challenge to the idea that stimulus energy is being integrated over time. They argued that their data were more consistent with the idea that pressure was the integrated quantity.

The statistical account, however, suggests that nothing is being integrated. On the contrary, it affirms that longer stimuli are more likely to be detected simply because a longer stimulus provides a longer time window in which a rare stimulus-triggered event might occur. This is a fundamentally different interpretation from the "energy integration" notion, and deserves careful consideration because of its wide implications across the biological spectrum of sentient organisms.

Heil and Neubauer (2003) have, themselves, moved on to a more statistical approach and widened the scope of the discussion to include psychophysically measured thresholds by indicating parallels with both AN and cortical first-spike latencies. Their link between AN first-spike latency and absolute threshold was the motivation for the final evaluation

a)Electronic mail: rmeddis@essex.ac.uk

of the peripheral model in Meddis (2006). Here, it was explained that the model showed the expected relationship between absolute threshold and the duration of a stimulus. At least for the shorter stimuli, there was a close parallel between the model and the psychophysical data. The (unspoken) implication was that calcium accumulation was also a candidate explanation for this phenomenon. However, Krishna's comments suggest that this might be another inappropriate conclusion. In the computer model (Meddis, 2006), some integration was taking place at the level of calcium accumulation at the AN synapse, but this is on a much shorter timescale than required by the psychophysical data. Krishna's comments point the way to a more radical theory of the dependence of auditory thresholds on stimulus duration.

Unfortunately, in auditory psychophysics, the term "temporal integration" has now become synonymous with the phenomenon of lower thresholds for longer stimuli. This is an unfortunate example of how the name of a putative mechanism is used to indicate a phenomenon that it might (or might not) explain. The name has become so entrenched in conventional thinking that it becomes difficult to consider the possibility that any alternative mechanism might be involved; i.e., "temporal integration does not require a temporal integrator!" However, that is exactly what needs to be considered.

This alternative view can be stated very simply as follows. A near-threshold stimulus of steady intensity establishes a (low) probability of a neural event whose occurrence will signal the presence of the stimulus. When the probability of occurrence is low, it is possible that the event will not occur before the end of the tone, and the stimulus will go undetected. The longer the stimulus, the greater the chance of detection. When the stimulus is more intense, the probability of the event will be greater and the duration of a detectable stimulus can be shorter. Stimulus thresholds are lower for longer stimuli and no integration is required.

It is not appropriate here to debate, at length, whether or not this theory is the correct one. However, the case is made that the relevance of the statistical inevitability of the association of longer delays with less probable events should be taken seriously, as indicated in Krishna's letter. At least for this author, this applies not only to discussions of first-spike latencies but also to the relationship between auditory threshold and stimulus duration. It also implies that we should abandon the habit of using the mechanism-implying descriptor of temporal integration to refer to the empirical phenomenon of lower thresholds for longer stimuli.

Heil, P., and Neubauer, H. (**2001**). "Temporal integration of sound pressure determines thresholds of auditory-nerve fibers," J. Neurosci. **21**, 7404–7415.

Heil, P., and Neubauer, H. (**2003**). "A unifying basis of auditory thresholds based on temporal summation," Proc. Natl. Acad. Sci. U.S.A. **100**, 6151–6156.

Krishna, B. S. (**2002**). "A unified mechanism for spontaneous-rate and first-spike timing in the auditory nerve," J. Comput. Neurosci. **13**, 71–91.

Krishna, B. S. (**2006**). "Comment on 'Auditory nerve first-spike latency and auditory absolute threshold: A computer model'," J. Acoust. Soc. Am., in press.

Meddis, R. (**2006**). "Auditory-nerve first-spike latency and auditory absolute threshold: A computer model." J. Acoust. Soc. Am. **119**, 406–417.

Viemeister, N. F. and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–856.

Watson, A. B. (1969). "Probability summation over time," Vision Res. **19**, 515–522.

Energy exchange in uncorrelated ray fields of vibroacoustics

Alain Le Bot[a)]
Laboratoire de Tribologie et Dynamique des Systèmes, CNRS, École Centrale de Lyon, 36, Avenue Guy de Collongue 69134 Ecully, France

(Received 31 May 2005; revised 29 March 2006; accepted 22 June 2006)

This paper is concerned with the basic equations governing energy and intensity in incoherent ray fields. Some fictitious sources are distributed on the boundary of the domain but also on diffracting wedges and peaks. Their powers are determined by some appropriate boundary integral equations. Once these powers are known, energy and intensity inside the domain are given by a simple superposition of contributions of these sources. All paths of propagation are taken into account including direct, reflected, refracted, transmitted, and diffracted rays, but also, radiation by surface, edge or corner modes, and the reciprocal paths for structural response. This theory unifies several fields from the "radiosity method" in room acoustics which determines the reverberation time to the "radiative transfer method" in structural dynamics which gives the repartition of vibrational energy inside subsystems of built-up structures. This is therefore a candidate for an alternative to statistical energy analysis when fields are nondiffuse.
[DOI: 10.1121/1.2227372]

PACS number(s): 43.20.Dk, 43.20.El, 43.20.Tb, 43.40.Rj [MO] Pages: 1194–1208

List of Symbols

$\gamma_0=2,\ 2\pi$ or 4π	solid angle of space of n-dimensional space $n=1,\ 2$, or 3
$\gamma=\int\cos\theta du=1,\ 2$ or π	hemispherical integral of $\cos\theta$ for $n=1,\ 2$ or 3
Ω	domain
Γ	regular boundary
Δ	edge of boundary
Υ	vertices of boundary
$\Gamma^0,\Delta^0,\Upsilon^0$	set of directions from Γ, Δ or Υ to $\mathbf{p}$
$\mathbf{s},\mathbf{r}$	source point, receiver point
$\mathbf{p},\mathbf{q}$	reflection or diffraction points
$\mathbf{u},\theta,\alpha$	emission direction, emission angles
$\mathbf{v},\varphi,\beta$	incidence direction, incidence angles
$\delta(\mathbf{r}),\delta(t),\delta_{\Delta^0}(\mathbf{v})$	Dirac function of space, time, unit sphere with support Δ^0
c,c',c_0	group speed, phase speed, sound speed
m	attenuation factor
$I(\mathbf{r},\mathbf{u},t)$	radiative intensity
$W(\mathbf{r},t)$	energy density
$\mathbf{I}(\mathbf{r},t)$	intensity vector
G	energy density of direct field
$\mathbf{H},H$	intensity of direct field, magnitude
$R(\mathbf{v},\mathbf{u}),R(\mathbf{v})$	bidirectional reflectivity, hemispherical reflectivity
$R_{ji}(\mathbf{v},\mathbf{u}),R_{ji}(\mathbf{v})$	bidirectional transmittivity, hemispherical transmittivity
$D(\mathbf{v},\mathbf{u})$	bidirectional diffractivity
A_{sa},A_0	radiation coefficient, forcing coefficient

[a)]Electronic mail: alain.le-bot@ec-lyon.fr

D_{sa}, D_{as}	diffractivity for radiation, for structural response
ρ	power density of volume sources
σ	power density of surface sources
λ	power density of line sources
μ	power of point sources

I. INTRODUCTION

The use of energy as a primary variable to describe vibrational fields is an idea widely spread in high frequency modeling. When waves are uncorrelated, the linear superposition applies on energy allowing the summation of energy of individual waves. In addition, if fields are assumed to be diffuse, the application of power balance leads to some simple equations.

In room acoustics, Sabine's formula and other related relationships from Millington and Eyring[1] give the reverberation time in terms of absorption of walls. This is quite a convincing example of the power of the energy approach. When the diffuse field assumption is no longer valid, geometrical acoustics is an alternative. The ray-tracing technique allows us to compute impulse responses in any room and therefore gives the reverberation time as well as many other acoustical criteria. The ray-tracing technique is however a numerical method (very efficient in practice) which is sometimes not convenient for a theoretical purpose. In case of diffusely reflecting walls and applying the factor view method. Kuttruff[2] has derived an integral equation on reverberation time valid for rooms of arbitrary shape. As Sabine's formula, this equation stems from the power balance, rays are again uncorrelated but now the field may be not diffuse.

In vibroacoustics, statistical energy analysis (SEA) (Ref. 3) also starts from diffuse fields. Systems are divided into subsystems in which the diffuse field assumption holds. SEA gives the steady-state energy levels whereas transient SEA predicts the decrease of energy after sources have been switched off. SEA is entirely based on the application of power balance and does not express anything else than power balance. (A very interesting attempt to introduce the second principle of thermodynamics in SEA is given in Ref. 4.) A simple linear equation is obtained which relates injected power and energy levels and whose parameters are damping and coupling loss factors. In this context, the present study aims to avoid the diffuse field assumption in SEA. We attempt to generalize SEA in the same way than Kuttruff's integral equation on reverberation time generalizes Sabine's formula. Rays are always assumed to be uncorrelated and the exchange of energy between two points of the boundary requires a factor view but also some energy conversion factors which generalize the coupling loss factors in SEA.

The outline of the paper is as follows: In Sec. II, fictitious sources for reflection and diffraction are introduced. Energy density and intensity of the ray field are then given in terms of these sources. In Sec. III, the general equation for the reflection sources is derived and then, the particular cases of diffuse and specular reflections are presented. In Sec. IV, these results are generalized to refraction and transmission. In Sec. V, the equations for diffraction sources are derived in the cases of wedge and peak. In Sec. VI, all previous equations are applied to sound radiated by surface, edge, and corner modes. Finally, structural response is tackled in Sec. VII.

II. ENERGY OF RAY FIELDS

Geometrical acoustics and its straightforward generalization, Geometrical Theory of Diffraction[5] are the natural framework for defining and describing rays. Many vibrational fields may be described in terms of rays including optics indeed but also acoustics, vibration of structures such as beams, plates, shells, and so on. For transient problems, the concept of wave packet or sound particle is more appropriate. For instance, it is usual in room acoustics to compute the impulse response with a ray-tracing algorithm by following the trajectory of sound particles emitted at an initial time. For the steady-state problem considered as being a particular case of transient problem, rays may be viewed as stationary flow of sound particles. In what follows, the term ray is employed for both steady-state and transient cases.

The energy density of the ray field is noted as W, whereas intensity, defined as the power per unit surface normal to the ray, is noted as $\mathbf{I}$. Energy density and intensity are always the sum of energies and intensities of individual rays. Interference effects are neglected considering that rays are uncorrelated. This is the main difference with Geometrical Theory of Diffraction where a phase is attached to rays.

Let consider a source point $\mathbf{s}$ in an n-dimensional space ($n=1$, 2, or 3). After an impulse at time τ, sound particles are provided and move away from the source $\mathbf{s}$. Energy density of this spherical ($n=3$), cylindrical ($n=2$), or plane ($n=1$) wave is noted as G while intensity is noted as $\mathbf{H}$. Expressions for this direct field are,

$$G(\mathbf{s},\tau;\mathbf{r},t) = G(\mathbf{s},\mathbf{r})\,\delta(t-\tau-r/c), \tag{1}$$

$$\mathbf{H}(\mathbf{s},\tau;\mathbf{r},t) = \mathbf{H}(\mathbf{s},\mathbf{r})\,\delta(t-\tau-r/c), \tag{2}$$

where $r=|\mathbf{s}-\mathbf{r}|$ is the source-receiver distance. $G(\mathbf{s};\mathbf{r})$ and $\mathbf{H}(\mathbf{s};\mathbf{r})$ are the stationary fields given by

$$G(\mathbf{s},\mathbf{r}) = \frac{e^{-mr}}{\gamma_0 c r^{n-1}} V(\mathbf{s},\mathbf{r}), \tag{3}$$

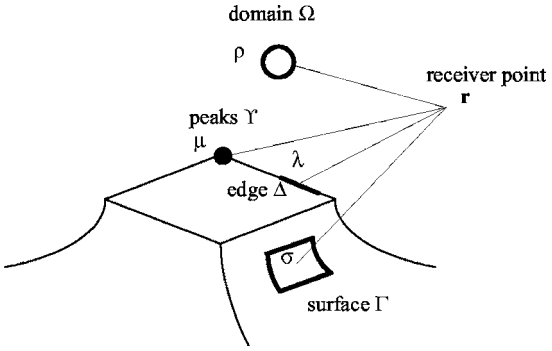


FIG. 1. Actual and fictive sources contributing to the energy at point **r**. Volume sources ρ in the domain Ω stand for actual noise sources, surface sources σ on the regular boundary Γ stand for reflection, line sources λ on edges Δ of boundary and point sources μ on singular points Υ of boundary stand for diffraction.

$$\mathbf{H}(\mathbf{s},\mathbf{r}) = cG(\mathbf{s},\mathbf{r})\mathbf{u}, \tag{4}$$

where $\gamma_0=2$, 2π or 4π is the solid angle of n-dimensional space, c is the group speed, m is the attenuation factor of the medium, and $\mathbf{u}$ is the unit vector from $\mathbf{s}$ to $\mathbf{r}$. $V(\mathbf{s},\mathbf{r})$ is the visibility function whose value is zero if an obstacle blocks the path between the source point $\mathbf{s}$ and the receiver point $\mathbf{r}$, and one otherwise. In some equations, $H=cG$ will denote the magnitude of vector $\mathbf{H}$.

Equations (1) and (2) verify the power balance with an impulse excitation,

$$\mathbf{div}_{\mathbf{r}}\,\mathbf{H} + mcG + \frac{\partial G}{\partial t} = \delta(\mathbf{r}-\mathbf{s})\,\delta(t-\tau). \tag{5}$$

Equations (1) and (2) are found to be the unique outgoing solution of Eq. (5).[6] Similarly, the power balance in the steady state condition for Eqs. (3) and (4) is[7]

$$\mathbf{div}_{\mathbf{r}}\,\mathbf{H} + mcG = \delta(\mathbf{r}-\mathbf{s}). \tag{6}$$

When considering a ray field in a domain Ω of three-dimensional space, Ω may be bounded or not, energy fields W and $\mathbf{I}$ result from a linear superposition of direct fields stemming from volume sources with power density ρ (W/m^3) located inside domain Ω, from surface sources with power density σ(W/m^2) located on the regular boundary Γ, from some line sources λ(W/m) on the set Δ of diffracting edges of the boundary and also from some point sources with power μ (W) on the set Υ of singular points of the boundary (Fig. 1). ρ are generally physical noise sources, whereas σ is associated with reflection on boundary, radiation by structure or transmission through walls, λ is related to diffraction by wedges, and μ to diffraction by peaks. In two-dimensional space (structure), ρ (W/m^2) is the power per unit surface of driving forces or acoustical pressure, σ is the power per unit length (W/m) of reflection sources or sources transmitted through edges common with other structural elements, and μ is the power (W) of diffracting points such as corners, driving points, small holes, rivets, bolts or any other singular points. No line sources λ are considered in structures. By summing the energy contributions of all these sources, energy density and intensity at any receiver point $\mathbf{r}$ are found to be,

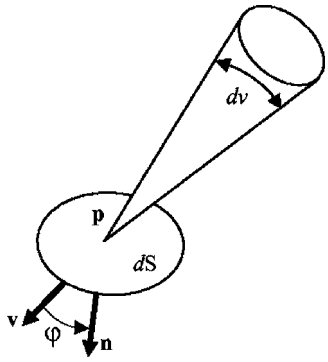


FIG. 2. Radiative intensity at point $\mathbf{p}$ in direction $\mathbf{v}$ crossing the surface dS. $\mathbf{n}$ is the normal to dS and φ the incidence angle.

$$\begin{aligned} W(\mathbf{r},t) = &\int_{\Omega} \rho(\mathbf{s},t-r/c)G(\mathbf{s},\mathbf{r})d\Omega_{\mathbf{s}} \\ &+ \int_{\Gamma} \sigma(\mathbf{p},\mathbf{v},t-r/c)G(\mathbf{p},\mathbf{r})d\Gamma_{\mathbf{p}} \\ &+ \int_{\Delta} \lambda(\mathbf{p},\mathbf{v},t-r/c)G(\mathbf{p},\mathbf{r})d\Delta_{\mathbf{p}} \\ &+ \sum_{\mathbf{p}\in\Upsilon} \mu(\mathbf{p},\mathbf{v},t-r/c)G(\mathbf{p},\mathbf{r}), \end{aligned} \tag{7}$$

$$\begin{aligned} \mathbf{I}(\mathbf{r},t) = &\int_{\Omega} \rho(\mathbf{s},t-r/c)\mathbf{H}(\mathbf{s},\mathbf{r})d\Omega_{\mathbf{s}} \\ &+ \int_{\Gamma} \sigma(\mathbf{p},\mathbf{v},t-r/c)\mathbf{H}(\mathbf{p},\mathbf{r})d\Gamma_{\mathbf{p}} + \int_{\Delta} \lambda(\mathbf{p},\mathbf{v},t \\ &- r/c)\mathbf{H}(\mathbf{p},\mathbf{r})d\Delta_{\mathbf{p}} + \sum_{\mathbf{p}\in\Upsilon} \mu(\mathbf{p},\mathbf{v},t-r/c)\mathbf{H}(\mathbf{p},\mathbf{r}), \end{aligned} \tag{8}$$

where $r=|\mathbf{s}-\mathbf{r}|$ or $|\mathbf{p}-\mathbf{r}|$ is the source-receiver distance and $\mathbf{v}$ the unit vector from $\mathbf{p}$ to $\mathbf{r}$. In these integrals and all subsequent ones, the term $\int\lambda G d\Delta$ is cancelled in dimension two. A local power balance applies inside Ω for fields W and $\mathbf{I}$,

$$\mathbf{div}\,\mathbf{I} + mcW + \frac{\partial W}{\partial t} = \rho, \tag{9}$$

where mcW is the power density being dissipated and ρ is the power density being injected.

Another concept useful to describe ray fields is the *radiative intensity*[8] $I(\mathbf{p},\mathbf{v},t)$ also called *specific intensity*.[9] Consider a point $\mathbf{p}$ on the infinitesimal surface dS and an infinitesimal solid angle dv about $\mathbf{v}$. The angle between $\mathbf{n}$ and $\mathbf{v}$ is noted φ (Fig. 2). The radiative intensity is the power per unit solid angle and per unit area normal to the ray,

$$I(\mathbf{p},\mathbf{v},t) = \frac{1}{\cos\varphi}\frac{d\mathcal{P}}{dS\,dv}. \tag{10}$$

Integration of radiative intensity over all directions gives the intensity vector,

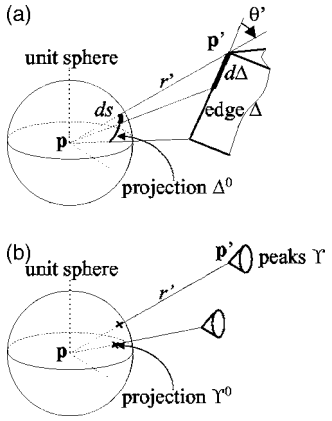


FIG. 3. (a) The set Δ^0 is the projection of Δ onto the unit sphere centered on $\mathbf{p}$. It is a curve with curvilinear abscissa s. The emission angle θ' is measured with the tangent to Δ. (b) The set Υ^0 is the projection of the diffracting peaks Υ onto the unit sphere. This is a discrete set.

$$\mathbf{I}(\mathbf{p},t)=\int I(\mathbf{p},\mathbf{v},t)\mathbf{v}dv. \tag{11}$$

All volume, surface, line, and point sources located inside the incident cone dv contribute to the radiative intensity. Therefore,

$$I(\mathbf{p},\mathbf{v},t)dv=\int \rho H d\Omega+\sigma H d\Gamma+\lambda H d\Delta+\sum \mu H, \tag{12}$$

where the sources ρ, σ, λ, μ of the right-hand side are those located inside the incident cone.

Let us begin by developing the first integral of Eq. (12). In spherical coordinates $(r,\mathbf{v})$ centered on $\mathbf{p}$, the infinitesimal volume is $d\Omega=r^{n-1}drdv$. With Eqs. (3) and (4), the first integral of Eq. (12) then reads $dv\int \rho e^{-mr}dr/\gamma_0$ where the integration is performed over the line beginning at $\mathbf{p}$ and with direction $-\mathbf{v}$. When Ω, is unbounded in direction $-\mathbf{v}$, no boundary sources σ,λ,μ contribute to the radiative intensity and the right-hand side of Eq. (12) reduces to its first integral. The radiative intensity is therefore,

$$I(\mathbf{p},\mathbf{v},t)=\frac{1}{\gamma_0}\int_{\mathbf{p}}^{\mathbf{p}-\infty.\mathbf{v}}\rho(\mathbf{s},t-r/c)e^{-mr}dr. \tag{13}$$

But when Ω is bounded in the direction $-\mathbf{v}$, the line beginning at $\mathbf{p}$ encounters the boundary at point $\mathbf{p}'$. Depending on the position of $\mathbf{p}'$ on Γ, Δ or Υ, a single term among the last three terms of Eq. (12) survives.

Before expanding them, let us introduce the set Γ^0 (resp. Δ^0 and Υ^0) defined as the set of unit vectors pointing from Γ (resp. Δ and Υ) to $\mathbf{p}$. These are subsets of the unit sphere centered in $\mathbf{p}$. The characteristic function χ_{Γ^0} is defined by $\chi_{\Gamma^0}(\mathbf{v})=1$ if $\mathbf{v}\in\Gamma^0$ and $\chi_{\Gamma^0}(\mathbf{v})=0$ otherwise. The set Δ^0 is the projection of Δ onto the unit sphere. Δ^0 is a curve, or union of curves, plotted on the unit sphere. Its curvilinear abscissa is noted as s [Fig. 3(a)]. The Dirac function δ_{Δ^0} is defined such as $\int \delta_{\Delta^0} f dv=\int_{\Delta^0} f ds$ for any function f of the unit sphere. It is said to have the support is Δ^0 because $\int \delta_{\Delta^0} f dv=0$ if f does not encounter Δ^0, that is $f(\mathbf{v})=0$ when $\mathbf{v}\in\Delta^0$. Similarly, Υ^0 is the projection of Υ onto the unit sphere. Υ^0 is a discrete set [Fig. 3(b)]. The Dirac function δ_{Υ^0} with support Υ^0 is defined such as $\int \delta_{\Upsilon^0} f dv=\Sigma_{\Upsilon^0} f$ any function f of the unit sphere. The support is Υ^0 meaning that $\int \delta_{\Upsilon^0} f dv=0$ whenever $f(\mathbf{v})=0$ if $\mathbf{v}\in\Upsilon^0$.

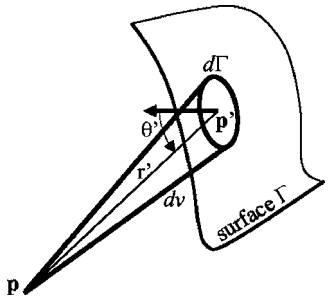


FIG. 4. Relation between the cone angle dv and the part $d\Gamma$ of the boundary enclosed in the cone. θ' is the emission angle measured with the normal to the surface and r' is the source-receiver distance.

First, let us assume that $\mathbf{p}'\in\Gamma$ and let develop the second term of the right-hand side of Eq. (12). The condition $\mathbf{p}'\in\Gamma$ is equivalent to $\mathbf{v}\in\Gamma^0$ and therefore $\chi_{\Gamma^0}(\mathbf{v})=1$. The infinitesimal surface $d\Gamma$ enclosed in the cone is $d\Gamma=r'^{n-1}\chi_{\Gamma^0}dv/\cos\theta'$ (Fig. 4), where θ' is the emission angle at $d\Gamma$ toward $\mathbf{p}$ and $r'=|\mathbf{p}'-\mathbf{p}|\cdot\ \theta'$ is measured between the emission direction and the normal to the surface $d\Gamma$. This normal makes sense since $\mathbf{p}'$ is assumed to be regular. This relationship is trivial when $n=1$. The second term of the right-hand side of Eq. (12) then reads $dv\sigma e^{-mr'}\chi_{\Gamma^0}/(\cos\theta'\gamma_0)$.

Secondly, when $\mathbf{p}'\in\Delta(n=3)$, the length $d\Delta$ is related to its projection ds on the unit sphere by $d\Delta=r'ds/\sin\theta'$ where $r'=|\mathbf{p}'-\mathbf{p}|$ and θ' is the emission angle now measured between the tangent to Δ and $\mathbf{v}$ [Fig. 3(a)]. But the Dirac function δ_{Δ^0} verifies $\int \delta_{\Delta^0}dv=\int_{\Delta^0}ds$ and then $ds=\delta_{\Delta^0}dv$. The infinitesimal length is therefore $d\Delta=r'\delta_{\Delta^0}dv/\sin\theta'$. The third term of the right-hand side of Eq. (12), existing in the only case $n=3$, becomes $dv\lambda e^{-mr'}\delta_{\Delta^0}/(r'\sin\theta'\gamma_0)$.

Finally, when $\mathbf{p}'\in\Upsilon$, the last sum of Eq. (12) may be written $\Sigma\mu H=\delta_{\Upsilon^0}\mu H dv$. This is just a particular case of the definition equation of the function δ_{Υ^0}. $\delta_{\Upsilon^0}dv$ is the number of points of Υ enclosed in cone dv [Fig. 3(b)]. The fourth term of the right-hand side of Eq. (12) is therefore $dv\mu e^{-mr'}\delta_{\Upsilon^0}/(r'^{n-1}\gamma_0)$.

When Ω is bounded in direction $-\mathbf{v}$, Eq. (12) reads,

$$\begin{aligned}I(\mathbf{p},\mathbf{v},t)=\frac{1}{\gamma_0}\Bigg[&\int_{\mathbf{p}}^{\mathbf{p}'}\rho(\mathbf{s},t')e^{-mr}dr\\&+\frac{\sigma(\mathbf{p}',\mathbf{v},t')}{\cos\theta'}e^{-mr'}\chi_{\Gamma^0}(\mathbf{v})\\&+\frac{\lambda(\mathbf{p}',\mathbf{v}',t')}{\sin\theta'}\frac{e^{-mr'}}{r'}\delta_{\Delta^0}(\mathbf{v})\\&+\mu(\mathbf{p}',\mathbf{v},t')\frac{e^{-mr'}}{r'^{n-1}}\delta_{\Upsilon^0}(\mathbf{v})\Bigg],\end{aligned} \tag{14}$$

where $t'=t-r/c$ is the time delayed by the flight duration from the source to the receiver. In this equation, the functions χ_{Γ^0}, δ_{Δ^0}, and δ_{Υ^0} have disjoint supports. It means that depending on the position of the point $\mathbf{p}'$ on the boundary,

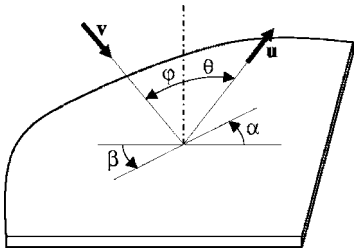


FIG. 5. Reflection on plane surface for incidence direction $\mathbf{v}=\varphi,\beta$ and reflection direction $\mathbf{u}=\theta,\alpha$. Elevation angles φ, θ are measured with the normal to the surface and azimuthal angles β, α are measured in the plane of the surface.

only one function among $\chi_{\Gamma^0}(\mathbf{v})$, $\delta_{\Delta^0}(\mathbf{v})$, $\delta_{\Upsilon^0}(\mathbf{v})$ is not null. For instance, if $\mathbf{p}'\in\Delta$ or equivalently $\mathbf{v}\in\Delta^0$, then $\mathbf{p}'\notin\Gamma$ and $\mathbf{p}'\notin\Upsilon$ and therefore $\chi_{\Gamma^0}(\mathbf{v})=\delta_{\Upsilon^0}(\mathbf{v})=0$. The right-hand side of Eq. (14), can just have one nonvanishing term among the last three terms (the third term is always canceled in dimension two). For this reason, it is not disturbing that θ' has different definitions in Γ and Δ. Equation (14) gives the radiative intensity in direction $\mathbf{v}$ in terms of all sources, ρ,σ,λ, and μ located on the path with direction $-\mathbf{v}$.

III. REFLECTION

Reflection of rays on the boundary may occur in different manners. Two extreme situations are of interest: diffuse reflection and specular reflection. The former is often encountered in acoustics when walls are rough and the second situation stands for perfect mirror for instance a plane hard wall or a straight free edge for structural rays. The case of general reflectivity is first considered before both particular cases are detailed. In all cases, a fictitious source layer σ being the density of reflected power is introduced on the surface Γ. It is then determined by an appropriate equation deduced from the power balance.

Let the *bidirectional reflectivity*[8] $R(\mathbf{v},\mathbf{u})$ of the boundary be the ratio of radiative intensity $I(\mathbf{p},\mathbf{u},t)$ reflected in direction $\mathbf{u}$ and the incident flux $I(\mathbf{p},\mathbf{v},t)\cos\varphi$ from direction $\mathbf{v}$ with incidence φ. At any point $\mathbf{p}$ of the boundary, the leaving radiative intensity, $I=d\mathcal{P}/\cos\theta dSdu$ in direction $\mathbf{u}$ with emission angle θ, is the sum of all incident fluxes $I\cos\varphi$ times the reflectivity R. Angles φ, θ and directions $\mathbf{v}$, $\mathbf{u}$ are defined in Fig. 5. It yields

$$I(\mathbf{p},\mathbf{u},t)=\frac{1}{\cos\theta}\frac{d\mathcal{P}}{dSdu}=\int R(\mathbf{v},\mathbf{u})I(\mathbf{p},\mathbf{v},t)\cos\varphi dv. \quad (15)$$

This is the so-called *detailed power balance*[10] which gives the power in any direction $\mathbf{u}$ from contributions of other directions $\mathbf{v}$. Consider an infinitesimal surface source $\mathbf{p}$ of area $d\Gamma$ with power $d\Gamma\sigma$. The flux of intensity $d\Gamma\sigma H$ through the infinitesimal solid angle du is $d\Gamma du\sigma/\gamma_0$. The meaning of σ is now apparent, σ/γ_0 is the reflected power per unit area of boundary and unit solid angle. Since the area normal to the ray is $d\Gamma\cos\theta$, the radiative intensity leaving the source $\mathbf{p}$ in direction $\mathbf{u}$ is from Eq. (10),

$$\frac{1}{\cos\theta}\frac{d\mathcal{P}}{dSdu}=\frac{\sigma(\mathbf{p},\mathbf{u},t)}{\gamma_0\cos\theta}. \quad (16)$$

Substitution of Eqs. (12) and (16) into Eq. (15) gives,

$$\begin{aligned}\frac{\sigma(\mathbf{p},\mathbf{u},t)}{\gamma_0\cos\theta}=&\int_\Omega R(\mathbf{v},\mathbf{u})\rho(\mathbf{s},t')H(\mathbf{s},\mathbf{p})\cos\varphi d\Omega_{\mathbf{s}}\\&+\int_\Gamma R(\mathbf{v},\mathbf{u})\sigma(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})\cos\varphi d\Gamma_{\mathbf{q}}\\&+\int_\Delta R(\mathbf{v},\mathbf{u})\lambda(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})\cos\varphi d\Delta_{\mathbf{q}}\\&+\sum_{\mathbf{q}\in\Upsilon}R(\mathbf{v},\mathbf{u})\mu(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})\cos\varphi. \quad (17)\end{aligned}$$

This is an integral equation which gives σ at any point $\mathbf{p}$ and in any direction $\mathbf{u}$ in terms of other volume, surface, line, and point sources.

Let us define the *hemispherical reflectivity*[8] $R(\mathbf{v})$ be the total flux leaving the boundary, for any unit incident flux from $\mathbf{v}$,

$$R(\mathbf{v})=\int R(\mathbf{v},\mathbf{u})\cos\theta du, \quad (18)$$

where the integral runs over the hemisphere of inward directions. By multiplying Eq. (15) by $\cos\theta$ and integrating over du gives,

$$\int I(\mathbf{p},\mathbf{u},t)\cos\theta du=\int R(\mathbf{v})I(\mathbf{p},\mathbf{v},t)\cos\varphi dv. \quad (19)$$

It is then apparent that the hemispherical reflectivity $R(\mathbf{v})$ is the ratio of the reflected power and the incident power from direction $\mathbf{v}$. This is a non-negative number less than 1 sometimes called *reflection efficiency*[11] in structural wave literature. Some surfaces have the property of constant bidirectional reflectivity for any fixed incidence $\mathbf{v}$. In this case, the bidirectional directivity is given by the hemispherical directivity $R(\mathbf{v},\mathbf{u})=R(\mathbf{v})/\gamma$, where $\gamma=\int\cos\theta du=1$, 2 or π for $n=1$, 2 or 3. For such surfaces, the reflection is said to be *diffuse*. When R does not depend on $\mathbf{u}$, the right-hand side of Eq. (17) does not depend on $\mathbf{u}$, too. Thus, the power density $\sigma(\mathbf{p},\mathbf{u},t)$ has the directivity given by Lambert's law,

$$\sigma(\mathbf{p},\mathbf{u},t)=\sigma(\mathbf{p},t)\cos\theta. \quad (20)$$

Multiplying Eq. (17) by $\cos\theta$ and integrating over the hemisphere of all emission directions $\mathbf{u}$ leads to,

$$\begin{aligned}\frac{\gamma}{\gamma_0}\sigma(\mathbf{p},t)=&\int_\Omega R(\mathbf{v})\rho(\mathbf{s},t')H(\mathbf{s},\mathbf{p})\cos\varphi d\Omega_{\mathbf{s}}\\&+\int_\Gamma R(\mathbf{v})\sigma(\mathbf{q},t')\cos\theta' H(\mathbf{q},\mathbf{p})\cos\varphi d\Gamma_{\mathbf{q}}\\&+\int_\Delta R(\mathbf{v})\lambda(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})\cos\varphi d\Delta_{\mathbf{q}}\\&+\sum_{\mathbf{q}\in\Upsilon}R(\mathbf{v})\mu(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})\cos\varphi, \quad (21)\end{aligned}$$

where θ' is the emission angle at point $\mathbf{q}$. This is a Fredholm's integral equation of second kind on σ. The first two terms of this equation were first derived by Kuttruff[2,12] in the context of room acoustics. The equation was originally,

$$B(\mathbf{r},t) = B_0(\mathbf{r},t) + \int_S (1-\alpha) B\left(\mathbf{r}, t - \frac{R}{c}\right) K(\mathbf{r},\mathbf{r}')dS', \tag{22}$$

where $B=\sigma 1/(1-\alpha)$ is the irradiation density used as unknown, S is the enclosure, $1-\alpha$ is the reflection coefficient [noted as R in Eq. (21)], B_0 is the contribution of direct sources is [first integral in Eq. (21)], R the source-receiver distance, and $K=\cos\theta\cos\theta'/\pi R^2$. Assuming that $B(\mathbf{r},t) = B(\mathbf{r})e^{-\lambda t}$, Eq. (22) leads to an integral equation on reverberation time of rooms which applies beyond the validity of Sabine's formula and especially for rooms having atypical shapes. Some algorithms have been proposed for solving this integral equation[13–16] and even an original closed-form solution was found for spherical enclosures.[17] It was also numerically solved for early decaying of sound in Ref. 15, where it is also proved the existence and uniqueness of reverberation time. On the other hand, Eq. (22) is also useful to compute the SPL map in the steady-state condition[18–21] and then Eq. (22) is an alternative to the ray-tracing technique. More generally, Eq. (22) and its generalization to diffracting sources Eq. (21), embody all geometrical acoustics with diffuse reflecting surfaces.

In the case of specular reflection, the bidirectional reflectivity is given by,

$$R(\mathbf{v},\mathbf{u}) = R(\mathbf{v})\frac{\delta(\mathbf{v}-\mathbf{u}')}{\cos\varphi}, \tag{23}$$

where $\mathbf{u}'=\mathbf{u}-\mathbf{2}(\mathbf{u}\cdot\mathbf{n})\mathbf{n}$ is the incident direction which specularly reflects in $\mathbf{u}\cdot\mathbf{n}$ is the unit outward normal to the boundary. The hemispherical reflectivity is from Eq. (18), $\int R(\mathbf{v})\delta(\mathbf{v}-\mathbf{u}')\cos\theta/\cos\varphi du$, where $\mathbf{u}=\theta,\alpha$ and $\mathbf{v}=\varphi,\beta$ (Fig. 5). The change of variable $\mathbf{u}=\theta,\alpha\rightarrow\mathbf{u}'=\theta,\alpha+\pi$ gives $du'=du$ and the hemispherical reflectivity is therefore $\int R(\mathbf{v})\delta(\mathbf{v}-\mathbf{u}')\cos\theta/\cos\varphi du'=R(\mathbf{v})$. Substitution of Eq. (23) into Eq. (15) gives the equality $I(\mathbf{p},\mathbf{u}',t)=R(\mathbf{u}')I(\mathbf{p},\mathbf{u},t)$. For a perfect mirror $R(\mathbf{v})=1$ and therefore the incident radiative intensity $I(\mathbf{p},\mathbf{u}',t)$ equals the reflected radiative intensity $I(\mathbf{p},\mathbf{u},t)$. But for an absorbing mirror, $R(\mathbf{u}')$ is the ratio of reflected and incident powers. Expanding $I(\mathbf{p},\mathbf{u},t)$ with Eq. (16) and $I(\mathbf{p},\mathbf{u}',t)$ with Eq. (14), the above equality leads to

$$\begin{aligned}\frac{\sigma(\mathbf{p},\mathbf{u},t)}{\cos\theta} = R(\mathbf{u}')\Bigg[&\int_{\mathbf{p}}^{\mathbf{p}'}\rho(\mathbf{s},t')e^{-mr}dr \\ &+ \frac{\sigma(\mathbf{p}',\mathbf{u}',t')}{\cos\theta'}e^{-mr'}\chi_{\Gamma^0}(\mathbf{u}') \\ &+ \frac{\lambda(\mathbf{p}',\mathbf{u}',t')}{\sin\theta'}\frac{e^{-mr'}}{r'}\delta_{\Delta^0}(\mathbf{u}') \\ &+ \mu(\mathbf{p}',\mathbf{u}',t')\frac{e^{-mr'}}{r'^{n-1}}\delta_{Y^0}(\mathbf{u}')\Bigg],\end{aligned} \tag{24}$$

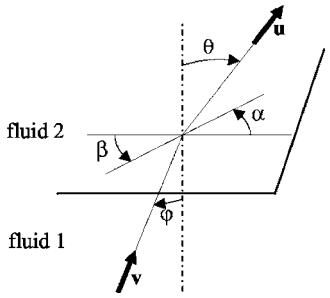


FIG. 6. Refraction at plane interface for incidence direction $\mathbf{v}=\varphi,\beta$ and refraction direction $\mathbf{u}=\theta,\alpha$.

where $\mathbf{p}'$ is the point located on the boundary in direction $-\mathbf{u}'$ from $\mathbf{p}$. This is a functional equation on σ. The first two terms of Eq. (24) were derived in Ref. 22 as an alternative to Eq. (21) for specular reflection. However, in case of partially diffuse reflecting surfaces and then, partially specular reflecting surfaces, the most widely spread solution is rather based on an algorithm which jointly uses the view factor method and the image-source technique[23,24] In Ref. 22, it was pointed out that Eq. (24) can be solved by the image-source technique and that Eq. (24) is more generally equivalent to the ray-tracing technique with specular reflection. But to find an algorithm similar to the collocation method and valid beyond the image-source technique limited in practice to simple polyhedra shape, remains an open question.

IV. TRANSMISSION AND REFRACTION

When waves impinge on the interface separating two media with different phase speeds, they are partially reflected and partially refracted. This is the case for optical or acoustical rays passing through the air-water interface for instance or structural waves in plates at discontinuity of thickness. In some cases, the media are the same on both sides but, the interface is material and then also gives rise to reflection and transmission of waves. Transparency of acoustical waves through walls but also transmission of structural waves through the common edge of right-angled plates are some examples. A fictitious source layer σ_i, where i is a subscript referring to the medium, is introduced on each side of the interface. Equations on σ_i are a simple generalization of previous ones in case of reflection. These equations apply for any transmission and refraction, the details of the particular system at hand, coincidence frequencies, double-leaf panel resonance, air-gap resonance and so on, are rejected in expressions of transmittivity.

Let $R_{ji}(\mathbf{v},\mathbf{u})$ be the *bidirectional transmittivity* from medium j to medium i defined as in Sec. III. R_{ii} is simply the reflectivity of the boundary in medium i. The angles φ and θ are defined in Fig. 6. The leaving radiative intensity $I_i = d\mathcal{P}_i/\cos\theta dSdu$ in medium i is the sum of all fluxes incident from all media. The detailed power balance now reads,

$$I_i(\mathbf{p},\mathbf{u},t)=\frac{1}{\cos\theta}\frac{d\mathcal{P}i}{dSdu}=\sum_i\int R_{ji}(\mathbf{v},\mathbf{u})I_j(\mathbf{p},\mathbf{v},t)\cos\varphi dv. \tag{25}$$

This equation generalizes Eq. (15). As in Sec. III, the equation on σ_i, is obtained by substituting Eqs. (12) and (16) into Eq. (25),

$$\begin{aligned}\frac{\sigma_i(\mathbf{p},\mathbf{u},t)}{\gamma_0\cos\theta}=&\sum_j\int_{\Omega_j}R_{ji}(\mathbf{v},\mathbf{u})\rho_j(\mathbf{s},t')H_j(\mathbf{s},\mathbf{p})\cos\varphi d\Omega_{\mathbf{s}}\\&+\int_{\Gamma_j}R_{ji}(\mathbf{v},\mathbf{u})\sigma_j(\mathbf{q},\mathbf{v},t')H_j(\mathbf{q},\mathbf{p})\cos\varphi d\Gamma_{\mathbf{q}}\\&+\int_{\Delta_j}R_{ji}(\mathbf{v},\mathbf{u})\lambda_j(\mathbf{q},\mathbf{v},t')H_j(\mathbf{q},\mathbf{p})\cos\varphi d\Delta_{\mathbf{q}}\\&+\sum_{\mathbf{q}\in\Upsilon_j}R_{ji}(\mathbf{v},\mathbf{u})\mu_j(\mathbf{q},\mathbf{v},t')H_j(\mathbf{q},\mathbf{p})\cos\varphi.\end{aligned} \tag{26}$$

This equation gives σ_i in terms of other sources.

When emitted energy is diffuse, the bidirectional transmittivity does not depend on the emission direction, $R_{ji}(\mathbf{v},\mathbf{u})=R_{ji}(\mathbf{v})/\gamma$ for all $\mathbf{u}$ where $R_{ji}(\mathbf{v})$ is the *hemispherical transmittivity or transmission efficiency* defined as in Eq. (18). The reflection and transmission sources follow Lambert's law (20) and Eq. (26) on unknown σ_i becomes,

$$\begin{aligned}\frac{\gamma}{\gamma_0}\sigma_i(\mathbf{p},t)=&\sum_j\int_{\Omega_j}R_{ji}(\mathbf{v})\rho_j(\mathbf{s},t')H_j(\mathbf{s},\mathbf{p})\cos\varphi d\Omega_{\mathbf{s}}\\&+\int_{\Gamma_j}R_{ji}(\mathbf{v})\sigma_j(\mathbf{q},t')\cos\theta' H_j(\mathbf{q},\mathbf{p})\cos\varphi d\Gamma_{\mathbf{q}}\\&+\int_{\Delta_j}R_{ji}(\mathbf{v})\lambda_j(\mathbf{q},\mathbf{v},t')H_j(\mathbf{q},\mathbf{p})\cos\varphi d\Delta_{\mathbf{q}}\\&+\sum_{\mathbf{q}\in\Upsilon_j}R_{ji}(\mathbf{v})\mu_j(\mathbf{q},\mathbf{v},t')H_j(\mathbf{q},\mathbf{p})\cos\varphi,\end{aligned} \tag{27}$$

θ' is the emission angle at $\mathbf{q}$ and $t'=t-r/c$. This is a set of Fredholm's integral equations of the second kind. This set of equations turns out to be a powerful tool to predict the repartition of energy in assembled plates in high frequency range.[6,7,25] The solving of this set of equations in steady-state condition is done by a collocation method. It then possible to compute the vibrational energy of each component of built-up structures but also to get the repartition of energy inside each subsystem. This is an improvement of (SEA) which just provides the total vibrational energy of subsystems. Indeed, this method is more time-computation consuming than SEA but significantly less time-consuming than FEM or BEM applied to the classical governing equations.

In case of perfect refraction, incoming rays $\mathbf{u}_j'$ with incidence φ_j in medium j which are refracted into a single ray $\mathbf{u}$ with angle θ in medium i are linked by Snell-Descartes' law of refraction, $\sin\varphi_j/c_j'=\sin\theta/c_i'$, where c_j' the phase speed in medium j. Rays $\mathbf{u}_j'$ and $\mathbf{u}$ are coplanar. The bidirectional transmittivity must read in such a manner that incident power from direction $\mathbf{u}_j'$ is refracted in the only direction $\mathbf{u}$ of medium i,

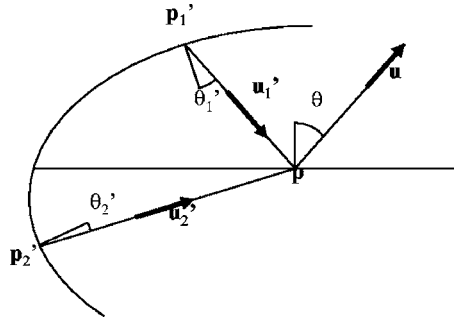


FIG. 7. The power emitted in direction $\mathbf{u}=\theta$, α is the sum of the reflection from direction $\mathbf{u}_1'$ and the transmission from direction $\mathbf{u}_2'$. $\mathbf{p}_i'$ is the point of the boundary which emits in the direction $\mathbf{u}_i'$ with emission angle θ_i'.

$$R_{ji}(\mathbf{v},\mathbf{u})=\left(\frac{c_j'}{c_i'}\right)^{n-1}R_{ji}(\mathbf{v})\frac{\delta(\mathbf{v}-\mathbf{u}_j')}{\cos\varphi}. \tag{28}$$

To show that $R_{ji}(\mathbf{v})$ is the hemispherical transmittivity, we substitute Eq. (28) into Eq. (18), $\int(c_j'/c_i')^{n-1}R_{ji}(\mathbf{v})\delta(\mathbf{v}-\mathbf{u}_j')\cos\theta/\cos\varphi du$. The solid angles attached to the change of variables $\mathbf{u}=\theta$, $\alpha\rightarrow\mathbf{u}_j'=\varphi_j$, $\alpha+\pi$ are $du=\sin\theta d\theta d\alpha$ and $du_j'=\sin\theta_j' d\varphi_j' d\alpha$. They are related by $\cos\theta du/c_i'^{n-1}=\cos\varphi_j' du_j'/c_j'^{n-1}$ (derive the square of Snell-Descartes' equality). The hemispherical transmittivity is therefore $\int R_{ji}(\mathbf{v})\delta(\mathbf{v}-\mathbf{u}_j')\cos\varphi_j/\cos\varphi du_j'=R_{ji}(\mathbf{v})$ which justifies the form of Eq. (28).

Substitution of Eq. (28) into Eq. (25) leads to $I_i(\mathbf{p},\mathbf{u},t)=\Sigma_j(c_j'/c_i')^{n-1}R_{ji}(\mathbf{u}_j')I_j(\mathbf{p},\mathbf{u}_j',t)$. Thus, further substitution of Eqs. (16) and (14) into this equality gives some functional equations for σ_i,

$$\begin{aligned}\frac{\sigma_i(\mathbf{p},\mathbf{u},t)}{\cos\theta}=\sum_j\left(\frac{c_j'}{c_i'}\right)^{n-1}R_{ji}(\mathbf{u}_j')\Bigg[&\int_{\mathbf{p}}^{\mathbf{p}_j'}\rho_j(\mathbf{s},t')e^{-m_j r}dr\\&+\frac{\sigma_j(\mathbf{p}_j',\mathbf{u}_j',t')}{\cos\theta_j'}e^{m_j r'}\chi_{\Gamma_j^0}(\mathbf{u}_j')\\&+\frac{\lambda_j(\mathbf{p}_j',\mathbf{u}_j',t')}{\sin\theta_j'}\frac{e^{-m_j r'}}{r'}\delta_{\Delta_j^0}(\mathbf{u}_j')\\&+\mu_j(\mathbf{p}_j',\mathbf{u}_j',t')\frac{e^{-m_j r'}}{r'^{n-1}}\delta_{\Upsilon_j^0}(\mathbf{u}_j')\Bigg],\end{aligned} \tag{29}$$

where $\mathbf{p}_j'$ is the first point of the boundary Γ_j encountered in direction $-\mathbf{u}_j'$ from $\mathbf{p}$ and θ_j' is the emission angle at $\mathbf{p}_j'$ (Fig. 7). This is a set of functional equations on unknowns σ_i. Equation (29) has been solved for a couple of plates in Ref. 26, whereas some features of Eq. (29) have been discussed in Ref. 22. In particular, it has been shown that Eq. (29) is not symmetrical under time reversing. This is due to the underlying assumption that rays are uncorrelated. To neglect the phase between incoming rays is not equivalent that to neglect the phase between outcoming rays.

Two expressions of transmission efficiency are given in the Appendix for the cases of refraction between two acoustical media and transmission through single walls. Many other cases are tackled in the literature, for instance, transmission through double walls is studied in Refs. 27 and 28, transmission of structural waves at joint of assembled beams in Refs. 29 and 30, and at a joint of assembled plates in Refs. 31 and 32.

V. DIFFRACTION

Geometrical Theory of Diffraction (GTD) introduces diffraction effects in geometrical acoustics. Whereas the condition for existence of classical rays is given by Fermat's principle expressing the stationarity of ray path, existence of diffracted rays follows from the generalized Fermat's principle which states that ray paths with constraints have an extremum optical length.[5] It leads to the existence of a new class of rays diffracted by wedges, peaks, and corners, but also creeping rays in case of diffraction by smooth obstacles. In this section, we just consider diffraction by wedges and corners but not diffraction by smooth obstacles. A fictitious source layer λ is introduced along diffracting edges and some fictitious sources μ are laid on vertices of peaks and corners of plates.

Let define the *bidirectional diffractivity* $D(\mathbf{v},\mathbf{u})$ with incidence $\mathbf{v}$ and emission direction $\mathbf{u}$, by analogy with the bidirectional reflectivity. For diffraction by corners and peaks, $D(\mathbf{v},\mathbf{u})$ is defined as the ratio of the emitted power $d\mathcal{P}$ per unit solid angle du about $\mathbf{u}$ and the incident radiative intensity I in direction $\mathbf{v}$. The detailed power balance is thus,

$$\frac{d\mathcal{P}}{du} = \int D(\mathbf{v},\mathbf{u}) I(\mathbf{p},\mathbf{v},t) dv. \tag{30}$$

The infinitesimal emitted power $d\mathcal{P}$ for a point source μ is the flux of $\mu\mathbf{H}$ through a small area $\epsilon^{n-1}du$. of the sphere of radius ϵ, $d\mathcal{P}=\mu e^{-m\epsilon}/\gamma_0\epsilon^{n-1}\times\epsilon^{n-1}du$. When ϵ goes to zero,

$$\frac{d\mathcal{P}}{du} = \frac{\mu(\mathbf{p},\mathbf{u},t)}{\gamma_0}. \tag{31}$$

The equation on μ is then derived by substituting Eqs. (12) and (31) into Eq. (30),

$$\begin{aligned}\frac{\mu(\mathbf{p},\mathbf{u},t)}{\gamma_0} = &\int_\Omega D(\mathbf{v},\mathbf{u})\rho(\mathbf{s},t')H(\mathbf{s},\mathbf{p})d\Omega_\mathbf{s} \\ &+ \int_\Gamma D(\mathbf{v},\mathbf{u})\sigma(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})d\Gamma_\mathbf{q} \\ &+ \int_\Delta D(\mathbf{v},\mathbf{u})\lambda(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})d\Delta_\mathbf{q} \\ &+ \sum_{\mathbf{q}\in\Upsilon} D(\mathbf{v},\mathbf{u})'\mu(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p}).\end{aligned} \tag{32}$$

This equation gives the power μ in terms of powers of other sources.

For diffraction by wedge whose edge has a length measure noted ν, the *bidirectional diffractivity* $D(\mathbf{v},\mathbf{u})$ is defined as the ratio of the emitted power $d\mathcal{P}$ per unit solid angle du about $\mathbf{u}$ and per unit length $\sin\theta d\nu$ normal to the ray, and the power density $I(\mathbf{p},\mathbf{v},t)\sin\varphi$ incident on the edge,

$$\frac{1}{\sin\theta}\frac{d\mathcal{P}}{d\nu du} = \int D(\mathbf{v},\mathbf{u}) I(\mathbf{p},\mathbf{v},t)\sin\varphi dv. \tag{33}$$

Directions $\mathbf{v}=\varphi,\beta,\mathbf{u}=\theta,\alpha$ are defined in Fig. 8(a). For a line source λ, the infinitesimal emitted power per unit length $d\mathcal{P}/d\nu$ is the flux of $\lambda\mathbf{H}$ through a part $\epsilon^2 du$ of sphere and therefore, $d\mathcal{P}/d\nu=\lambda e^{-m\epsilon}/4\pi\epsilon^2\times\epsilon^2 du$. When ϵ goes to zero,

$$\frac{1}{\sin\theta}\frac{d\mathcal{P}}{d\nu du} = \frac{\lambda(\mathbf{p},\mathbf{u},t)}{4\pi\sin\theta}. \tag{34}$$

The equation on λ is obtained by substitution of Eqs. (12) and (34) into Eq. (33),

$$\begin{aligned}\frac{\lambda(\mathbf{p},\mathbf{u},t)}{4\pi\sin\theta} = &\int_\Omega D(\mathbf{v},\mathbf{u})\rho(\mathbf{s},t')H(\mathbf{s},\mathbf{p})\sin\varphi d\Omega_\mathbf{s} \\ &+ \int_\Gamma D(\mathbf{v},\mathbf{u})\sigma(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})\sin\varphi d\Gamma_\mathbf{q} \\ &+ \int_\Delta D(\mathbf{v},\mathbf{u})\lambda(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})\sin\varphi d\Delta_\mathbf{q} \\ &+ \sum_{\mathbf{q}\in\Upsilon} D(\mathbf{v},\mathbf{u})\mu(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})\sin\varphi.\end{aligned} \tag{35}$$

This is an integral equation on the unknown λ.

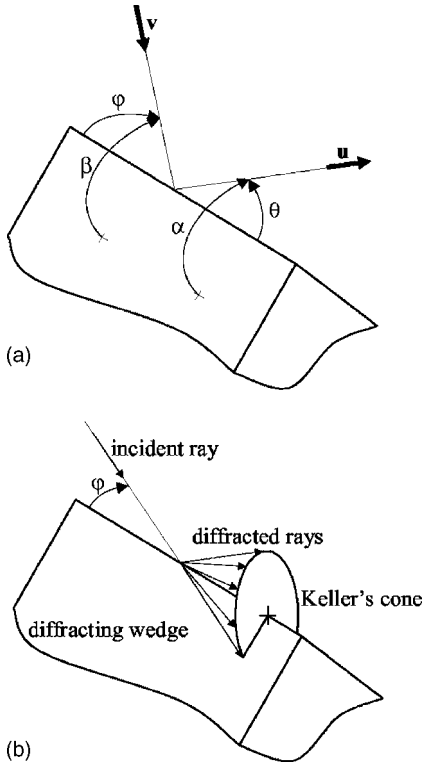


FIG. 8. Diffraction by wedge. (a) Elevation angles and φ, θ are measured with the tangent to the edge and azimuthal angles and β, α are measured in the plane normal to the edge of the wedge. (b) Keller's cone is the set of emission directions for which $\theta=\varphi$.

Until now, it has been tacitly assumed that rays impinging on wedges, peaks, and corners may be diffracted in any direction. However, the generalized Fermat's principle specifies which ray paths are admissible for diffraction. It is found that corners and peaks diffract in all directions, whereas wedges only diffract in the so-called *Keller's cone*. For an incident ray with direction $\mathbf{v}=\varphi,\beta$, Keller's cone is the set of all emission directions $\mathbf{u}=\theta,\ \alpha$ verifying $\theta=\varphi$ [Fig. 8(b)]. The equality of incidence and emission angles is known as Keller's law of diffraction. Let us introduce the *reciprocal Keller's cone* K of direction $\mathbf{u}$ as being the set of all incidence directions $\mathbf{v}$ whose incidence angle φ is equal to the emission angle θ. All energy emerging from the wedge in direction $\mathbf{u}$ stems from the reciprocal Keller's cone. The following form of the bidirectional diffractivity satisfies this condition,

$$D(\mathbf{v},\mathbf{u}) = D(\beta,\alpha)\frac{\delta(\varphi-\theta)}{\sin\varphi\sin\theta}. \tag{36}$$

$D(\beta,\alpha)$ is related to the classical diffraction coefficient d at normal incidence used in GTD by $D(\beta,\alpha)=|d(\beta,\alpha)|^2$. Some expressions for the diffractivity $D(\beta,\alpha)$ are given in the Appendix.

For the particular diffractivity given in Eq. (36), the equation on the unknown λ is obtained by multiplying Eq. (35) by $\sin\theta$ and by substituting Eq. (36). Four integrals then appear in right-hand side which must be carefully evaluated. Since the diffractivity of Eq. (36) contains a Dirac function, the integrands $d\Omega$, $d\Gamma$, $d\Delta$ of these integrals reduces to the surface of reciprocal Keller's cone. Then, let introduce the notation K for the reciprocal Keller's cone, $L=K\cap\Gamma$ for the trace of the reciprocal Keller's cone on the regular boundary and $M=K\cap\Delta$ the discret set of points of Δ lying on the reciprocal Keller's cone. The surface measure on the reciprocal Keller's cone is $dK=r\sin\varphi d\beta dr$ [Fig. 9(a)]. Since the infinitesimal volume is $d\Omega=r^2\sin\varphi d\varphi d\beta dr$ in spherical coordinates, it related to dK by $d\Omega=rd\varphi dK$. The first integral of Eq. (35) multiplied by $\sin\theta$ becomes $\int\delta(\varphi-\theta)D\rho H\times rd\varphi dK=\int_K D\rho Hr dK$, where D of the right-hand side designates $D(\beta,\alpha)$ of Eq. (36). For the second integral, the infinitesimal surface is $d\Gamma=r^2\sin\varphi d\beta/\cos\theta'$, where as usual θ' is the emission angle (Fig. (4)). The length measure is $dL=r\sin\varphi d\beta/\sin\chi$, where χ is the angle between L and the generating line of K [Fig. 9(b)]. Thus, $d\Gamma=r\sin\chi/\cos\theta'\times d\varphi dL$. The second integral is therefore $\int_L D\sigma Hr\sin\chi/\cos\theta' dL$. The third integral reads $\int\delta(\varphi-\theta)D\lambda Hd\Delta=\Sigma D\lambda H\times d\Delta/d\varphi$ where the sum runs over the set $M=K\cap\Delta$. But $d\Delta/d\varphi=r\cos\psi$, where ψ is the angle between Δ and the normal to Keller's cone [Fig. 9(c)]. Finally, the fourth term is $\Sigma\delta(\varphi-\theta)D\mu H$ and the equation on λ for diffraction by wedges is

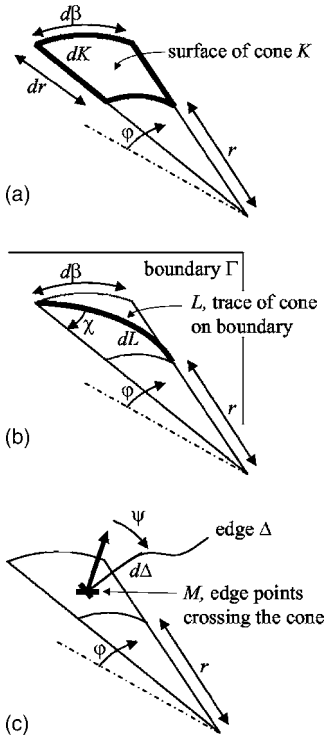


FIG. 9. (a) Surface measure on the reciprocal Keller's cone K. (b) Length measure on the curve $L=K\cap\Gamma$, χ is the angle between the generating line of the cone and the line L. (c) Discrete set $M=K\cap\Delta$, ψ is the angle measured between the tangent to Δ and the normal to K.

$$\begin{aligned}\lambda(\mathbf{p},\mathbf{u},t) = &\int_K D(\beta,\alpha)\rho(\mathbf{s},t')\frac{e^{-mr}}{r}dK_{\mathrm{s}}\\ &+\int_L D(\beta,\alpha)\sigma(\mathbf{q},\mathbf{v},t')\frac{\sin\chi}{\cos\theta'}\frac{e^{-mr}}{r}dL_{\mathbf{q}}\\ &+\sum_{\mathbf{q}\in M} D(\beta,\alpha)\frac{\lambda(\mathbf{q},\mathbf{v},t')}{\cos\psi}\frac{e^{-mr}}{r}\\ &+\sum_{\mathbf{q}\in \mathrm{Y}} D(\beta,\alpha)\mu(\mathbf{q},\mathbf{v},t')\frac{e^{-mr}}{r^2}\delta(\varphi-\theta).\end{aligned} \tag{37}$$

This is a functional equation on λ.

Diffraction by the top of noise barriers or around buildings are two common examples of multidiffraction where rays can be diffracted more than once. These problems are usually solved by BEM for low frequencies and by GTD for higher frequencies. For instance, Pierce[33] gives an approximate expression for double-edge diffraction by thick three-sided barriers using GTD. This solution is assessed for wide barriers by Kurze[34] while Medwin *et al.*[35] generalize this result in the time domain. However, all these methods require a finite number of rays between the diffracting edges and then are limited to a finite order of diffraction. Equations (32) and (37) are an alternative ray method for multiple diffraction. They lead to a finite set of linear equations where the unknowns are the diffracted powers. Solving these equations allows to account the infinite number of diffractions in a single step[36,37] but does not account for interference effects between edges.

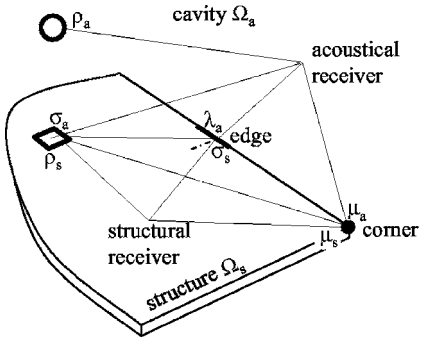


FIG. 10. Actual and fictitious sources contributing to the energy in structure and in acoustics. Structural sources are noted as ρ_s, reflection sources on edge are σ_s, and diffraction sources of corners are μ_s. Acoustical sources are noted as ρ_a, surface radiation sources are σ_a (surface mode radiation), line radiation sources are λ_a (edge mode radiation), and point radiation sources are μ_a (corner mode radiation).

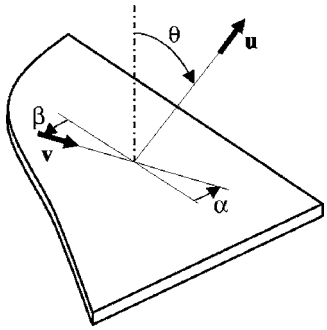


FIG. 11. Surface mode radiation. Elevation angle θ is measured with the normal to the surface and azimuthal angles β, α are measured in the plane of the surface.

VI. RADIATION OF SOUND

Radiation of sound is usually described using two different approaches. The modal approach[38,39] introduces a radiation factor for any mode of the radiating structure. It is well adapted to identify which modes are responsible of the radiation. The wave approach is more appropriate to ray theories. From this point of view, radiation occurs in three situations.[40] First, structural rays continuously loss their energy when traveling and then give rise to radiation of acoustical rays. This is the so-called surface mode radiation. Second, structural rays when impinging on edges are partially reflected and partially diffracted into acoustics. This is the edge mode radiation. Finally, singular points of structures are also responsible of diffraction of structural rays. This is the corner mode radiation. The subscript s is introduced for quantities related to structure and a for acoustics. The sound speed is noted c_0. Then, a fictitious source layer σ_a, density of radiated power, is introduced at any point of the radiating surface. For edge mode radiation, a fictitious source σ_s is introduced on the edge for reflection into structure and a source λ_a for radiation. Finally, some point sources μ_s for structure and μ_a a for acoustics are laid on each singular point of the structure (Fig. 10).

Radiation by surface mode only occurs beyond the coincidence frequency when structural waves are supersonic. Let us introduce a *bidirectional radiation coefficient* $A_{sa}(\mathbf{v},\mathbf{u})$ as the acoustical radiative intensity in direction $\mathbf{u}$ for a unit incident structural radiative intensity in direction $\mathbf{v}$. The directions $\mathbf{v}=\beta$ and $\mathbf{u}=\theta$, α and their related angles are drawn in Fig. 11. The detailed power balance then states that the radiated intensity is the sum of all contributions of incident structural rays,

$$I(\mathbf{p},\mathbf{u},t)=\frac{1}{\cos\theta}\frac{d\mathcal{P}}{dSdu}=\int A_{sa}(\mathbf{v},\mathbf{u})I(\mathbf{p},\mathbf{v},t)dv. \tag{38}$$

In the mean time, the attenuation factor m_s for structure is the sum of a term for internal losses $\eta\omega/c_s$, where η is the damping loss factor and ω is the circular frequency, and an additional term for radiation. Consider a structural plane wave with radiative intensity $I=I_0\delta(\mathbf{v}-\mathbf{v}_0)$. The decrease of intensity thickness dx is $-dI_0=m_sI_0dx$. Internal losses are $\eta\omega/c_s\times I_0dx$, whereas radiation losses are $2d\mathcal{P}/dS\times dx$. The factor 2 stems from the presence of fluid on both sides of the structure. From Eq. (38), $d\mathcal{P}/dS\times dx=\int A_{sa}\cos\theta du\times I_0dx$ and therefore the power balance states,

$$m_s=\frac{\eta\omega}{c_s}+2\int A_{sa}(\mathbf{v}_0,\mathbf{u})\cos\theta du. \tag{39}$$

The integral does not depend on direction $\mathbf{v}_0$ for isotropic structure. Now, the acoustical surface sources σ_a distributed over the structure provide this energy lost by the structure. Equation (16) gives the left-hand side of Eq. (38), whereas the right-hand side is obtained by substituting Eq. (12),

$$\begin{aligned}\frac{\sigma_a(\mathbf{p},\mathbf{u},t)}{4\pi\cos\theta}&=\int_{\Omega_s}A_{sa}(\mathbf{v},\mathbf{u})\rho_s(\mathbf{s},t')H(\mathbf{s},\mathbf{p})d\Omega_{\mathbf{s}}\\&+\int_{\Gamma_s}A_{sa}(\mathbf{v},\mathbf{u})\sigma_s(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})d\Gamma_{\mathbf{q}}\\&+\sum_{\mathbf{q}\in\Upsilon_s}A_{sa}(\mathbf{v},\mathbf{u})\mu_s(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p}).\end{aligned} \tag{40}$$

In this equation, Ω_s denotes the structural domain, Γ_s its boundary, and Υ_s the set of diffracting points. Snell-Descartes' law states that the emission direction $\mathbf{u}$ has a polar angle θ_0, measured with the normal to the surface, such as $1/c_s'=\sin\theta_0/c_0$. The radiated ray $\mathbf{u}$ has also an azimuthal angle β, measured in the plane of structure, equal to the azimuthal angle α of structural ray $\mathbf{v}$. Then, the bidirectional radiation coefficient A_{sa} reduces to the particular form,

$$A_{sa}(\mathbf{v},\mathbf{u})=A_{sa}\delta(\alpha-\beta)\frac{\delta(\theta-\theta_0)}{\sin\theta_0}, \tag{41}$$

where A_{sa} in the second-hand side is a parameter depending on structure and fluid properties. This parameter and the related attenuation factor m_s are given in the Appendix for lightly loaded thin plates while the most general case of fluid-loaded thick plates is discussed in Ref. 41. Now substitution of Eq. (41) into Eq. (40) leads to the equation on σ_a,

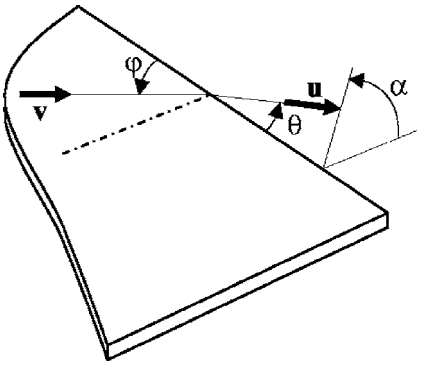


FIG. 12. Edge mode radiation. Elevation angle φ, θ is measured with the tangent to the edge and azimuthal angle α is measured in the plane of the edge.

$$\frac{\sigma_a(\mathbf{p},\mathbf{u},t)}{2\cot\theta_0} = A_{sa}\delta(\theta-\theta_0)\times\left[\int_{\mathbf{p}}^{\mathbf{p}'}\rho_s(\mathbf{s},t')e^{-m_s r}dr + \frac{\sigma_s(\mathbf{p}',\mathbf{u}',t')}{\cos\theta'}e^{-m_s r'}\chi_{\Gamma_s^0(\mathbf{u}')} + \mu_s(\mathbf{p}',\mathbf{u}',t')\frac{e^{-m_s r'}}{r'}\delta_{\Upsilon_s^0}(\mathbf{u}')\right], \tag{42}$$

where $\mathbf{u}'=\alpha$ and $\mathbf{u}=\theta,\alpha$. $\mathbf{p}'$ is as usual the first point of the boundary in direction $-\mathbf{u}'$. This equation gives the fictitious sources σ_a for surface mode radiation in terms of structural sources ρ_s, σ_s, and μ_s. It is solved in Ref. 42 for the case of a beam radiating in a two-dimensional acoustic medium. Radiation only occurs in direction θ_0 and therefore, sound is emitted in two strips. The intersection of these strips is a double zone with a higher sound pressure level and outside of the strips is a shadow zone.

Radiation by edge mode occurs at any frequency. This is a particular case of diffraction with a structural ray diffracted into fluid. The structure to acoustic bidirectional diffractivity for edge is defined as in Eq. (33),

$$\frac{1}{\sin\theta}\frac{d\mathcal{P}}{d\nu du} = \int D_{sa}^e(\mathbf{v},\mathbf{u})I(\mathbf{p},\mathbf{v},t)\sin\varphi dv, \tag{43}$$

where the incident direction $\mathbf{v}=\varphi$ belongs to the plane of structure and the radiation direction $\mathbf{u}=\theta$, α is in the fluid (Fig. 12). Structural waves are partially reflected into structure itself and partially diffracted into acoustics. Thus, the edge has a reflection efficiency R_s less than unity. Consider a structural plane wave with radiative intensity $I=I_0\delta(\mathbf{v}-\mathbf{v}_0)$ incident upon the edge. The diffracted power per unit length is given by Eq. (43), $d\mathcal{P}/d\nu=I_0\sin\varphi_0\int D_{sa}\sin\theta du$. This power is not reflected into structure and therefore the power balance imposes,

$$R_s(\mathbf{v}_0) = 1-\int D_{sa}^e(\mathbf{v}_0,\mathbf{u})\sin\theta du. \tag{44}$$

This reflection efficiency R_s depends on the incident direction $\mathbf{v}_0$. The exact expression for the reflection coefficient R_s of membranes is derived in Ref. 43 while the case of baffled and nonbaffled plates is solved in Ref. 44. The unknown σ_s is determined by applying Eq. (17) with the reflection efficiency R_s. Indeed Eqs. (21) and (24) must be preferred in case of diffuse and specular reflection. The energy converted into acoustical waves is emanated by some acoustical sources distributed along the edge of the structure. Their power per unit length λ_a is determined by,

$$\frac{\lambda_a(\mathbf{p},\mathbf{u},t)}{4\pi\sin\theta} = \int_{\Omega_s} D_{sa}^e(\mathbf{v},\mathbf{u})\rho_s(\mathbf{s},t')H(\mathbf{s},\mathbf{p})\sin\varphi d\Omega_{\mathbf{s}} + \int_{\Gamma_s} D_{sa}^e(\mathbf{v},\mathbf{u})\sigma_s(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})\sin\varphi d\Gamma_{\mathbf{q}} + \sum_{\mathbf{q}\in\Upsilon_s} D_{sa}^e(\mathbf{v},\mathbf{u})\mu_s(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})\sin\varphi. \tag{45}$$

This is the most general equation governing the line source λ_a. But as in diffraction by wedges, acoustical rays are radiated in Keller's cone. However, incidence and emission angles measured with the tangent to the edge are now related by $\cos\varphi/c_s'=\cos\theta/c_0$. Since the incident direction is restricted to be in the plane of structure, the reciprocal Keller's cone of a radiated ray $\mathbf{u}$ now reduces to a single direction noted $\mathbf{u}'=\varphi'$. The bidirectional diffractivity takes the form,

$$D_{sa}^e(\mathbf{v},\mathbf{u}) = \frac{c_s'}{c_0}D_{sa}^e(\alpha)\frac{\delta(\varphi-\varphi')}{\sin^2\varphi}, \tag{46}$$

where $D_{sa}^e(\alpha)$ is the square of the classical diffraction coefficient at normal incidence. When introducing Eq. (46) into Eq. (45) and, as usual, developing the infinitesimal surface $d\Omega=rdrd\varphi$ and the infinitesimal boundary length $d\Gamma=rd\varphi/\cos\theta'$, the equation on λ_a is obtained,

$$\frac{\lambda_a(\mathbf{p},\mathbf{u},t)}{2\sin\theta} = \frac{c_s'}{c_0}D_{sa}^e(\alpha)\frac{1}{\sin\varphi'}\left[\int_{\mathbf{p}}^{\mathbf{p}'}\rho_s(\mathbf{s},t')e^{-mr}dr + \frac{\sigma_s(\mathbf{p}',\mathbf{u}',t')}{\cos\theta'}e^{-mr'}\chi_{\Gamma_s^0(\mathbf{u}')} + \mu_s(\mathbf{p}',\mathbf{u}',t')\frac{e^{-mr'}}{r'}\delta_{\Upsilon_s^0}(\mathbf{u}')\right]. \tag{47}$$

The power density λ_a is then related to structural sources ρ_s, σ_s, and μ_s.

Radiation by corner mode also occurs at any frequency. The bidirectional diffractivity, as in Eq. (30), is defined by the following detailed power balance:

$$\frac{d\mathcal{P}}{du} = \int D_{sa}^c(\mathbf{v},\mathbf{u})I(\mathbf{p},\mathbf{v},t)dv, \tag{48}$$

where $\mathbf{v}$ is in the structure and $\mathbf{u}$ is any direction in the fluid (Fig. 13). The left-hand side of Eq. (48) is given by Eq. (31) and the right-hand side by Eq. (12),

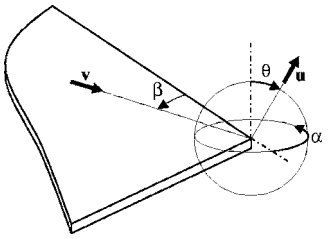


FIG. 13. Corner mode radiation. Elevation angle θ is measured with the normal to the surface and azimuthal angles β, α are measured in the plane of the surface.

$$\frac{\mu_a(\mathbf{p},\mathbf{u},t)}{4\pi} = \int_{\Omega_s} D^c_{sa}(\mathbf{v},\mathbf{u})\rho_s(\mathbf{s},t')H(\mathbf{s},\mathbf{p})d\Omega_s + \int_{\Gamma_s} D^c_{sa}(\mathbf{v},\mathbf{u})\sigma_s(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})d\Gamma_{\mathbf{q}} + \sum_{\mathbf{q}\in Y_s} D^c_{sa}(\mathbf{v},\mathbf{u})\mu_s(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p}). \quad (49)$$

The fictitious acoustical sources μ_a are related to structural sources ρ_s, σ_s, and μ_s. Concerning the sources μ_s at the same corner, this is a structural diffraction problem. Thus, μ_s is given by Eq. (32) with the appropriate diffraction coefficient.

VII. STRUCTURAL RESPONSE

Reflection, transmission, and sound absorption by walls has been addressed in Secs. III and IV. But in some cases, the acoustical energy lost during reflection is converted into structural energy. This phenomenon followed by radiation of sound is named *resonant transparency* in SEA literature meaning that structural modes are involved. An overview of sound transmission through infinite and finite structures is available in Ref. 45. Acoustic to structure conversion mode is the reciprocal problem of sound radiation and since all ray paths can be traveled in both directions, any radiation mode *a priori* gives rise to a conversion mode and conversely. The case of rays impinging on the surface of the structure is first considered. Diffraction of acoustical rays into structure by edges is modeled by introducing some fictitious sources σ_s for the structure and λ_a for acoustics while diffraction by edges or any singular point requires some point sources μ_s and μ_a.

This is a well-known result that a plane wave impinging on an infinite structure is reflected, transmitted, and that a forced term appears in the structure. Fictitious sources σ_a for reflection and transmission are given by Eqs. (26), (27), and (29) depending on the type of transmittivity. The incident power is totally reflected and transmitted and no energy is supplied to the structure. This result holds even at θ_0 incidence and thus, seems to state that the reciprocal path of surface radiated rays does not exist. However it does exist, but for an incident wave having a complex wave number. In the presence of fluid on both sides of the structure, a further requirement is that two rays simultaneously strike the structure at θ_0 incidence and at the same point. This phenomenon is possible but improbable. Anyway, within the framework of the present theory, correlation of rays have been neglected and thus, this type of conversion is assumed to be never realized.

Equations (7) and (8) give the contribution of the free waves propagating inside the structure. But, the presence of a forced term in the structure can lead to a significant increase of vibrational level and therefore cannot be neglected. Let define the forcing coefficient A_0 as the ratio of the vibrational energy of the forced term and the incident acoustical power of a plane wave (Appendix . The total vibrational energy of the forced term W_0 is then the sum,

$$W_0(\mathbf{r},t) = \int A_0(\varphi)I(\mathbf{r},\mathbf{v},t)\cos\varphi dv, \quad (50)$$

for all incident waves. Introducing Eq. (12), the forced term becomes,

$$W_0(\mathbf{r},t) = \int_{\Omega_a} A_0(\varphi)\rho(\mathbf{s},t')H(\mathbf{s},\mathbf{r})\cos\varphi d\Omega_{\mathbf{s}} + \int_{\Gamma_a} A_0(\varphi)\sigma(\mathbf{p},\mathbf{v},t')H(\mathbf{p},\mathbf{r})\cos\varphi d\Gamma_{\mathbf{p}} + \int_{\Delta_a} A_0(\varphi)\lambda(\mathbf{p},\mathbf{v},t')H(\mathbf{p},\mathbf{r})\cos\varphi d\Delta_{\mathbf{p}} + \sum_{\mathbf{p}\in Y_a} A_0(\varphi)\mu(\mathbf{p},\mathbf{v},t')H(\mathbf{p},\mathbf{r})\cos\varphi. \quad (51)$$

The vibrational energy in the structure is therefore the sum $W(\mathbf{r},t)+\mathbf{W}_0(\mathbf{r},t)$, where W is the energy of the free term given by Eq. (7) and W_0 the energy of the forced term given by Eq. (51).

Conversion by edge is once again a diffraction problem. The bidirectional diffractivity $D_{as}(\mathbf{v},\mathbf{u})$ is defined as in Eq. (33),

$$\frac{1}{\sin\theta}\frac{d\mathcal{P}}{dvdu} = \int D^e_{as}(\mathbf{v},\mathbf{u})I(\mathbf{p},\mathbf{v},t)\sin\varphi dv. \quad (52)$$

All the energy impinging on the edge is either diffracted into structural wave or diffracted into acoustical wave. The acoustical diffraction sources λ_a have ever been found in Eq. (37) and the structural diffraction sources σ_s distributed along the edge are given by

$$\frac{\sigma_s(\mathbf{p},\mathbf{u},t)}{2\pi\sin\theta} = \int_{\Omega_a} D^e_{as}(\mathbf{v},\mathbf{u})\rho_a(\mathbf{s},t')H(\mathbf{s},\mathbf{p})\sin\varphi d\Omega_{\mathbf{s}} + \int_{\Gamma_a} D^e_{as}(\mathbf{v},\mathbf{u})\sigma_a(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})\sin\varphi d\Gamma_{\mathbf{q}} + \int_{\Delta_a} D^e_{as}(\mathbf{v},\mathbf{u})\lambda_a(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})\sin\varphi d\Delta_{\mathbf{q}} + \sum_{\mathbf{q}\in Y_a} D^e_{as}(\mathbf{v},\mathbf{u})\mu_a(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})\sin\varphi. \quad (53)$$

The power densities σ_s are then related to acoustical sources $\rho_a,\sigma_a,\lambda_a$, and μ_a. For any ray $\mathbf{u}$ in the structure is attached a reciprocal Keller's cone whose axis is the edge and angle φ'

is defined by the law of diffraction $\cos\varphi'/c_0=\cos\theta/c_s'$. The bidirectional diffractivity taking into account this law of diffraction is,

$$D_{as}^{e}(\mathbf{v},\mathbf{u})=\frac{c_0}{c_s'}D_{as}^{e}(\beta)\frac{\delta(\varphi-\varphi')}{\sin^2\varphi}. \tag{54}$$

Substitution of Eq. (54) into Eq. (53) is done in same condition as for wedge diffraction. The measures $d\Omega$, $d\Gamma$, and $d\Delta$ are developed in spherical coordinates with the result,

$$\begin{aligned}2\frac{\sigma_s(\mathbf{p},\mathbf{u},t)}{\sin\theta}=\frac{c_0}{c_s'}\frac{1}{\sin\varphi'}\Bigg[&\int_K D_{as}^{e}(\beta)\rho_a(\mathbf{s},t')\frac{e^{-mr}}{r}dK_{\mathbf{s}}\\&+\int_L D_{as}^{e}(\beta)\sigma_a(\mathbf{q},\mathbf{v},t')\frac{\sin\chi}{\cos\theta'}\frac{e^{-mr}}{r}dL_{\mathbf{q}}\\&+\sum_{\mathbf{q}\in M} D_{as}^{e}(\beta)\frac{\lambda_a(\mathbf{q},\mathbf{v},t')}{\cos\psi}\frac{e^{-mr}}{r}\\&+\sum_{\mathbf{q}\in \Upsilon_a} D_{as}^{e}(\beta)\mu_a(\mathbf{q},\mathbf{v},t')\frac{e^{-mr}}{r^2}\delta(\varphi-\varphi')\Bigg],\end{aligned} \tag{55}$$

where as for wedged diffraction, dK is the surface measure on the reciprocal Keller's cone K whose semiangle is φ', dL is the length measure on $L=K\cap\Gamma_a$, θ' is the emission angle, and χ is the angle between L and the emission direction, $M=K\cap\Delta_a$ is the set of points where the line Δ_a crosses the cone K and ψ is the related angle between the tangent to Δ_a and the normal to K. The bidimensional Eq. (55) jointly with Eq. (47) is solved in Ref. 46 for the case of a finite baffled plate with simply supported edges. Results of this method are compared with some reference results from BEM.

Finally, conversion by corner is also a diffraction problem. The detailed power balance similar to Eq. (30) reads

$$\frac{d\mathcal{P}}{du}=\int D_{as}^{c}(\mathbf{v},\mathbf{u})I(\mathbf{p},\mathbf{v},t)dv, \tag{56}$$

for any direction $\mathbf{u}$ into the structure. The structural fictitious source μ_s emits a power per unit angle $d\mathcal{P}/du=\mu_s/2\pi$ and, thus, once again μ_s is determined by an integral equation,

$$\begin{aligned}\frac{\mu_s(\mathbf{p},\mathbf{u},t)}{2\pi}=&\int_{\Omega_a} D_{as}^{c}(\mathbf{v},\mathbf{u})\rho_a(\mathbf{s},t')H(\mathbf{s},\mathbf{p})d\Omega_{\mathbf{s}}\\&+\int_{\Gamma_a} D_{as}^{c}(\mathbf{v},\mathbf{u})\sigma_a(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})d\Gamma_{\mathbf{q}}\\&+\int_{\Delta_a} D_{as}^{c}(\mathbf{v},\mathbf{u})\lambda_a(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p})d\Delta_{\mathbf{q}}\\&+\sum_{\mathbf{q}\in \Upsilon_a} D_{as}^{c}(\mathbf{v},\mathbf{u})\mu_a(\mathbf{q},\mathbf{v},t')H(\mathbf{q},\mathbf{p}),\end{aligned} \tag{57}$$

where μ_s is related to acoustical sources ρ_a, σ_a, λ_a, and μ_a. The sources μ_a for the same problem are given by Eq. (32) with the acoustics to acoustics diffraction coefficient.

It is remarkable that the intensity of the forced term does not appear in Eq. (17) giving σ_s the power being reflected neither in Eq. (47) giving λ_a the power being radiated by edges. Its seems that this forced intensity is not reflected neither radiated and thus the power balance seems to be violated. However, it is not. The diffraction coefficients D_{as} and the related one D_{aa} for the diffraction of acoustical waves interacting with the edge of the structure take into account the presence of the forced vibration. And therefore, the energy of the forced term is ever accounted for in the fictitious source σ_s of Eq. (53) (same for λ_a).

VIII. CONCLUSION

In this paper, the basic equations governing energy of uncorrelated ray fields have been presented. All classical phenomena of vibroacoustics are accounted for, that is reflection, refraction, transmission, diffraction, radiation by surface, edge or corner modes and structural response. These equations are derived within the framework of the Geometrical Theory of Diffraction which is the natural theory to describe these rays. The method is based on the use of some fictitious sources each time rays are reflected, refracted, transmitted or diffracted. The power of a fictitious source is the sum of powers of all individual rays being deviated at this point. This is the main difference with the ray-tracing technique which requires that all ray paths from a source to a receiver are determined.

Besides, the proposed formalism is entirely based on energy variables. Indeed, the underlying assumption is that all rays are uncorrelated. In general, this is not true in low frequency range but this is a relevant assumption in high frequency range especially if rms values of the field in wideband are expected. The spirit of Statistical Energy Analysis is preserved: a description of vibrational fields in terms of energy well suited for high frequencies. However, the present equations do not assume that fields are diffuse in all subsystems and even, fields can be largely nondiffuse. The present equations can thus be considered as an extension of SEA.

Energy equations of vibrational fields are useful for several purposes. The "radiosity method" in room acoustics is an efficient method to determine the reverberation-time beyond the validity of Sabine's formula. But similar equations in steady-state condition can further give the repartition of vibrational energy inside largely nondiffuse fields. The present theory embodies these two methods with, in addition, the contribution of diffraction.

APPENDIX

The reflection and transmission efficiencies at the interface between two fluids with different acoustical impedances $Z_i=\rho_i c_i$, $i=1,2$, where ρ_i is the mass per unit volume and c_i is the sound speed are[47]

$$R_{11}(\varphi)=\left|\frac{Z_2\sec\theta-Z_1\sec\varphi}{Z_2\sec\theta+Z_1\sec\varphi}\right|^2, \tag{A1}$$

$$R_{12}(\varphi)=\frac{c_1\sec\varphi}{c_2\sec\theta}\times\left|\frac{2Z_2\sec\theta}{Z_2\sec\theta+Z_1\sec\varphi}\right|^2, \tag{A2}$$

where φ is the incidence angle and θ is the refracted angle measured with the normal to the interface in Figs. 5 and 6,

$\sec\varphi = 1/\cos\varphi$ is the secant function. These angles are related by Snell-Descartes' law of refraction $\sin\varphi/c_1' = \sin\theta/c_2'$. These equations apply for instance at the water-air interface.

Reflection and transmission of acoustical waves through walls with same fluid of impedance $Z_0 = \rho_0 c_0$ on both sides are given by the following relationships:[11]

$$R_{11}(\varphi) = \left| \frac{Z_s}{Z_s + 2Z_0 \sec\varphi} \right|^2, \tag{A3}$$

$$R_{12}(\varphi) = \left| \frac{2Z_0 \sec\varphi}{Z_s + 2Z_0 \sec\varphi} \right|^2, \tag{A4}$$

where $Z_s = [B(\omega \sin\varphi/c_s')^4 - m\omega^2]/i\omega$ is the mechanical impedance of the wall at incidence φ, B being the bending stiffness and m the mass per unit area of the wall. The forced term in the wall is not a structural ray. Its wave number matches with the trace of the acoustical wave number and the forcing coefficient is,

$$A_0(\varphi) = \frac{8B\omega^2 Z_0 \sec\varphi \sin^3\varphi}{c_0^3 |Z_s + 2Z_0 \sec\varphi|^2}. \tag{A5}$$

The diffractivity of a wedge at normal incidence $\varphi = \theta = \pi/2$ is[5,48,49]

$$D(\beta,\alpha) = \frac{\nu^2 \sin^2(\nu\pi)}{2\pi k} \left[\frac{1}{\cos(\nu\pi) - \cos[\nu(\alpha+\beta)]} + \frac{1}{\cos(\nu\pi) - \cos[\nu(\alpha-\beta)]} \right]^2, \tag{A6}$$

where $\nu = \pi/\phi$ is the wedge index, ϕ being the outer angle of the wedge, and k is the acoustical wave number. The angles α and β are measured in the normal plane to the wedge as defined in Fig. 8. This expression is singular for $\alpha = \pm\beta$ and does not predict the correct values near these angles. An alternative expression for D valid both within and outside the transition regions is given by the Uniform Theory of Diffraction,[49]

$$D(\beta,\alpha) = \frac{\nu^2}{8\pi k} \left| \cot\left[\frac{\nu(\pi+\alpha-\beta)}{2} F(kLa^+(\alpha-\beta)) \right] + \cot\left[\frac{\nu(\pi-\alpha+\beta)}{2} F(kLa^-(\alpha-\beta)) \right] + \cot\left[\frac{\nu(\pi+\alpha+\beta)}{2} F(kLa^+(\alpha+\beta)) \right] + \cot\left[\frac{\nu(\pi-\alpha-\beta)}{2} F(kLa^-(\alpha+\beta)) \right] \right|^2, \tag{A7}$$

$L = r\sin^2\varphi$, r being the source-receiver distance and φ the incidence angle and $a^{\pm}(\zeta) = 2\cos^2(\pi N^{\pm}/\nu - \zeta/2)$, $N^{\pm}$ is an integer which more nearly satisfies the equality $2\pi N^{\pm}/\nu - \zeta = \pm\pi$. $F(X) = 2i\sqrt{X}e^{iX}\int_{\sqrt{X}}^{\infty} e^{-i\tau^2} d\tau$ is a transition function which involves a Fresnel integral, and $\cot x$ is the cotangent function. It can be checked that Eqs. (A6) and (A7) agree well for large X.[49]

The radiation coefficient A_{sa} introduced in Eq. (41) is found by solving the classical governing equation of fluid-loaded Love's plate, with an incident structural plane wave. For a light fluid (air),

$$A_{sa} = \frac{\rho_0}{2m} \frac{M}{M^2 - 1} \quad M > 1, \tag{A8}$$

where $M = c_s'/c_0$ is the Mach number of the structural wave. When the structural wave is substance ($M < 1$) there is no radiation, i.e., $A_{sa} = 0$. The attenuation factor m_s is calculated with Eq. (39),

$$m_s = \frac{\eta\omega}{c_s} + \frac{\rho_0}{m} \frac{1}{\sqrt{M^2 - 1}} \quad M > 1. \tag{A9}$$

This last equality is also twice the imaginary part of the structural wave number given by the dispersion of the fluid-loaded plate.

[1] See any standard book in acoustics for instance, H. Kuttruff, *Room Acoustics*, 3rd ed. (Elsevier Applied Science, London, 1991). An historical review is given in W. Joyce, "Sabine's reverberation time and ergodic auditoriums," J. Acoust. Soc. Am. **58**, 643–655 (1975).

[2] H. Kuttruff, "Simulierte nachhallkurven in Rechteckräumen mit diffusem schallfeld," Acustica **25**, 333–342 (1971).

[3] R. H. Lyon, *Statistical Energy Analysis of Dynamical Systems: Theory and Application* (MIT Press, Cambridge, Massachusetts, 1975).

[4] A. Carcaterra, "Entropy in vibration, energy sharing among linear and nonlinear freely vibrating systems," 10th International Congress on Sound and Vibration, Stockholm Sweden (2003), pp. 2025–2032.

[5] J. B. Keller, "Geometrical theory of diffraction," J. Opt. Soc. Am. **62**, 116–130 (1962).

[6] A. Le Bot, "Energy transfer for high frequencies in built-up structures," J. Sound Vib. **250**, 247–275 (2002).

[7] A. Le Bot, "A vibroacoustic model for high frequency analysis," J. Sound Vib. **211**, 537–554 (1998).

[8] M. F. Modest, "*Radiative Heat Transfer*" (Mc Graw-Hill, New York, 1993).

[9] S. Chandrasekhar, *Radiative Transfer* (Dover, New York, 1960).

[10] W. Joyce, "Sabine's reverberation time and ergodic auditoriums," J. Acoust. Soc. Am. **58**, 643–665 (1975).

[11] L. Cremer, M. Heckl, and E. E. Ungar, *Structure-Borne Sound*, 2nd ed. (Springer-Verlag, Berlin, 1988).

[12] H. Kuttruff, *Room Acoustics*, 3rd ed. (Elsevier Applied Science, London, 1991).

[13] R. Gerlach and V. Mellert, "Der nachhallvorgang als markoffsche kette theorie und erste experimentelle überprüfung," Acustica **32**, 211–227 (1975).

[14] E. N. Gilbert, "An interactive calculation of auditorium reverberation," J. Acoust. Soc. Am. **69**, 178–184 (1981).

[15] R. Miles, "Sound field in a rectangular enclosure with diffusely reflecting boundaries," J. Sound Vib. **92**, 203–226 (1984).

[16] H. Kuttruff, "A simple iteration scheme for the computation of decay constants in enclosures with diffusely reflecting boundaries," J. Acoust. Soc. Am. **98**, 288–293 (1995).

[17] M. Carroll and C. Chien, "Decay of reverberent sound in a spherical enclosure," J. Acoust. Soc. Am. **62**, 1442–1446 (1977).

[18] A. Le Bot and L. Ricol, "Integral equation instead of heat conduction equation for medium and high frequencies," Inter-Noise'95, Newport Beach (1995), pp. 579–582.

[19] H. Kuttruff, "Energetic sound propagation in rooms," Acust. Acta Acust. 83, 622–628 (1997).

[20] A. Le Bot and A. Bocquillet, "Comparison of an integral equation on energy and the ray-tracing technique for room acoustics," J. Acoust. Soc. Am. **108**, 1732–1740 (2000).

[21] L. P. Franzoni, D. B. Bliss, and J. W. House, "An acoustic boundary element method based on energy and intensity variables for prediction of high frequency broadband sound fields." J. Acoust. Soc. Am. **110**, 3071–3080 (2001).

[22] A. Le Bot, "A functional equation for the specular reflection of rays," J.

Acoust. Soc. Am. **112**, 1276–1287 (2002).
[23]E. R. Eckert and E. M. Sparrow, "Radiative heat exchange between surfaces with specular reflection," Int. J. Heat Mass Transfer **3**, 42–54 (1961).
[24]H. Kuttruff, "Stationary propagation of sound energy in flat enclosures with partially diffuse surface reflection," Acust. Acta Acust. 86, 1028–1033 (2000).
[25]E. Sarradj, "High frequency boundary integral method as an alternative to statistical energy analysis," 10th International Congress on Sound and Vibration, Stockholm Sweden (2003), pp. 1131–1138.
[26]V. Cotoni and A. Le Bot, "Specular and diffuse of rays in coupled thin plates at high frequencies" J. Sound Vib. **265**, 23–41 (2003).
[27]A. London, "Transmission of reverberant sound through double walls," J. Acoust. Soc. Am. **22**, 270–279 (1950).
[28]A. J. Price and M. J. Crocker, "Sound transmission through double panels using Statistical Energy Analysis," J. Acoust. Soc. Am. **47**, 683–693 (1970).
[29]J. L. Horner and R. G. White, "Prediction of vibrational power transmission through jointed beams," Int. J. Mech. Sci. **32**, 215–223 (1990).
[30]R. C.N. Leung and R. J. Pinnington, "Wave propagation through right-angled joints with compliance-flexural incident wave," J. Sound Vib. **142**, 31–46 (1990).
[31]W. Wöhle, T. Beckmann, and H. Schreckenbach, "Coupling loss factors for Statistical Energy Analysis of sound transmission at rectangular slab joints, Part I," J. Sound Vib. **77**, 323–334 (1981).
[32]W. Wöhle, T. Beckmann, and H. Schreckenbach, "Coupling loss factors for Statistical Energy Analysis of sound transmission at rectangular slab joints, Part II," J. Sound Vib. **77**, 335–344 (1981).
[33]A. D. Pierce, "Diffraction around corners and over barriers," J. Acoust. Soc. Am. **55**, 941–955 (1974).
[34]U. J. Kurze, "Noise reduction by barriers," J. Acoust. Soc. Am. **55**, 504–518 (1974).
[35]H. Medwin, E. Childs, and G. M. Jebsen, "Impulse studies of double diffraction: A discrete Huyghens interpretation," J. Acoust. Soc. Am. **72**, 1005–1013 (1982).
[36]E. Reboul, A. Le Bot, and J. Perret-Liaudet, "Introduction of acoustical diffraction in the radiative transfer method," C. R. Mec. **332**, 505–511 (2004).
[37]E. Reboul, A. Le Bot, and J. Perret-Liaudet, "Radiative transfer equation for multiple diffraction," J. Acoust. Soc. Am. **118**, 1326–1334 (2004).
[38]P. W. Smith, "Response and radiation of structural modes excited by sound," J. Acoust. Soc. Am. **34**, 640–647 (1962).
[39]C. Lesueur, *Rayonnement Acoustique des Structures* (Eyrolles, Paris, 1988).
[40]D. G. Crighton, "The 1988 Rayleigh medel lecture: Fluid loading - the interaction between sound and vibration," J. Sound Vib. **133**, 1–27 (1989).
[41]M. Pierucci and T. S. Graham, "A study of bending waves in fluid-loaded thick plates," J. Acoust. Soc. Am. **65**, 1190–1197 (1979).
[42]V. Cotoni and A. Le Bot, "Radiation of plane structures at high frequency using an energy method," Int. J. Acoust. Vib. **6**, 209–214 (2001).
[43]H. G. Davies, "Natural motion of a fluid-loaded semi-infinite membrane," J. Acoust. Soc. Am. **55**, 213–219 (1974).
[44]D. G. Crighton and D. Innes, "The modes, resonances and forced response of elastic structures under heavy fluid loading," Philos. Trans. R. Soc. London, Ser. A, Ser. A **312**, 291–341 (1984).
[45]M. Heckl, "The tenth Sir Richard fairey memorial lecture: Sound transmission in buildings," J. Sound Vib. **77**, 165–189 (1981).
[46]V. Cotoni, A. Le Bot, and L. Jezequel, "Sound transmission through a plate by an energy flow approach," Acust. Acta Acust. 88, 827–836 (2002).
[47]W. C. Elmore and M. A. Heald, *Physics of Waves* (Dover, New York, 1985).
[48]A. D. Pierce, *Acoustics - An Introduction to its Physical Principles and Applications.* (Acoustical Society of America, American Institute of Physics, New York, 1991).
[49]R. G. Kouyoumjian and P. H. Pathak, "A uniform geometrical theory of diffraction for an edge a perfectly conducting surface," Proc. IEEE **62**, 1448–1461 (1974).

Coupled integral equations for sound propagation above a hard ground surface with trench cuttings

Gong Li Wang and Weng Cho Chew
Center for Computational Electromagnetics and Electromagnetics Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1406 West Green Street, Urbana, Illinois 61801-2991

Michael J. White
Engineering Research and Development Center/Construction Engineering Research Laboratory, U.S. Army Corps of Engineers, P.O. Box 9005, Champaign, Illinois 61826-9005

(Received 11 June 2005; revised 20 June 2006; accepted 21 June 2006)

A set of coupled integral equations is formulated for the investigation of sound propagation from an infinitesimal harmonic line source above a hard ground surface corrugated with cuttings. Two half-space Green's functions are employed in the formulation. The first one defined for the upper half space is used to reduce the problem size and eliminate the edge effect resulting from the boundary truncation; the other one for the lower half space is to simplify the representation of the Neumann-Dirichlet map. As a result, the unknowns are only distributed over the corrugated part of the surface, which leads to substantial reduction in the size of the final linear system. The computational complexity of the Neumann-Dirichlet map is also reduced. The method is used to analyze the behavior of sound propagation above textured surfaces the impedance of which is expectedly altered. The effects of number and opening of trench cuttings, and the effect of source height are investigated. The conclusions drawn can be used for reference in a practical problem of mitigating gun blast noise. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2225997]

PACS number(s): 43.20.Fn, 43.20.El, 43.20.Bi [JBS] Pages: 1209–1216

I. INTRODUCTION

Acoustic wave propagation from an infinitesimal harmonic line source above a locally-perturbed half space has been of interest to the acoustical society due to its practical importance and its concise mathematical description. The problem arises in different areas, e.g., the study of the effect of oscillation in harbors (Hwang and Tuck, 1970; Lee, 1971; Shaw, 1971), the suppression of traffic noise using barriers (Peplow and Chandler-Wilde, 1999), and the effect of nonflat terrain on sound propagation (Alpera *et al.*, 2003), etc.

The problem addressed in this paper comes from the attempt of mitigating gun blast noise above a hard ground surface by changing it into an artificial soft (or pressure release) surface using trench cuttings. A number of studies have been conducted on using various structures to deflect sound propagation along the surface of a hard ground (Swenson *et al.*, 1992). We suggest texturing the hard surface with trench cuttings, which leads to a change in the surface impedance. If the geometry of cuttings satisfies some conditions, the surface will exhibit low impedance. As a result, more noise is deflected away from the cuttings and less is propagated along the surface, and the adverse effect on the surrounding environment is suppressed.

The use of cuttings to alter surface impedance to change the property of sound propagation is not new to the acoustical society. Tol and Holties (1999) have studied the use of low, close to track barriers and absorptive layers on a slab track to lower the noise from train traffic. Attenborough and Boulanger (2002) have investigated the effective impedance of hard rough surfaces.

The idea of impedance surface has also been used in the electromagnetic society. Readers are referred to Qian *et al.* (Qian *et al.*, 1998) for details. Readers are also referred to some textbooks or notes (Collin, 1990; Kong, 1998; Chew, 2002) for properties of corrugated surfaces.

The objective of this paper is to provide an efficient mathematical description for the problem and to preliminarily study the application of the artificial soft surface to mitigate noise traveling along a hard ground surface. As mentioned above, an infinitesimal harmonic line source is used as excitation throughout the paper, considering that the radiated field can be expressed as a summation of Fourier components. For the theoretical model to be as close as possible to the real world problem, a quasiperiodic structure is employed in which a finite number of cuttings are included instead of using an infinitely periodic structure. Plane waves are often employed in discussing the behavior of waves above a periodic structure, where the Floquet-Bloch theorem is always used to represent the scattered field. As is known that the radiation of a line source can be expressed in terms of the superposition of different plane waves, therefore the behavior of the line source can be obtained by investigating the behavior of the plane waves. For purposes of numerical computation, however, the implementation is not straightforward and some approximation may be needed (Lam, 1999).

This problem is easily formulated using the boundary integral equation method or the boundary element method. The geometry used is shown in Fig. 1, where $\Omega_1=\{(x,z)|z>0\}$ is the upper half space, and Ω_2 denotes the inside of the cutting in the hard ground surface. The boundary of Ω_1 is

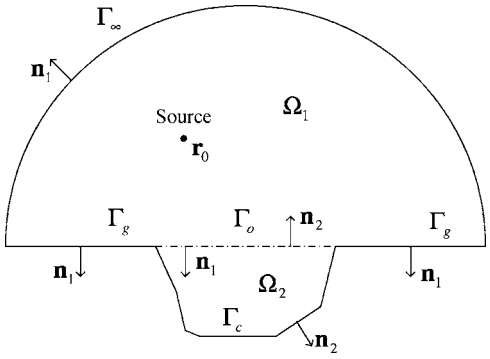


FIG. 1. Geometry of the model for the boundary integral equation formulation.

$\partial\Omega_1=\Gamma_\infty+\Gamma_g+\Gamma_o$, where Γ_∞ is the upper semicircle of infinitely large radius; Γ_o is a part of the plane $z=0$, coinciding with the opening of the cutting; $\Gamma_g=\{(x,z)|z=0\}\backslash\Gamma_o$. The boundary of Ω_2 is $\partial\Omega_2=\Gamma_o+\Gamma_c$, where Γ_c connects the two end points of Γ_o and is entirely below the plane $z=0$, denoting the sides and bottom of Ω_2. The solution domain is $\Omega=\Omega_1+\Omega_2+\Gamma_o$, the boundary of which is $\partial\Omega=\Gamma_\infty+\Gamma_g+\Gamma_c$.

Hwang and Tuck (1970) use a single-layer potential method to study the effect of wave-induced oscillations by solving for the unknown equivalent sources distributed along the boundary of the harbor (Γ_c) and coastline (Γ_g). Lee (1971) uses the Green's second theorem in both the regions inside and outside the harbor, Ω_1 and Ω_2, respectively. This method is also used by Shaw (1971). Peplow and Chandler-Wilde (1999) present a method for the case when the boundary of the solution domain is penetrable and absorptive. For simplicity, the impedance boundary condition is used and accordingly a special Green's function is employed to account for the finite acoustic impedance of the lower half space. When the boundary is rigid, their formulation reduces to that of Hwang and Tuck. Alpera *et al.* (2002) use Peplow and Chandler-Wilde's method to study the effect of nonflat terrain on sound propagation.

In the formulations proposed by Hwang and Tuck, and Peplow and Chandler-Wilde, a half-space Green's function is used in the region of upper half space Ω_1, while a free-space Green's function is used in the cutting Ω_2. The use of the half space Green's function in Ω_1 reduces the number of unknowns substantially because Γ_g, the flat part of the hard ground surface is eliminated from the integral equation, and avoids the edge effect resulting from the boundary truncation when a free-space Green's function is used in Ω_1. In our formulation, two half-space Green's functions instead of only one are involved, one is used in Ω_1, and the other is used in Ω_2. The use of the second half-space Green's function in Ω_2, as will be shown later, simplifies the representation of the Neumann-Dirichlet map, and hence reduces the complexity of its computation.

The paper is organized as follows: Sec. II describes the formulation of the coupled boundary integral equations. The difference from the previous one of Peplow and Chandler-Wilde is depicted in Sec. III. Section IV discusses the validation using an analytical method and two previous numerical methods. Section V studies the effects of number and opening of cuttings on the propagation of sound waves over quasiperiodic structures. The effect of source height is also addressed. The conclusions are drawn from the numerical analysis, which can be used for reference in designing a practical artificial soft surface to suppress gun blast noise.

II. BOUNDARY INTEGRAL EQUATION FORMULATION

The governing equation and boundary conditions for the acoustic wave propagation from an infinitesimal harmonic line source above a locally perturbed rigid plane is

$$\begin{cases} \nabla^2 p(\mathbf{r}) + k^2 p(\mathbf{r}) = -\delta(\mathbf{r}-\mathbf{r}_0), \mathbf{r}\in\Omega, \mathbf{r}_0\in\Omega_1 \\ \left.\dfrac{\partial p(\mathbf{r})}{\partial n}\right|_{\Gamma_g\cup\Gamma_c} = 0 \\ \lim\limits_{r\to\infty}\left[-ik\sqrt{r}p(\mathbf{r}) + \sqrt{r}\dfrac{\partial p(\mathbf{r})}{\partial r}\right] = 0, \end{cases} \tag{1}$$

where p is the pressure, $k=\omega/c$ is the wave number, and $\mathbf{r}_0$ is the location of the line source. The time dependence is $e^{-i\omega t}$ and is suppressed. We study the case when the source is above the ground surface, or $\mathbf{r}_0\in\Omega_1$. The unit vector $\mathbf{n}$ is the outward normal to $\Gamma_g\cup\Gamma_c$. For the case when the source is in the cutting, or $\mathbf{r}_0\in\Omega_2$, the derivation is similar.

Consider two half-space Green's functions as follows:

$$\begin{cases} \nabla^2 g_l(\mathbf{r},\mathbf{r}') + k^2 g_l(\mathbf{r},\mathbf{r}') = -\delta(\mathbf{r}-\mathbf{r}') \\ \left.\dfrac{\partial g_l(\mathbf{r},\mathbf{r}')}{\partial \gamma_l}\right|_{z=0} = 0, l=1,2, (-1)^{l-1}z\geqslant 0, (-1)^{l-1}z'\geqslant 0. \end{cases} \tag{2}$$

Clearly, $g_1(\mathbf{r},\mathbf{r}')$ is for the upper half space and $g_2(\mathbf{r},\mathbf{r}')$ for the lower half space, both satisfying Neumann condition on the plane $z=0$. Their expressions are readily obtained such that $g_l(\mathbf{r},\mathbf{r}\prime)=g_0(\mathbf{r},\mathbf{r}\prime)+g_0(\mathbf{r},\mathbf{r}_l')$, $l=1,2$, where $g_0(\mathbf{r},\mathbf{r}\prime)=i/4H_0^{(1)}(k|\mathbf{r}-\mathbf{r}\prime|)$ is the free-space Green's function. Here, $\mathbf{r}'=(x',z')$ is the location of a source, and $\mathbf{r}_l'=(x',-z')$ is the location of its image. The unit vectors $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are normal to the plane $z=0$, with $\boldsymbol{\gamma}_1$ pointing downward and $\boldsymbol{\gamma}_2$ pointing upward.

Using the vector identity $\nabla\cdot(u\mathbf{A})=u\nabla\cdot\mathbf{A}+\nabla u\cdot\mathbf{A}$ and the sift property of the delta function, we have for any point in Ω_1 (pp. 430–433, Chew, 1995)

$$p_{\text{inc}}(\mathbf{r}) = p(\mathbf{r}) - \int_{\Gamma_o} d\Gamma\, g_1(\mathbf{r},\mathbf{r}')\frac{\partial p(\mathbf{r}')}{\partial n_1'}. \tag{3}$$

Here, $\mathbf{n}_1$ is the outward normal to Ω_1. Notice that the integration over Γ_∞ vanishes due to the radiation condition in Eq. (1); over Γ_o only the single layer potential is left due to the Neumann condition in Eq. (2). Similarly, for any point in Ω_2, we have

$$0 = p(\mathbf{r}) - \int_{\Gamma_o} d\Gamma\, g_2(\mathbf{r},\mathbf{r}')\frac{\partial p(\mathbf{r}')}{\partial n_2'} + \int_{\Gamma_c} d\Gamma\, p(\mathbf{r}')\frac{\partial g_2(\mathbf{r},\mathbf{r}')}{\partial n_2'}. \tag{4}$$

Here, $\mathbf{n}_2$ is the outward normal to Ω_2. Equations (3) and (4)

are used to calculate the field in Ω_1 and Ω_2 once the equivalent sources over Γ_o and Γ_c are determined. To solve for the unknown sources, boundary integral equations have to be set up over Γ_o and Γ_c.

Consider a point, denoted by $\mathbf{r}$, on Γ_o. The part of Γ_o in the vicinity of $\mathbf{r}$ is indented above by a semicircle centered about $\mathbf{r}$, denoted by Γ_ε. Then $\mathbf{r}$ is outside of $\Omega_1 \backslash \Omega_\varepsilon$, Ω_ε being the area enclosed by Γ_ε and Γ_o. According to the extinction theorem (p. 432, Chew, 1995)

$$p_{\text{inc}}(\mathbf{r}) = \lim_{\varepsilon\to 0}\int_{\Gamma_\varepsilon} d\Gamma p(\mathbf{r}')\frac{\partial g_1(\mathbf{r},\mathbf{r}')}{\partial n_1'} - \lim_{\varepsilon\to 0}\int_{\Gamma_\varepsilon} d\Gamma g_1(\mathbf{r},\mathbf{r}')\frac{\partial p(\mathbf{r}')}{\partial n_1'} - \lim_{\varepsilon\to 0}\int_{\Gamma_o\backslash\Gamma_\varepsilon} d\Gamma g_1(\mathbf{r},\mathbf{r}')\frac{\partial p(\mathbf{r}')}{\partial n_1'}. \tag{5}$$

For the first term, when $z'\to 0$, the source and its image coincide with each other; therefore $g_1=2g_0$. It is readily shown that it reduces to $p(\mathbf{r})$ by using the asymptotic property of $H_1^{(1)}(k|\mathbf{r}-\mathbf{r}'|)$. The second term vanishes because $\partial p(\mathbf{r})/\partial\varepsilon$ is continuous over Γ_ε. The third term is a principal value integral. Consequently Eq. (5) is simplified as

$$p_{\text{inc}}(\mathbf{r}) = p(\mathbf{r}) - P.V.\int_{\Gamma_o} d\Gamma g_1(\mathbf{r},\mathbf{r}')\frac{\partial p(\mathbf{r}')}{\partial n_1'}. \tag{6}$$

When the part of Γ_o in the vicinity of the point $\mathbf{r}$ is indented below, $\mathbf{r}$ is outside of $\Omega_2\backslash\Omega_\varepsilon$, then we have by the same token

$$0 = p(\mathbf{r}) - P.V.\int_{\Gamma_o} d\Gamma g_2(\mathbf{r},\mathbf{r}')\frac{\partial p(\mathbf{r}')}{\partial n_2'} + \int_{\Gamma_c} d\Gamma p(\mathbf{r}')\frac{\partial g_2(\mathbf{r},\mathbf{r}')}{\partial n_2'}. \tag{7}$$

Similarly, for a point on Γ_c

$$0 = \frac{1}{2}p(\mathbf{r}) - \int_{\Gamma_o} d\Gamma g_2(\mathbf{r},\mathbf{r}')\frac{\partial p(\mathbf{r}')}{\partial n_2'} + P.V.\int_{\Gamma_c} d\Gamma p(\mathbf{r}')\frac{\partial g_2(\mathbf{r},\mathbf{r}')}{\partial n_2'}. \tag{8}$$

Notice that the constant factors before $p(\mathbf{r})$ in Eqs. (7) and (8) are different. The reason is that for $p(\mathbf{r})$ in Eq. (8), the contribution from the image source vanishes when $\varepsilon\to 0$. When Γ_c is not smooth at some points, the constant factor in Eq. (8) will be $\alpha/2\pi$ (El-Zafrany, 1993), where α is the angle subtended by the indented part relative to $\mathbf{r}$.

Using the operator representation for single layer potential and double layer potential similar to those introduced by Peplow and Chandler-Wilde (1999), Eqs. (6)–(8) are rewritten as

$$p_{o,\text{inc}} = p_o - \mathcal{S}_{oo}^1\frac{\partial p_o}{\partial n_1}, \tag{9a}$$

$$0 = p_o - \mathcal{S}_{oo}^2\frac{\partial p_o}{\partial n_2} + \mathcal{K}_{co}^2 p_c, \tag{9b}$$

$$0 = \frac{1}{2}p_c - \mathcal{S}_{oc}^2\frac{\partial p_o}{\partial n_2} + \mathcal{K}_{cc}^2 p_c. \tag{9c}$$

Here, $p_j, \partial p_j/\partial n_l \in L_2(\Gamma_j)$, $j=o,c$, $l=1,2$. The single layer potential is defined as

$$\mathcal{S}_{jk}^l[\ ] = \int_{\Gamma_j} d\Gamma g_l(\mathbf{r},\mathbf{r}')[\ ], \quad \mathbf{r}\in\Gamma_k, \quad l=1,2, \quad j,k=o,c; \tag{10}$$

and the double layer potential is defined as

$$\mathcal{K}_{jk}^l[\ ] = \int_{\Gamma_j} d\Gamma \frac{\partial g_l(\mathbf{r},\mathbf{r}')}{\partial n_l'}[\ ], \quad \mathbf{r}\in\Gamma_k, \quad l=1,2, \quad j,k=o,c. \tag{11}$$

When $j=k$, the two potentials are understood under the meaning of principal value integral.

By inspection, it is seen that from Eqs. (9b) and (9c)

$$p_o = \mathcal{Z}_2\frac{\partial p_o}{\partial n_2}, \tag{12}$$

where

$$\mathcal{Z}_2 = \left[\mathcal{S}_{oo}^2 - \mathcal{K}_{co}^2\left(\mathcal{K}_{cc}^2 + \frac{1}{2}I\right)^{-1}\mathcal{S}_{oc}^2\right]. \tag{13}$$

Here, $\mathcal{Z}_2$ is an impedance operator or a Neumann-Dirichlet map projecting the derivative of a function to itself; I is a unit operator mapping a function to itself. The operator $\mathcal{S}_{oo}^1$ in Eq. (9a) is also an impedance operator. Denoting it as $\mathcal{Z}_1$, and substituting Eq. (12) in Eq. (9a), we have

$$p_{o,\text{inc}} = (\mathcal{Z}_1 + \mathcal{Z}_2)\frac{\partial p_o}{\partial n_2}, \tag{14}$$

where the condition of $\mathbf{n}_1=-\mathbf{n}_2$ on Γ_o is used.

Albeit the formulation is derived for one cutting, the application to multiple cuttings is straightforward. Suppose the number of cuttings is P, then $\Gamma_o=\cup_{j=1}^P\Gamma_{o_j}$, $\Gamma_c=\cup_{j=1}^P\Gamma_{c_j}$. The nicety here is that the Neumann-Dirichlet map can be evaluated for each cutting separately before solving Eq. (14) because there is no direct coupling between different cuttings. If all the cuttings have the same geometry, the evaluation is done only once, so the workload is reduced substantially. Equations (9a)–(9c) can also be solved directly; this however results in a larger workload. Suppose that after discretization, there are N elements for the opening of each cutting, and M elements for the two sides and floor of each cutting, then the total number of unknowns for the three coupled equations is $P(2N+M)$, while the number of unknowns for Eq. (14) is PN. Their ratio is thus $\beta=2+M/N$, and the ratio of their computational complexity is $O(\beta^2)$ if an iterative solver is used or $O(\beta^3)$ if a direct solver is used.

In the numerical implementation, Γ_o and Γ_c are first partitioned into boundary elements: $\Gamma_o=\cup_{j=1}^N e_{o,j}$, $\Gamma_c=\cup_{j=1}^M e_{c,j}$, N and M being the numbers of elements on Γ_o and Γ_c, re-

spectively; then a pulse basis function is associated with each element. The discretized integral equations are then tested at the center of each element. The solution of the resulting linear system includes two steps: (1) the inverse of the discretized counterpart of $K_{cc}^2+I/2$ is first evaluated for each cutting using the Gaussian elimination method; (2) the linear system resulting from Eq. (9a) is then solved by the same method or by the conjugate gradient method.

In the discretization, two kinds of integrals need to be evaluated, viz.,

$$\int_{e_k} g_1(\mathbf{r}_j,\mathbf{r}')d\Gamma, \quad \mathbf{r}_j \in e_j, \quad e_j,e_k \subset \Gamma_o, \tag{15a}$$

$$\int_{e_k} g_2(\mathbf{r}_j,\mathbf{r}')d\Gamma, \quad \mathbf{r}_j \in e_j, \quad e_j,e_k \subset \Gamma_o \cup \Gamma_c. \tag{15b}$$

When $\mathbf{r}_j$ and $\mathbf{r}'$ are far away from each other, the midpoint rule is used in the evaluation; when they are close to each other, to guarantee the accuracy, either a high-order Gaussian quadrature rule is used; or as an alternative, one element is first partitioned into a number of small pieces, then a lower-order Gaussian quadrature rule is used for each piece.

III. DIFFERENCE FROM THE FORMULATION OF PEPLOW AND CHANDLER-WILDE

Without loss of generality, we consider the case with only one cutting. The boundary integral equations presented by Peplow and Chandler-Wilde (1999) are listed as follows:

$$p_{o,\mathrm{inc}} = p_o - S_{oo}^1 \frac{\partial p_o}{\partial n_1}, \tag{16a}$$

$$0 = \frac{1}{2}p_o - S_{oo}^0 \frac{\partial p_o}{\partial n_2} + K_{co}^0 p_c, \tag{16b}$$

$$0 = \frac{1}{2}p_c + K_{oc}^0 p_o - S_{oc}^0 \frac{\partial p_o}{\partial n_2} + K_{cc}^o p_c, \tag{16c}$$

where the superscript 0 of S,K in Eqs. (16a)–(16c) means that the underlying Green's function is that of free space g_0. The Neumann-Dirichlet map is derived as

$$Z_2' = \left[\frac{1}{2}I - K_{co}^0\left(K_{cc}^0 + \frac{1}{2}I\right)^{-1}K_{oc}^0\right]^{-1} \times \left[S_{oo}^0 - K_{co}^0\left(K_{cc}^0 + \frac{1}{2}I\right)^{-1}S_{oc}^0\right]. \tag{17}$$

Here, the prime is used to distinguish from that in Eq. (13). Equation (16a) is then simplified as

$$p_{o,\mathrm{inc}} = (Z_1 + Z_2')\frac{\partial p_o}{\partial n_2}. \tag{18}$$

Suppose that the numbers of unknowns over Γ_c and Γ_o are M and N, respectively, then the operation for the evaluation of discretized Z_2 in Eq. (13) is $O(NM^2+N^2M+M^3)$, and the total operation for solving $\partial p_o/\partial n_2$ using our formulation is $O(NM^2+N^2M+M^3+N^3)$. When the formulation of Peplow and Chandler-Wilde is used, an additional operation of $O(N^2M+2N^3)$ is needed to evaluate the inverse of the operator $1/2I-K_{co}^0(K_{cc}^0+1/2I)^{-1}K_{oc}^0$ and the product of the two terms in the square brackets in Eq. (17). If $N \ll M$, the difference is small, but when $N \gg M$, the difference will be large. When N is comparable to M, or $N \approx M$, the cost using their formulation is about 1.75 times that using our formulation.

IV. ANALYTICAL AND NUMERICAL VALIDATION

As an analytical validation, consider an extreme case in which the cutting is rectangular and the depth d is tending to zero. The limit of this case is the scattering from a plain hard ground surface the solution of which is $g_1(\mathbf{r},\mathbf{r}_o)$ or p_{inc}.

When $d \to 0$, only the floor of the cutting contributes, thus the two sides are excluded from Γ_c. The double-layer contribution in Eq. (9b) is rewritten as

$$K_{co}^2 p_c = \int_{\Gamma_c} \frac{\partial g_0(\mathbf{r},\mathbf{r}')}{\partial n_2'} p_c(\mathbf{r}')d\Gamma + \int_{\Gamma_c'} \frac{\partial g_I(\mathbf{r},\mathbf{r}')}{\partial \tilde{n}_2'} p_c(\mathbf{r}')d\Gamma, \tag{19}$$

where Γ_c' is the image of Γ_c; $g_I(\mathbf{r},\mathbf{r}')=g_0(\mathbf{r},\mathbf{r}_I')$ is the Green's function for the image source, $\mathbf{r}'=(x',z')$, $\mathbf{r}_I'=(x',-z')$; $\tilde{\mathbf{n}}_2$ is the image of $\mathbf{n}_2$, $\mathbf{n}_2=n_{2z}\hat{z}$, $\tilde{\mathbf{n}}_2=-n_{2z}\hat{z}$. All the images are about the plane $z=0$. When $d \to 0$, $\mathbf{r}$ approaches Γ_c from above, and approaches Γ_c' from below. In other words, $\mathbf{r}$ approaches the two double-layers both from the side where the negative charges are distributed, therefore

$$K_{co}^2 p_c \to -\frac{1}{2}p_c(\mathbf{r}) - \frac{1}{2}p_c(\mathbf{r}) = -p_c(\mathbf{r}), \quad d \to 0 \tag{20}$$

according to the jump condition for double-layer potential (Colton and Kress, 1983). For the double-layer contribution $K_{cc}^2 p_c$ in Eq. (9c), the same argument applies. However, now $\mathbf{r}$ is already on Γ_c. Therefore, we readily obtain $\partial H_0^{(1)}(kr)/\partial n_2=0$. As a result, the contribution from Γ_c vanishes, only Γ_c' contributes

$$K_{cc}^2 p_c \to -\frac{1}{2}p_c(\mathbf{r}), \quad d \to 0. \tag{21}$$

For the single-layer potential $S_{oc}^2 \partial p_o/\partial n_2$ in Eq. (9c), when $d \to 0$, $S_{oc}^2 \to S_{oo}^2$ according to the jump condition for single-layer potential (Colton and Kress, 1983), therefore, no additional term results.

According to Eqs. (21) and (9c), $S_{oo}^2 \partial p_o/\partial n_2=0$, which immediately leads to

$$\frac{\partial p_o}{\partial n_2} = 0 \tag{22}$$

by judiciously assuming that operator S_{oo}^2 is not singular. Substituting Eqs. (20) and (22) into Eq. (9b), we have $p_o=p_c$. Substituting Eq. (22) into Eqs. (3) and (9a), it follows that $p=p_{\mathrm{inc}}$ everywhere in $\bar{\Omega}_1(=\Omega_1 \cup \partial\Omega_1)$.

Next we use the following integral equation for a more general validation:

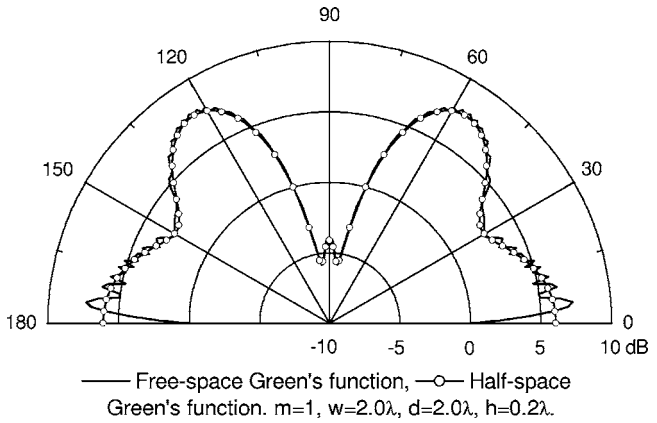


FIG. 2. Radiation pattern above one cutting computed with the present formulation compared to that computed using Eq. (23). Here, m is the number of cuttings, w is the opening size, d is the depth of cutting, h is the source height.

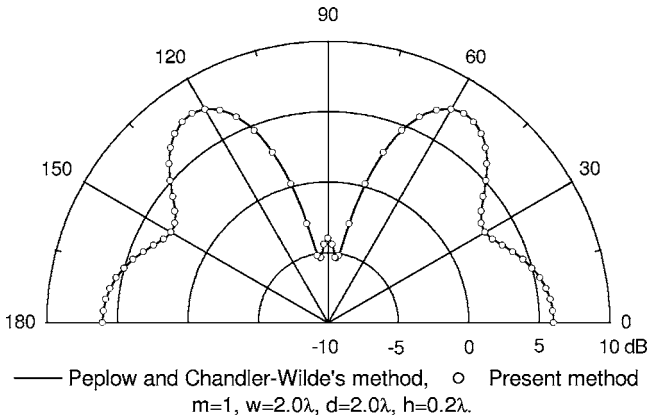


FIG. 3. Radiation pattern above one cutting computed with the present formulation compared to that computed using the formulation of Peplow and Chandler-Wilde. Here, m, w, d, and h are of the same meaning as in Fig. 2.

$$p_{\text{inc}}(\mathbf{r}) = \frac{1}{2}p(\mathbf{r}) - P\,.\,V\,.\int_{\Gamma_g \cup \Gamma_c} \frac{\partial g_0(\mathbf{r},\mathbf{r}')}{\partial n'} p(\mathbf{r}')d\Gamma, \quad (23)$$

which is obtained by applying the Green's second theorem in the entire solution domain Ω using the free space Green's function g_0. The unit vector $\mathbf{n}$ is the outward normal to the boundary $\Gamma_g \cup \Gamma_c$.

Figure 2 shows the radiation pattern over a corrugated surface computed with the new formulation. The radiation pattern is defined by

$$\text{RP} = 20\log_{10}\left|\frac{p(\mathbf{r})}{p_0(\mathbf{r})}\right|, \quad r \to \infty, \quad (24)$$

where $p_0(\mathbf{r}) = g_0(\mathbf{r},\mathbf{r}')$, $p(\mathbf{r})$ is given by Eq. (3). When $r \to \infty$, we have

$$\frac{p(\mathbf{r})}{p_0(\mathbf{r})} = 2\cos(k_z h) - 2\int_{\Gamma_o} d\Gamma e^{-i\mathbf{k}\cdot\mathbf{r}'}\frac{\partial p_o(\mathbf{r}')}{\partial n_2'}, \quad r \to \infty, \quad (25)$$

where it is assumed that $\mathbf{r}'=0$ in $g_0(\mathbf{r},\mathbf{r}')$ while $\mathbf{r}_0=(0,h)$ in Eq. (1), h being the height of the source above the ground; $\mathbf{k}=k_x\hat{x}+k_z\hat{z}$, $k_x=k\cos\theta$, $k_z=k\sin\theta$, $0\leqslant\theta\leqslant\pi$.

The main difference from the excess attenuation (Peplow and Chandler-Wilde, 1999) is that the radiation pattern is only for the far field, while the excess attenuation is for all field points. The other difference is that there is a minus sign before the excess attenuation expression.

Only one cutting is included in the model in Fig. 2. Its opening and depth are both 2.0λ. The source is 0.2λ above the opening center. When Eq. (23) is used, Γ_g is truncated at a distance of 64λ from the opening center on both sides. The element size for both methods is 0.1λ. When evaluating the integrals in Eqs. (15a) and (15b), each element is subsectioned into three pieces with equal size, and the midpoint rule is then used for each piece. It is seen that there are small ripples over the radiation pattern computed with Eq. (23). Also, appreciable deflections are observed near the surface. The ripples and deflections, called edge effect, have been shown to come from the interference of the radiation from the two ends of the truncated surface. Numerical experiments indicate that to suppress the edge effect, an extremely large truncated surface is needed, which increases the problem size and lowers the execution efficiency. In the radiation pattern computed with the coupled equations [Eqs. (9a)–(9c)], no ripple is observed, and the curve looks quite smooth. Furthermore, the inward deflection near the surface is eliminated. Note that the radiation pattern is a semicircle because the radiation is limited to the upper half space.

Finally we validate the present method using the formulation of Peplow and Chandler-Wilde [Eqs. (16a)–(16c)]. The same model is used as above. The element size for both methods is also 0.1λ, and the same integration method as above is applied. The results are shown in Fig. 3. Both methods use the half-space Green's function in the air region, so none of them suffers from the edge effect, and the agreement between the two results is excellent.

V. SOUND PROPAGATION ABOVE A HARD GROUND SURFACE WITH CUTTINGS

By incorporating a special texture on an acoustically hard surface, the acoustic impedance of the surface can be changed to meet some special requirements arising in practical applications (Tol and Holties, 1999; Attenborough and Boulanger, 2002). The alteration of surface impedance is generally obtained by cutting a periodic structure into the surface (Collin, 1990; Kong, 1998; Chew, 2002). The opening of the cuttings is made much less than the wavelength of the applied field so that only the dominant mode is propagating toward the floor, while others are evanescent. Therefore, if the cutting depth is chosen such that

$$kd = n\pi + \frac{\pi}{2}, \quad n = 0,1,\ldots, \quad (26)$$

the open circuit at the floor is transformed into a short circuit at the opening by the depth of cuttings. Generally, the depth is chosen as one-quarter wavelength. In this way, the impedance at the opening of cuttings is very low, and the opening behaves like a soft surface. This depth is referred to as critical depth.

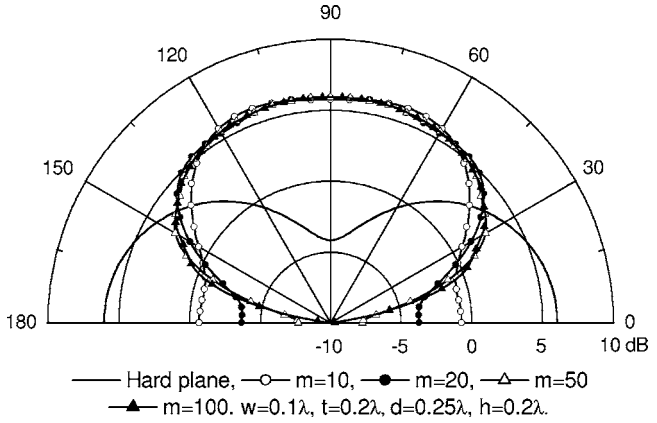


FIG. 4. Radiation pattern above corrugated planes with different numbers of cuttings. Here, *m*, *w*, *d*, and *h* are of the same meaning as in Fig. 2; *t* is the period of the quasiperiodic structure formed by cuttings.

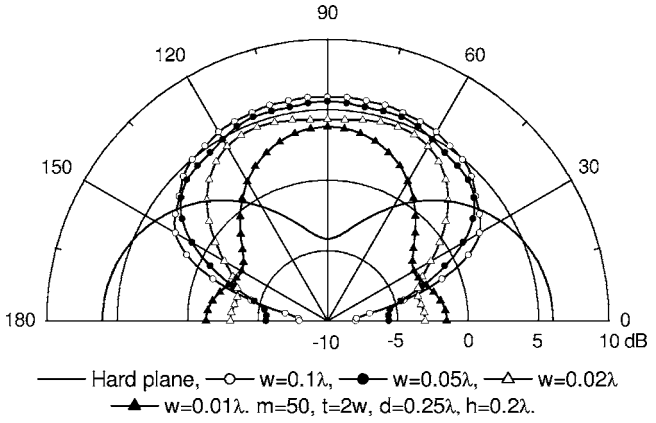


FIG. 6. Radiation pattern above corrugated planes with different cutting openings. Here, *m*, *w*, *t*, *d*, and *h* are of the same meaning as in Fig. 4.

The soft surface exhibits low impedance by accommodating energy flow freely into or out of the cuttings. This immediately means that less energy will flow along the surface. In other words, at the depths other than the critical depth or when there are no corrugations at the surface, much more energy will flow along the surface to infinity. In terms of impedance, at the critical depth, the surface impedance is lower than when at other depths or when the surface is not textured.

Figure 4 shows the effect of the number of cuttings on the radiation pattern. The source is located 0.2λ above the surface. It is seen that when the number of cuttings increases, the radiation pattern tends more closely to the one above a plain soft surface. For the sake of visibility, only the numbers of up to 100 are shown here. In fact, when the number is quite large, say larger than 100, only small differences are observed between the results for different numbers. Therefore it is unnecessary to use a large number of cuttings to achieve a good suppression effect.

Shown in Fig. 5 is the distribution of a pressure field on a corrugated surface. It corresponds to the case in Fig. 4 when there are 50 cuttings in the surface. The result is denoted by the line with circles on it. The line without circles is for the radiation pattern of the power flow when there is no cutting, which is calculated analytically. Clearly the field attenuates much faster above the corrugated surface than above the plain hard surface. This indicates that by choosing suitable geometric parameters for cuttings, a hard surface does exhibit low acoustic impedance so that sound propagation to distance along the surface is suppressed.

Figure 6 is for the effect of opening size of cuttings on the radiation pattern. In these results, the number of cuttings is fixed for different openings. We see that the mitigation effect near the surface is weakened as the opening size is decreased. If the frequency to be suppressed is 50 Hz, then 0.1λ is about 70 cm, too large to be applicable. Therefore using cuttings of small opening is of practical importance. To balance the unfavorable tendency when opening is small, one possibility is to increase the number of cuttings as indicated in Fig. 4.

Figure 7 shows the radiation patterns of a source of different heights above a corrugated surface. The corrugation is formed by making 50 cuttings into a hard surface. As above a plain hard surface, the higher the source height, the more complex the radiation pattern. Notice that the radiation intensity in the lateral direction becomes increasingly large when the source moves up gradually. This unfortunately shows that

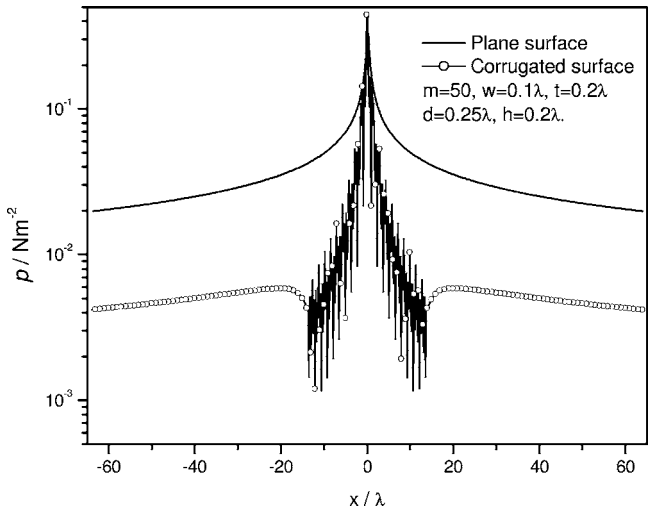


FIG. 5. Pressure field distribution on a corrugated surface. Here, *m*, *w*, *t*, *d*, and *h* are of the same meaning as in Fig. 4.

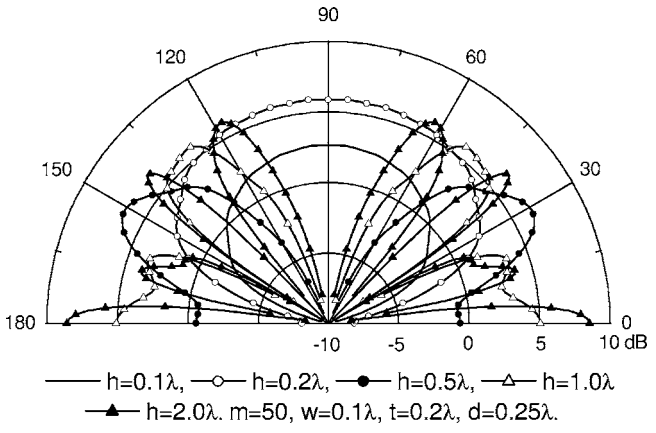


FIG. 7. Radiation pattern for different heights of source above a corrugated surface. Here, *m*, *w*, *t*, *d*, and *h* are of the same meaning as in Fig. 4.

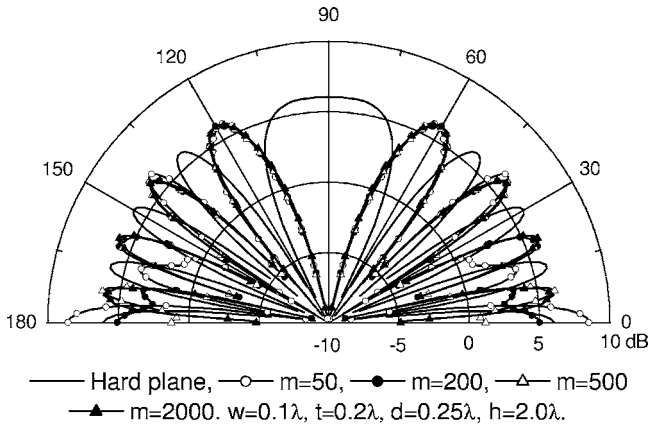


FIG. 8. Radiation pattern of a high source for different numbers of cuttings above a corrugated surface. Here, *m*, *w*, *t*, *d*, and *h* are of the same meaning as in Fig. 4.

the corrugated surface no longer acts as a soft surface for a high source. The situation is expected to be improved by increasing the number of cuttings. Shown in Fig. 8 are the radiation patterns of a source 2λ above a surface with different numbers of cuttings. Note that 2λ is the largest height used in Fig. 7. We see that increasing the number of cuttings does improve the radiation pattern in the lateral direction. In fact, the radiation pattern for 2000 cuttings is almost the same as that above a plain soft plane.

The above examples show that the radiation pattern of a line source is sensitive to both cutting geometry and source height. The reason is explained as follows. Consider the field distribution on a plain hard surface. It is known that the magnitude of the field is peaked at the point directly below the source and is lowered with the distance from the source. When the source is close to the surface, the field varies rapidly along the surface, and much energy is localized around the peak point. Even a small change around the peak point at the surface will have a strong influence on the field distribution. It is thus understood that a small number of cuttings are enough to have an appreciable mitigation effect. When the source is higher, the field varies more slowly, so the energy is distributed more evenly along the surface. Then the influence of the change around the peak point on the field is weakened. To have a strong enough mitigation effect, a large number of cuttings are therefore needed.

Generally speaking, when the source is high, the angle subtended by the cuttings to the source is reduced. Then the total impedance of the corrugated surface is increased, and the surface behaves more like a plain hard surface. To keep the surface impedance at a low level, the number of cuttings needs to be increased to enlarge the corrugation extent. Considering that the noise spectrum is broadband, to sufficiently suppress high frequency components, a large number of cuttings are needed. At high frequencies, a source is of acoustically large height though its physical height remains the same.

VI. CONCLUSIONS

A set of coupled integral equations using two half-space Green's functions is formulated to study sound propagation above a hard surface textured with cuttings. A Neumann-Dirichlet map is introduced over cutting openings. The resulting algorithm is more efficient than the one presented by Peplow and Chandler-Wilde. The use of the Neumann-Dirichlet map reduces the computational complexity as opposed to the methods where the coupled equations are solved directly, especially when multiple cuttings are studied.

The effect of cutting geometry on sound propagation is quantified with the new integral equation method. It turns out that the radiation pattern of a source sitting above a textured surface is strongly affected by both the number and opening of cuttings. The results show that 50 cuttings with an opening of 0.1λ are sufficient to create an artificial soft surface when the source is close to the surface. For practicability, the opening has to be made as small as possible. However, the radiation pattern is slightly affected when the opening is small. The number of cuttings has to increase for the surface impedance to stay at a low level.

Numerical results indicate that the radiation pattern is also a strong function of source height. More and more variations are observed in the radiation pattern with the increase in the source height. More nulls and peaks can be seen and the envelope becomes complicated. On the other hand, the effect of cuttings is weakened when the source moves away from the surface. This tendency can be balanced by increasing the number of cuttings (Chew, 2001; Kinsler, 2000; Sievenpiper, 1999).

ACKNOWLEDGMENT

This work was supported by CERL Grant No. ARMY W9132T-04-C-0019.

Alpera, S. Q., Jacobsen, F., Juhl, P. M., and Henriquez, V. C. (**2003**). "A BEM approach to validate a model for predicting sound propagation over non-flat terrain," Appl. Acoust. **64**, 781–791.

Attenborough, K., and Boulanger, P. (**2002**). "Effective impedance of hard rough surfaces," J. Acoust. Soc. Am. **112**, 2215.

Chew, W. C. (**1995**). *Waves and Fields in Inhomogeneous Media* (IEEE, New York).

Chew, W. C. (2002). ECE431 notes of University of Illinois at Urbana-Champaign, Urbana, Illinois.

Chew, W. C., Jin, J.-M., Michielssen, E., and Song, J. (**2001**). *Fast and Efficient Algorithms in Computational Electromagnetics* (Artech House, Boston and London).

Collin, R. E. (**1990**). *Field Theory of Guided Waves*, 2nd ed. (Wiley-IEEE, New York).

Colton, D., and Kress, R. (**1983**) *Integral Equation Methods in Scattering Theory* (Wiley-Interscience, New York).

El-Zafrany, A. (**1993**). *Techniques of the Boundary Element Method* (Ellis Horwood, New York).

Hwang, L.-S., and Tuck, E. O. (**1970**). "On the oscillations of harbours of arbitrary shape," J. Fluid Mech. **42**, 447–464.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**2000**). *Fundamentals of Acoustics*, 4th ed. (Wiley, New York).

Kong, J. A. (**1998**). *Electromagnetic Wave Theory* (EMW, Cambridge, Massachusetts).

Lam, Y. W. (**1999**). "A boundary integral formulation for the prediction of acoustic scattering from periodic structures," J. Acoust. Soc. Am. **105**, 762–769.

Lee, J.-J. (**1971**). "Wave induced oscillations in harbours of arbitrary geometry," J. Fluid Mech. **45**, 372–394.

Peplow, A. T., and Chandler-Wilde, S. N. (**1999**). "Noise propagation from a cutting of arbitrary cross-section and impedance," J. Sound Vib. **223**, 355–378.

Qian, Y., Sievenpiper, D., Radisic, V., Yablonovitch, E., and Itoh, T. (**1998**).

"A novel approach for gain and bandwidth enhancement of patch antennas," *IEEE Radio and Wireless Conference Dig*. Colorado Springs, CO, p. 221.

Shaw, R. P. (**1971**). "Long period forced oscillations," in *Topics in Ocean Engineering*, edited by C. I. Bretschneider (Gulf, Houston), pp. 29–40.

Sievenpiper, D., Zhang, L., Broas, R. F. J., Alexópolous, N. G., and Yablonovitch, E. (**1999**). "High_impedance electromagnetic surfaces with a forbidden frequency band," IEEE Trans. Microwave Theory Tech. **47**, 2059–2074.

Swenson, G. W. Jr., Sandeen, E. R., Pater, L. L., and Zhuang, H. C. (1992). "The potential for mitigation of gun blast noise through sheltering of the source," USACERL Technical Report No. N-92/09.

Tol, P. F. van, and Holties, H. A. (**1999**). "A reciprocal method to evaluate low, close to track noise barriers," J. Acoust. Soc. Am. **105**, 949.

An experimental investigation of guided wave propagation in corrugated plates showing stop bands and pass bands

Tribikram Kundu[a] and Sourav Banerjee[b]
Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, Arizona 85721

Kumar V. Jata[c]
Air Force Research Laboratory, AFRL/MLL, Wright Patterson Air Force Base, Ohio 45433

(Received 4 January 2006; revised 11 June 2006; accepted 13 June 2006)

Nonplanar surfaces are often encountered in engineering structures. In aerospace structures, periodically corrugated boundaries are formed by friction-stir-welding. In civil engineering structures, rebars used in reinforced concrete beams and slabs have periodic surface. Periodic structures are also being used to create desired acoustic band gaps. For health monitoring of these structures, a good understanding of the elastic wave propagation through such periodic structures is necessary. Although a number of research papers on the wave propagation in periodic structures are available in the literature, no one experimentally investigated the guided wave propagation through plates with periodic boundaries and compared the experimental results with theoretical predictions as done in this paper. The experimental results clearly show that elastic waves can propagate through the corrugated plate (waveguide) for certain frequencies called "pass bands," and find it difficult to propagate for some other frequencies called "stop bands." Stop bands are found to increase with the degree of corrugation. Experimental results are compared with the theoretical predictions, and good matching is observed for plates with small degree of corrugation. Only two parameters—the depth of corrugation and the wavelength of the periodicity—are sufficient for modeling the elastic wave propagation in slightly corrugated plates. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2221534]

PACS number(s): 43.20.Fn, 43.20.El, 43.20.Ye [LLT] Pages: 1217–1226

I. INTRODUCTION

The problem of elastic wave propagation in periodic structures has been investigated for over five decades. Brillouin wrote the classical book on this subject (1946). Dynamics of a wide variety of periodic structures has been presented in this book. Later, Mead and his co-workers (Mead, 1970, 1975, 1976, 1986; Mead and Markus, 1983; Mead and Bardell, 1987; Mead and Yaman, 1991) made significant contributions in this field of research. In these works, Mead *et al.* solved the elastodynamic problems involving periodically supported beams (Mead, 1970; Mead and Markus, 1983), periodic damped plates (Mead, 1976), damped plates with stiffeners (Mead, 1986; Mead and Yaman, 1991), and thin cylindrical shells with periodic circumferential stiffeners (Mead and Bardell, 1987). Like many other engineering problems, periodic structure problems have been also solved by the finite element method (Oris and Petyt, 1974). Following Brillouin's classical approach, recently Ruzzene and Baz (2000) analytically solved the one-dimensional problem—composite rods with shape memory alloy inserts, periodically embedded in the base material of the rod. Interested readers are referred to the article by Mester and Benaroya (1995) for a comprehensive review of wave propagation problems in periodic and near-periodic structures.

A common feature of the elastic wave propagation in periodic structures is the existence of distinct frequency bands—some of which allow wave propagation and others do not. Those frequencies, for which the waves can propagate through the structure, are called pass band frequencies, and other frequencies for which the waves are attenuated in the structure are called stop band frequencies or forbidden frequency bands (Vasseur *et al.*, 1998).

In none of the articles referred to above has the elastic wave propagation through free plates with periodic surface boundaries been analyzed. In the above papers, the periodicity inside the materials or in the support condition has been considered. For example, Brillouin (1946) in his classical book presented the solution of wave propagation problem through one-, two-, and three-dimensional lattices of point masses with various degrees of complexity, Vasseur *et al.* (1998) studied the wave transmission through two-dimensional binary solid/solid composite media composed of arrays of Duralumin cylindrical inclusions embedded in an epoxy resin matrix, Ruzzene and Baz (2000) solved the one-dimensional problem of a composite rod with periodical insertions.

The problem of wave propagation in structures made of homogeneous materials, but having nonplanar boundaries and interfaces, has been the topic of investigation in the last three decades (Nayfeh *et al.*, 1978; Boström, 1983, 1989; Sandström, 1986; Fokkema, 1980; Glass *et al.*, 1983; El-Bahrawy, 1994a, 1994b; Banerjee and Kundu, 2004; Declercq *et al.*, 2005). Stop bands and pass bands of the

[a]Electronic mail: tkundu@email.arizona.edu
[b]Electronic mail: sourav@email.arizona.edu
[c]Electronic mail: kumar.jata@wpafb.af.mil

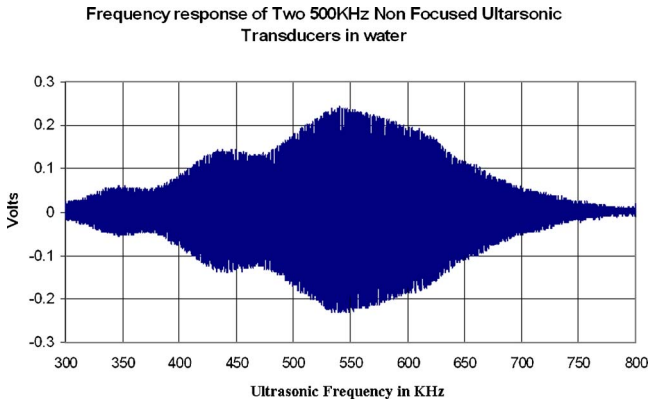


FIG. 1. (Color online) Received signal amplitude variation with frequency for the transmitter-receiver placed face to face.

Rayleigh-Lamb symmetric modes in sinusoidally corrugated waveguides have been studied by El-Bahrawy (1994a). Only recently have generalized dispersion equations for periodically corrugated waveguides been studied, and solutions for both symmetric and antisymmetric modes in a sinusoidally corrugated waveguide been presented (Banerjee and Kundu, 2006).

Although a number of theoretical papers have been published on elastic wave propagation in periodic structures, as mentioned above, very few experimental papers are available on this topic. The work of Vasseur *et al.* (1998) is the only two-dimensional experimental investigation available today. To the best of our knowledge, no investigator has yet experimentally measured the stop band and pass band frequencies in corrugated plates and compared the experimental results with the theoretical predictions as done in this paper.

II. EXPERIMENT

A. Transducer characterization

Two 1 in. diameter ultrasonic transducers were placed face to face. One transducer was excited by a signal frequency that continuously varied from 300 kHz to 800 kHz, while the second transducer recorded the received signal. The recorded signal is shown in Fig. 1. Note that the transducer resonance frequency is close to 540 kHz, although the transducers were labeled as having 500 kHz resonance frequency.

B. Specimens

Three aluminum plates were machined to produce three specimens with three different degrees of corrugation. A typical specimen is shown in Fig. 2. Figure 2(a) shows the full plate, Fig. 2(b) shows the side view of the corrugation, and Fig. 2(c) shows the period of corrugation D, highest plate thickness H_1, and lowest plate thickness H_2 in the corrugated region. Note that the average plate thickness $(2h)$ in the corrugated region is equal to $(H_1+H_2)/2$, and the corrugation depth $\varepsilon=(H_1-H_2)/4$. These dimensions for the three specimens are given in Table I.

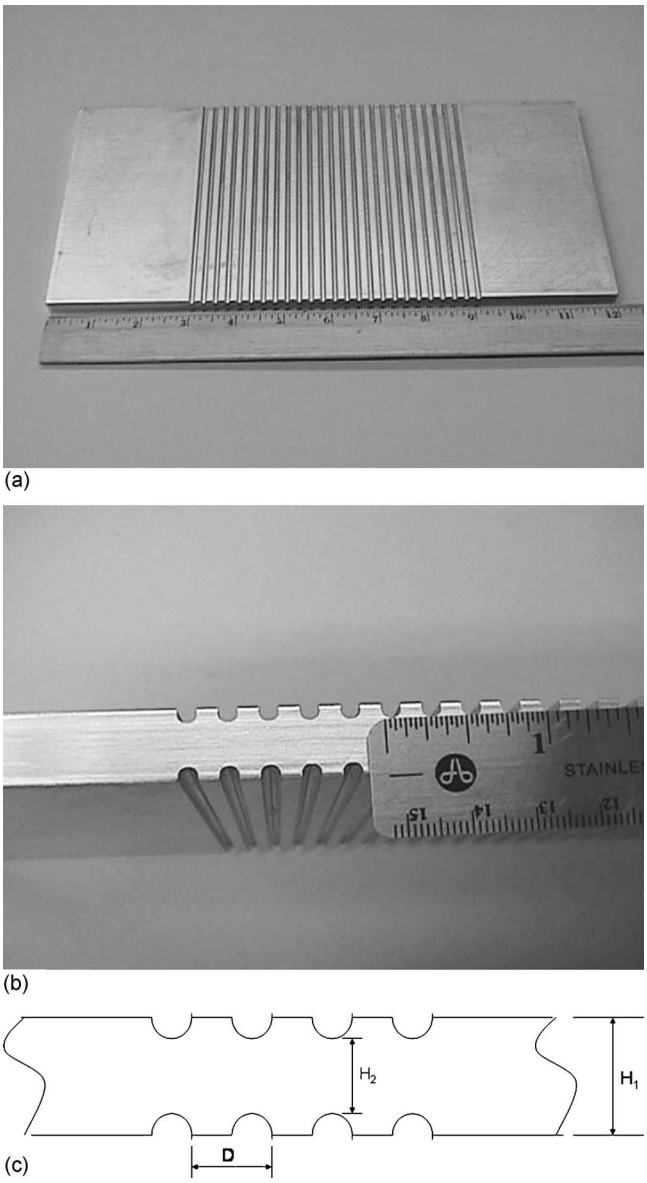


FIG. 2. Corrugated plate: (a) Top view, (b) side view, and (c) side view showing different dimensions.

C. Experimental setup

Two transducers are placed in the pitch-catch arrangement over the aluminum plate as shown in Fig. 3. Transducer T acts as the transmitter and the second transducer R acts as the receiver. Two transducers are inclined at an angle θ (clockwise and counterclockwise) with respect to the vertical axis as shown. The transducers are placed at a face to face distance of "d" and a height "h" above the aluminum plate. Transducers and the plate are immersed in water, which acts as the coupling fluid between the transducers and the plate—so that the ultrasonic energy can easily propagate from the transmitter to the plate and from the plate to the receiver. If the distance d is set such that the direct reflected beam (shown by dashed line in Fig. 3) cannot reach the receiver, then the ultrasonic energy must propagate through the plate for a length g_1 as the guided wave (shown by the bold arrow in Fig. 3) before leaking back into the coupling fluid and reaching the receiver R.

TABLE I. Dimensions of three corrugated plate specimens. All dimensions are given in in. and mm; mm values are given in parentheses.

Specimen No.	H_1 $(2h+2\varepsilon)$	H_2 $(2h-2\varepsilon)$	D	$2h$ $(H_1+H_2)/2$	ε $(H_1-H_2)/4$	$2h/D$	ε/D
1	0.5 (12.7)	0.416 (10.57)	0.425 (10.8)	0.458 (11.63)	0.021 (0.53)	1.078	0.049
2	0.5 (12.7)	0.3 (7.62)	0.37 (9.40)	0.4 (10.16)	0.05 (1.27)	1.081	0.135
3	0.5 (12.7)	0.187 (4.75)	0.38 (9.65)	0.344 (8.74)	0.078 (1.98)	0.905	0.205

D. Experimental results

Experiments are carried out for two different angles of incidence, $\theta=25°$ and 30°. These two angles are selected because for these inclination angles relatively strong guided waves could be generated in the corrugated plate in the frequency range of our interest. Experimental results for these two sets of incident angles are described in detail below.

1. Experimental results for 25° and 30° angles of incidence

Two transducers T and R, of 1 in. (25.4 mm) diameter, are placed above the smooth portion of the aluminum plate which is 0.5 in. (12.7 mm) thick. First, the transducers are positioned such that the directly reflected beam (shown by dashed line in Fig. 3) can reach the receiver R. This is the case when $h=3$ in. (76.2 mm) and $d=2.8$ in. (71.1 mm). The received signal strength as a function of the frequency for this transducer-receiver arrangement is shown in Fig. 4. Note that Figs. 1 and 4 are almost identical. Therefore, the receiving signal characteristics are not altered significantly when the transmitter and receiver are placed in the pitch-catch arrangement—with the receiver receiving the direct reflected beam. The received signal is plotted after attenuating the signal by a 37 dB attenuator. When h is reduced to 2.5 in. (63.5 mm) and d is proportionately reduced to 2.3 in. (58.4 mm), then the reflected beam showed similar strength variation with frequency.

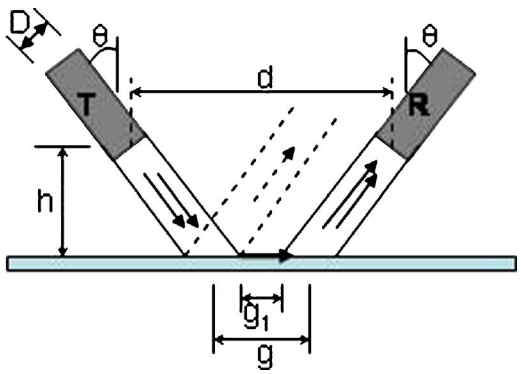


FIG. 3. Schematic of the transmitter (T), receiver (R) and the plate specimen arrangement. The direct reflected beam is shown by dashed lines. The receiver is placed beyond the direct reflection zone to detect the leaky guided waves.

Keeping h fixed at 2.5 in. (63.5 mm) when the transducer spacing is increased to 4.25 in. (108 mm), the received signal voltage versus frequency plot is changed significantly—as shown in Fig. 5. We will refer to the received signal voltage versus frequency plots as $V(f)$ curves. The $V(f)$ curve of Fig. 5 is plotted after attenuating the received signal by a 28 dB attenuator. Note that the peak near 540 kHz, observed in Figs. 1 and 4, is no longer present in Fig. 5. Also, two peaks of Fig. 5, near 430 kHz and 645 kHz, are absent in Figs. 1 and 4. It will be shown later that these two peaks correspond to two Lamb wave modes in the plate. A simple calculation with transducer diameter $D=1$ in. (25.4 mm), transducer spacing $d=4.25$ in. (108 mm), height $h=2.5$ in. (63.5 mm), and transducer inclination angle $\theta=25°$ gives g (see Fig. 3) =1.918 in. (48.72 mm) and g_1 (see Fig. 3) =0.815 in. (20.7 mm). Since g_1 is nonzero, the direct reflected beam cannot reach the receiver. Therefore, the ultrasonic energy must propagate through the plate as guided waves—for a certain distance greater than g_1 before leaking into the coupling fluid and being received by the receiver. It will be shown later that two frequencies, 430 kHz and 645 kHz, generate two guided wave modes for transducer inclination angle $\theta=25°$.

When the smooth plate is replaced by an aluminum plate with small corrugation ($\varepsilon/D=0.049$, Specimen 1 in Table I),

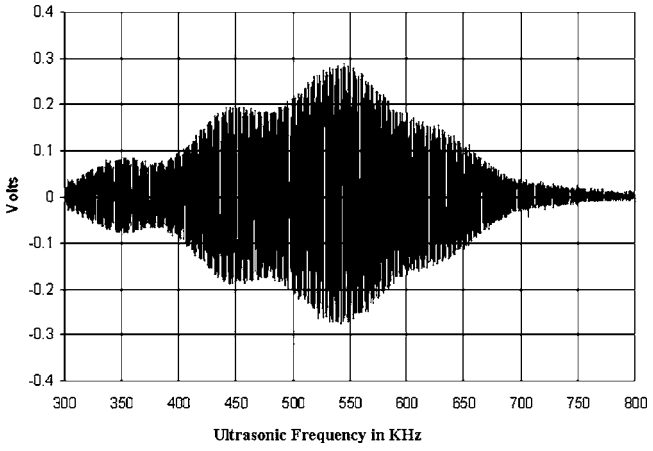


FIG. 4. Received signal voltage amplitude versus signal frequency curve, or $V(f)$ curve, for a smooth plate specimen when the receiver is placed in the direct reflection zone marked by dashed lines in Fig. 3. Note the similarities between Figs. 1(b) and 4. [For this figure, $\theta=25°$, $h=3$ in. (76.2 mm), $d=2.8$ in. (71.1 mm), and the signal attenuation is 37 dB. A similar plot is obtained for $h=2.5$ in. (63.5 mm) and $d=2.3$ in. (58.4 mm)].

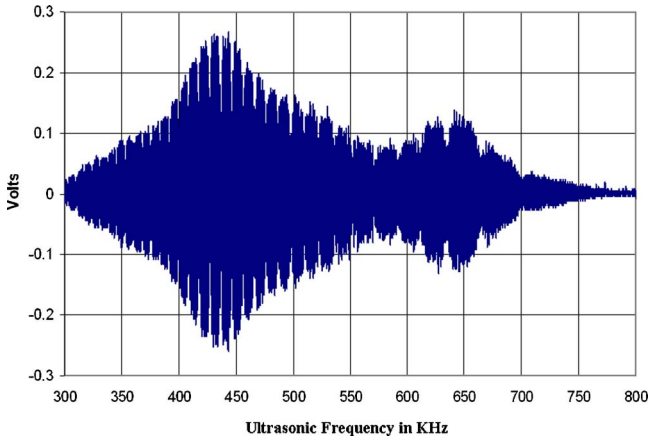


FIG. 5. (Color online) $V(f)$ curve for a smooth plate specimen when the receiver is placed beyond the direct reflection zone as shown in Fig. 3. Note the changes in the $V(f)$ curves of Figs. 4 and 5 in spite of the fact that the plate specimens for both figures are the same; the only difference is the horizontal distance (d) between the two transducers. For Fig. 5, the distance d is greater. [For this figure, $\theta=25°$, $h=2.5$ in. (63.5 mm), $d=4.25$ in. (108 mm), and the signal attenuation is 28 dB].

the $V(f)$ curve obtained for the setting $h=3$ in. (76.2 mm) and $d=7$ in. (177.8 mm) is shown in Fig. 6(a). For the transducer spacing $d=7$ in., the distance traveled by the guided wave in the corrugated plate is significantly greater than that for Fig. 5. Naturally, the received signal in Fig. 6(a) is much weaker than that in Fig. 5. Only an attenuation of strength 16 dB is applied to the received signal before plotting it in Fig. 6(a); while for Fig. 5, it was 28 dB.

A comparison between Figs. 6(a) and 4 shows some similarities between these two $V(f)$ curves—both have peaks between 500 and 550 kHz and the signal strength gradually decays to a very small value at low (~300 kHz) and high (~800 kHz) frequencies. However, a closer inspection also reveals some clear distinctions that will be discussed later.

Keeping all parameters (h,d,θ) unchanged, Specimen 1 is then replaced by Specimen 2 and finally by Specimen 3. The $V(f)$ curve for Specimen 2 (medium corrugation, $\varepsilon/D=0.135$) is shown in Fig. 6(b), and for Specimen 3 (large corrugation, $\varepsilon/D=0.205$) is shown in Fig. 6(c). To maintain the numerical value of the $V(f)$ peaks close to 0.3 in all plots, a 14 dB attenuator is used for Fig. 6(b), and an 18 dB attenuator is used for Fig. 6(c). Comparison of these two figures with Fig. 4 shows some distinctive features that are discussed later.

Similar experiments with the same three corrugated plate specimens are carried out again for the 30° angle of incidence and $V(f)$ curves for the three plates are recorded. Three $V(f)$ curves for the three corrugated plates for 30° angle of incidence are shown in Figs. 7(a)–7(c).

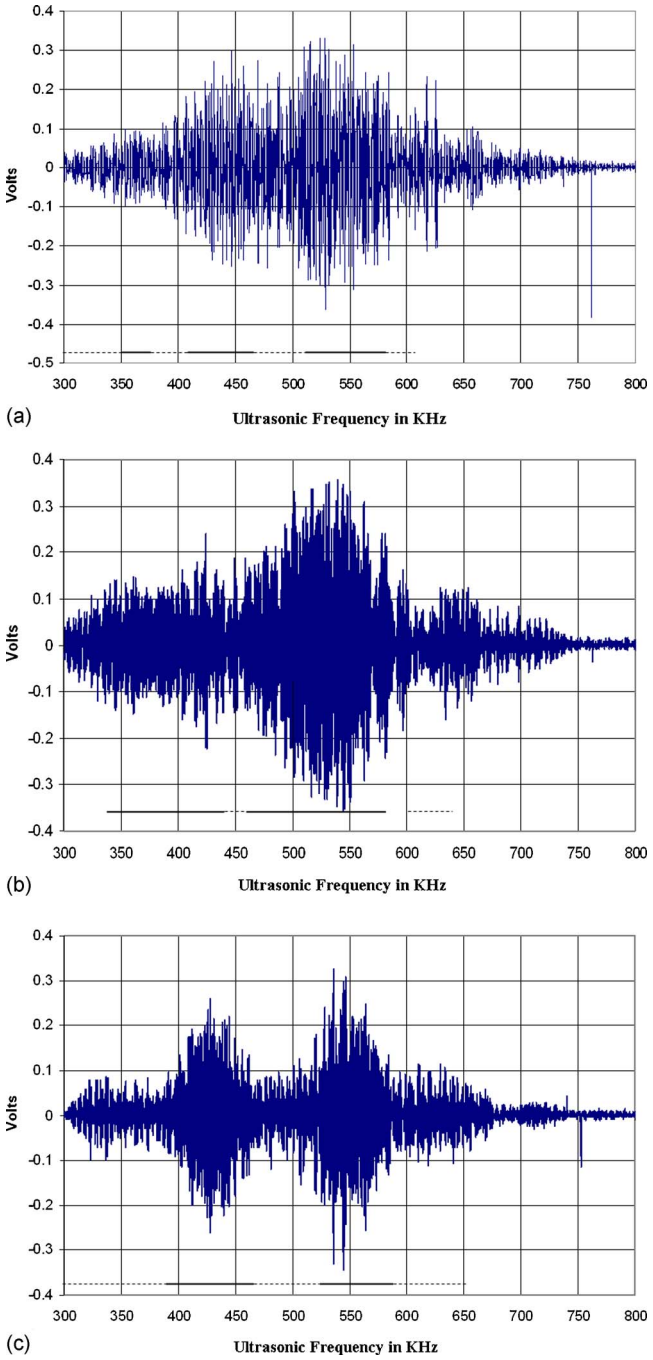


FIG. 6. (Color online) $V(f)$ curves for three corrugated plate specimens when the transducer inclination angle is $\theta=25°$ and the receiver is placed beyond the direct reflection zone as shown in Fig. 3. (a) $V(f)$ curve for Specimen No. 1 (low corrugation, see Table I) when $h=3$ in. (76.2 mm), $d=7$ in. (177.8 mm), and the signal attenuation is 16 dB. (b) $V(f)$ curve for Specimen No. 2 (medium corrugation, see Table I) when $h=3$ in. (76.2 mm), $d=7$ in. (177.8 mm), and the signal attenuation is 14 dB. (c) $V(f)$ curve for Specimen No. 3 (large corrugation, see Table I) when $h=3$ in. (76.2 mm), $d=7$ in. (177.8 mm), and the signal attenuation is 18 dB.

2. Distinctive features of V(f) curves of corrugated plates

A comparison of Figs. 4 and 7(a) reveals that, in Fig. 4, the signal strength is the maximum near 540 kHz and it decays almost monotonically for both higher and lower frequencies with a couple of local minima observed near 380 and 480 kHz, while that is not the case in Fig. 7(a). Although the $V(f)$ amplitude envelope has a decaying trend for both higher and lower frequencies, this trend is not as monotonic as in Fig. 4. Clearly, in Fig. 7(a), the amplitude envelope has two noticeable dips (almost global minima) near 380 and 480 kHz, as shown by dashed curved line in Fig. 7(a). A few other smaller dips may be noticed in the amplitude envelope, but the two strongest dips are near 380 and 480 kHz. Note

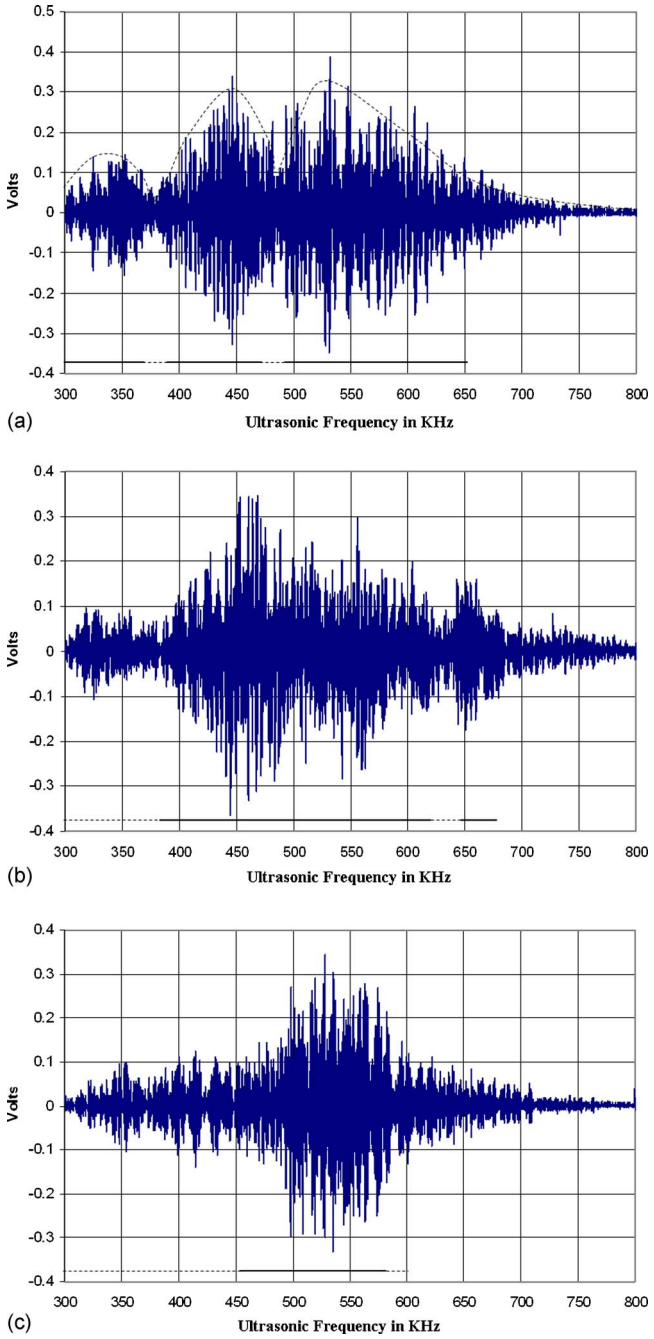


FIG. 7. $V(f)$ curves for three corrugated plate specimens when the transducer inclination angle $\theta=30°$ and the receiver is placed beyond the direct reflection zone as shown in Fig. 3. (a) $V(f)$ curve for Specimen No. 1 (low corrugation, see Table I) when $h=2.25$ in. (57.2 mm), $d=7.5$ in. (190.5 mm), and the signal attenuation is 14 dB. (b) $V(f)$ curve for Specimen No. 2 (medium corrugation, see Table I) when $h=2.25$ in. (57.2 mm), $d=7.5$ in. (190.5 mm), and the signal attenuation is 10 dB. (c) $V(f)$ curve for Specimen No. 3 (large corrugation, see Table I) when $h=2.25$ in. (57.2 mm), $d=7.5$ in. (190.5 mm), and the signal attenuation is 15 dB.

that the propagating signal amplitude is very small in the frequency ranges of 370–390 kHz and 470–490 kHz. Clearly, the ultrasonic signal finds it difficult to propagate in these two frequency ranges. The frequency bandwidths that block the ultrasonic wave propagation through the plate are called "stop bands" and the frequency bandwidths that do not cause such an obstruction to the wave propagation are called "pass bands." Therefore, for Specimen 1, for a 30° angle of incidence, the stop bands are 370–390 kHz and 470–490 kHz, while the frequency bandwidths of 300–370 kHz, 390–470 kHz, and 490–650 kHz constitute the pass bands. Near the bottom of Fig. 7(a), continuous and dashed horizontal lines are used to mark the pass band and stop band regions, respectively. Similarly, in Figs. 7(b), 7(c), and 6(a)–6(c) pass bands and stop bands are marked by continuous and dashed lines, respectively. In some figures, clear distinctions exist between the signals in pass band and stop band regions. For example, in Fig. 6(c), signals in the pass band zones are significantly stronger than those in the stop band zones. However, in some other figures, such as in Fig. 7(a), the signal strength variations in these two regions are not that distinct. In some cases, logical judgments have been used to decide pass band and stop band regions. For example, in Fig. 7(c), one can see that the signal strength is weak in the region from 300 to 490 kHz, and strong between 490 and 580 kHz. However, we denoted the stop band from 300 to 450 kHz instead of 490 kHz because the signal strength starts to increase after 450 kHz in Fig. 7(c); while in Fig. 4 (the flat plate case), the signal strength decreases from 450 to 480 kHz. Therefore, the corrugated surface is probably not creating a stop band between 450 and 490 kHz. It should be mentioned here that such subjective judgments and ambiguities may be overcome by employing sophisticated signal processing techniques, which can compare the strength of the received signals in different frequency ranges for the corrugated plates with those for the smooth plates and face to face orientations of the transducers.

Stop band and pass band frequencies for the three plates, obtained for 30° and 25° angles of incidence, are shown by continuous and dashed lines in Figs. 6 and 7, and their values are listed in Table II.

E. Dispersion curves for smooth plates

Before analyzing and understanding the experimental data for the corrugated plates, given in Figs. 6 and 7 and summarized in Table II, it is necessary to investigate first if the $V(f)$ curve for the smooth plate (Fig. 5) is reliable; in other words, whether the peaks of the $V(f)$ curve for the smooth plate—for which the guided wave propagation theory is well developed—appear at the right places. Figure 5 shows its two peaks near 430 kHz and 645 kHz; these peaks are not present in Fig. 4. Do these peaks correspond to the Lamb wave modes generated in the plate? To investigate this, the dispersion curves for the aluminum plate are theoretically computed. The P-wave speed (c_P) in aluminum is 6.2 km/s, its S-wave speed (c_S) is 3 km/s, and density (ρ) is 2.7 gm/cc. The plate thickness is 12.7 mm.

Lamb wave dispersion curves for a homogeneous isotropic elastic plate are obtained from the well-known dispersion equations [Kundu, 2004]:

TABLE II. Stop band and pass band frequencies in kHz for two striking angles and three corrugated plate specimens whose dimensions are given in Table I.

Striking angle	Specimen No. 1		Specimen No. 2		Specimen No. 3	
	Stop band	Pass band	Stop band	Pass band	Stop band	Pass band
30°	370–390 470–490	300–370 390–470 490–650	300–380 620–645	380–620 645–680	300–450 580–600	450–580
25°	300–350 375–410 465–510 580–610	350–375 410–465 510–580	440–460 600–640	340–440 460–580	300–390 470–530 590–650	390–470 530–590

$$\frac{\tanh\left(\omega h\sqrt{\frac{1}{c_L^2}-\frac{1}{c_P^2}}\right)}{\tanh\left(\omega h\sqrt{\frac{1}{c_L^2}-\frac{1}{c_S^2}}\right)}=\frac{\left(\frac{2}{c_L^2}-\frac{1}{c_S^2}\right)^2}{\frac{4}{c_L^2}\sqrt{\frac{1}{c_L^2}-\frac{1}{c_P^2}}\sqrt{\frac{1}{c_L^2}-\frac{1}{c_S^2}}}, \quad (1a)$$

$$\frac{\tanh\left(\omega h\sqrt{\frac{1}{c_L^2}-\frac{1}{c_P^2}}\right)}{\tanh\left(\omega h\sqrt{\frac{1}{c_L^2}-\frac{1}{c_S^2}}\right)}=\frac{\frac{4}{c_L^2}\sqrt{\frac{1}{c_L^2}-\frac{1}{c_P^2}}\sqrt{\frac{1}{c_L^2}-\frac{1}{c_S^2}}}{\left(\frac{2}{c_L^2}-\frac{1}{c_S^2}\right)^2}, \quad (1b)$$

where ω is the angular frequency ($\omega=2\pi f$) of the propagating wave; the signal frequency f is in MHz, and ω is in rad/μs. h is one-half of the plate thickness in mm, c_P and c_S are the P-wave speed and S-wave speed in the plate material, respectively, and c_L is the phase velocity of the propagating Lamb wave modes. All velocities are in km/s. Equations (1a) and (1b) correspond to the symmetric and antisymmetric Lamb modes, respectively. Dispersion curves generated by Eq. (1) are shown in Fig. 8.

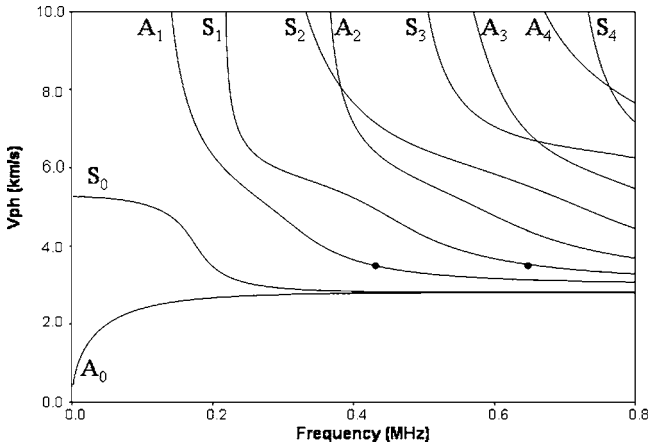


FIG. 8. Dispersion curves of 0.5 in. (12.7 mm) thick aluminum plate (c_P =6.2 km/s, c_S=3 km/s, and ρ=2.7 gm/cc). Two black circles are the experimental data points corresponding to the two peaks at 430 kHz and 645 kHz in the $V(f)$ curve of Fig. 5, corresponding phase velocity V_{ph} =3.526 km/s for 25° angle of incidence is obtained from Snell's law [Eq. (2)]. Antisymmetric and symmetric modes of order m are denoted by A_m and S_m, respectively.

The incident angle for the $V(f)$ curves of Fig. 5 is 25°. Therefore, the corresponding phase velocity from Snell's law is

$$c_L=\frac{c_f}{\sin\theta}=\frac{1.49}{\sin(25)}=3.526 \text{ km/s}, \quad (2)$$

where c_f is the acoustic wave speed in water (=1.49 km/s) and θ is the incident angle (=25°). Therefore, two peaks of Fig. 5 correspond to two points in the frequency-phase velocity plot of Fig. 8. The horizontal and vertical coordinates of these points are (430 kHz, 3.526 km/s) and (645 kHz, 3.526 km/s). These points are plotted in Fig. 8 by solid circles. Note that they coincide with the A_1 (first antisymmetric) and S_1 (first symmetric) modes. Thus, the reliability of the experimental $V(f)$ plots is established.

F. Dispersion curves for corrugated plates

Banerjee and Kundu (2006) presented a theoretical solution of elastic wave propagation in sinusoidal corrugated plates as shown in Fig. 9. Their approach is not based on the perturbation theory and can be applied equally well to both small and large corrugations. They obtained the dispersion equation by applying the traction-free boundary conditions. Solution of the dispersion equation gives both symmetric and antisymmetric modes. In a periodically corrugated waveguide, all possible spectral orders of wave numbers were considered for the analytical solution. It was observed that the truncation of the spectral order influenced the results. The

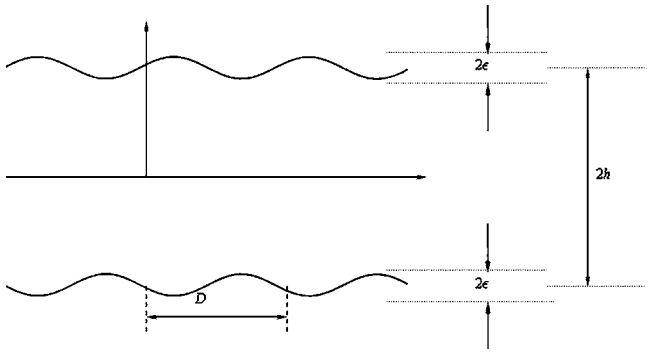


FIG. 9. Corrugated plate geometry with sinusoidal boundaries considered for the theoretical analysis. D=corrugation period, ε=corrugation depth, and $2h$=average plate thickness.

truncation number depends on the degree of corrugation and the frequency of the wave. Usually, increasing frequency requires increasing the number of terms in the series solution, or in other words, a higher truncation number. The dispersion equation for such plates with periodic boundary geometry can be written as

$$\mathrm{Det}[\mathbf{T}] = 0. \tag{3}$$

The dimension of the matrix $\mathbf{T}$ is $(2\mathrm{mod}(n)+1)\times(2\mathrm{mod}(n)+1)$, where n is the number of wave numbers developed for each frequency in the sinusoidally corrugated plate. If n varies from -1 to $+1$, the elements of the $\mathbf{T}$ matrix can be written as (Banerjee and Kundu, 2006)

$$BC1^{+}_{np} = -2\mu(k_n\eta_n)\lfloor De^{ih\eta_n}(J_0(\varepsilon\eta_n) + iH_0(\varepsilon\eta_n))\rfloor,$$

$$BC1^{-}_{np} = -2\mu(-k_n\eta_n)\lfloor De^{-ih\eta_n}(J_0(\varepsilon\eta_n) - iH_0(\varepsilon\eta_n))\rfloor,$$

$$BC1^{+}_{ns} = \mu(k_n^2-\beta_n^2)[De^{ih\beta_n}(J_0(\varepsilon\beta_n) + iH_0(\varepsilon\beta_n))],$$

$$BC1^{-}_{ns} = \mu(k_n^2-\beta_n^2)[De^{-ih\beta_n}(J_0(\varepsilon\beta_n) - iH_0(\varepsilon\beta_n))],$$

$$BC2^{+}_{np} = [\lambda(-k_n^2-\eta_n^2) + 2\mu(-\eta_n^2)][De^{ih\eta_n}(J_0(\varepsilon\eta_n) + iH_0(\varepsilon\eta_n))],$$

$$BC2^{-}_{np} = [\lambda(-k_n^2-\eta_n^2) + 2\mu(-\eta_n^2)][De^{-ih\eta_n}(J_0(\varepsilon\eta_n) - iH_0(\varepsilon\eta_n))],$$

$$BC2^{+}_{ns} = 2\mu(k_n\beta_n)\lfloor De^{ih\beta_n}(J_0(\varepsilon\beta_n) + iH_0(\varepsilon\beta_n))\rfloor,$$

$$BC2^{-}_{ns} = -2\mu(k_n\beta_n)\lfloor De^{-ih\beta_n}(J_0(\varepsilon\beta_n) - iH_0(\varepsilon\beta_n))\rfloor,$$

$$BC3^{+}_{np} = 2\mu(k_n\eta_n)\lfloor De^{-ih\eta_n}(J_0(\varepsilon\mu_n) - iH_0(\varepsilon\eta_n))\rfloor,$$

$$BC3^{-}_{np} = -2\mu(k_n\eta_n)\lfloor De^{ih\eta_n}(J_0(\varepsilon\mu_n) + iH_0(\varepsilon\eta_n))\rfloor,$$

$$BC3^{+}_{ns} = -\mu(k_n^2-\beta_n^2)[De^{-ih\beta_n}(J_0(\varepsilon\beta_n) - iH_0(\varepsilon\beta_n))],$$

$$BC3^{-}_{ns} = -\mu(k_n^2-\beta_n^2)[De^{ih\beta_n}(J_0(\varepsilon\beta_n) + iH_0(\varepsilon\beta_n))],$$

$$BC4^{+}_{np} = [\lambda(k_n^2+\eta_n^2) + 2\mu(\eta_n^2)][De^{-ih\eta_n}(J_0(\varepsilon\eta_n) - iH_0(\varepsilon\eta_n))],$$

$$BC4^{-}_{np} = [\lambda(k_n^2+\eta_n^2) + 2\mu(\eta_n^2)][De^{ih\eta_n}(J_0(\varepsilon\eta_n) + iH_0(\varepsilon\eta_n))],$$

$$BC4^{+}_{ns} = -2\mu(k_n\beta_n)\lfloor De^{-ih\beta_n}(J_0(\varepsilon\beta_n) - iH_0(\varepsilon\beta_n))\rfloor,$$

$$BC4^{-}_{ns} = 2\mu(k_n\beta_n)\lfloor De^{ih\beta_n}(J_0(\varepsilon\beta_n) + iH_0(\varepsilon\beta_n))\rfloor, \tag{4}$$

where the Struve function $H_n(z)$ appears in the solution of the inhomogeneous Bessel equation which for integer n has the form

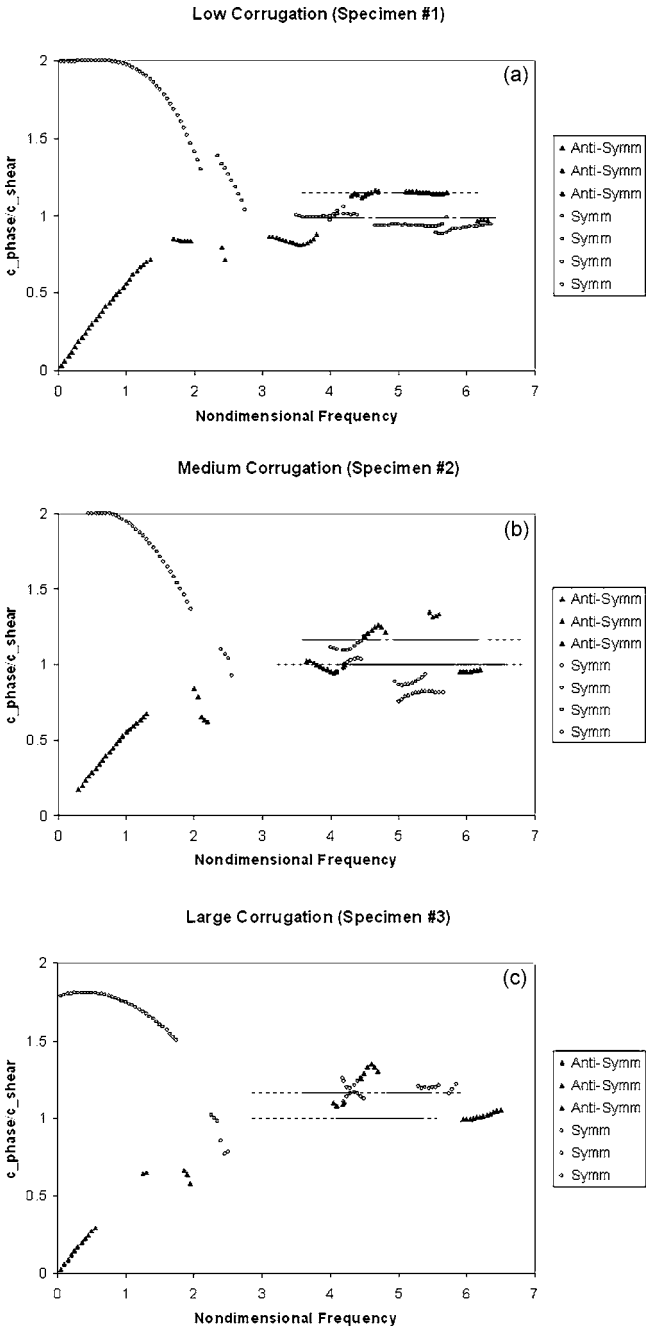


FIG. 10. Symmetric (circles) and antisymmetric (triangles) modes computed theoretically from Eq. for three plate specimens [(a)—small corrugation, Specimen No. 1; (b)—medium corrugation, Specimen No. 2; and (c)—large corrugation, Specimen No. 3]. See Table I for specimen dimensions. Experimentally obtained stop bands (dashed lines) and pass bands (continuous lines) for two normalized phase velocities (0.993 corresponds to 30° striking angle and 1.175 for 25° striking angle) are shown in each plot. In (a), pass bands match very well with the theoretical values. However, the matching between the theoretical and experimental values is not as good in (b) and (c).

$$z^2\frac{d^2y}{dz^2} + z\frac{dy}{dz} + (z^2-n^2)y = \frac{2}{\pi}\frac{z^{n+1}}{(2n-1)!!}, \tag{5}$$

the general solution of this equation consists of a linear combination of the Bessel functions $J_n(z)$ and the Struve functions $H_n(z)$.

Although the plate boundaries considered in the experiment are not pure sinusoidal, the geometry, shown in Fig. 9,

TABLE III. Nondimensional frequencies (Ω) for stop and pass bands for two striking angles and three corrugated plate specimens, whose dimensions are given in Table I.

Striking angle (θ)	30°	30°	25°	25°
$\frac{c_L}{c_S}=\frac{(1.49/\sin\theta)}{3}$	0.993	0.993	1.175	1.175
Specimen No. 1	Stop band frequency (MHz)	Pass band frequency (MHz)	Stop band frequency (MHz)	Pass band frequency (MHz)
Frequency (f) range (in MHz)	0.370–0.390 0.470–0.490	0.300–0.370 0.390–0.470 0.490–0.650	0.300–0.350 0.375–0.410 0.465–0.510 0.580–0.610	0.350–0.375 0.410–0.465 0.510–0.580
Nondimensional frequency (Ω) $\left(\Omega=\frac{2\pi fh}{c_S}=\frac{11.63\pi f}{3}=12.18f\right)$	4.51–4.75 5.72–5.97	3.65–4.51 4.75–5.72 5.97–7.92	3.65–4.26 4.57–4.99 5.66–6.21 7.06–7.43	4.26–4.57 4.99–5.66 6.21–7.06
Specimen No. 2	Stop band frequency (MHz)	Pass band frequency (MHz)	Stop band frequency (MHz)	Pass band frequency (MHz)
Frequency (f) range (in MHz)	0.300–0.380 0.620–0.645	0.380–0.620 0.645–0.680	0.440–0.460 0.600–0.640	0.340–0.440 0.460–0.580
Nondimensional frequency (Ω) $\left(\Omega=\frac{2\pi fh}{c_S}=\frac{10.16\pi f}{3}=10.64f\right)$	3.19–4.04 6.56–6.86	4.04–6.56 6.86–7.24	4.68–4.89 6.38–6.81	3.61–4.68 4.89–6.17
Specimen No. 3	Stop band frequency (MHz)	Pass band frequency (MHz)	Stop band frequency (MHz)	Pass band frequency (MHz)
Frequency (f) range (in MHz)	0.300–0.450 0.580–0.600	0.450–0.580	0.300–0.390 0.470–0.530 0.590–0.650	0.390–0.470 0.530–0.590
Nondimensional stop band frequency (Ω) $\left(\Omega=\frac{2\pi fh}{cS}=\frac{8.74\pi f}{3}=9.15f\right)$	2.75–4.12 5.35–5.49	4.12–5.31	2.75–3.57 4.35–4.85 5.40–5.95	3.57–4.30 4.85–5.40

is the closest geometry to our problem for which theoretical solutions are available today. A comparison between Figs. 2(c) (true plate geometry) and 9 (plate geometry that has analytical solution) shows some common features between these two geometries, such as both plates have a periodicity with wavelength D, both have a maximum plate thickness H_1, and a minimum plate thickness H_2. Then, the average plate thickness $2h$ is $(H_1+H_2)/2$ and the corrugation depth $2\varepsilon=(H_1-H_2)/2$.

Analytically computed dispersion curves for the fundamental symmetric and antisymmetric modes for the three plate geometries with $2h/D$ ratio equal to 1.078, 1.081, 0.905, and the corresponding ε/D ratio equal to 0.049, 0.135, and 0.205, respectively, are shown in Fig. 10. It should be noted here that although the geometry (Fig. 9) for the analytical solution is different from the specimen geometries (Fig. 2), two important parameters ($2h/D$ and ε/D) are the same for the analytical solution and the experimental investigation (listed in the right two columns of Table I). Three plate geometries for the analytical solution are denoted as Specimens 1, 2, and 3; similar to the three-plate specimens described in Table I. Figure 10 shows the analytically computed dispersion curves for the three corrugated plates. In Fig. 10, the phase velocity is normalized with respect to the shear wave speed (3 km/s) in the plate material. The nondimensional frequency (Ω) plotted along the horizontal axis is defined as

$$\Omega=\frac{\omega h}{c_S}, \tag{6}$$

where ω, h, and c_S are identical to those in Eq. (1).

In the dispersion curves of Fig. 10, one can observe several discontinuities that are not observed in the dispersion curves for a smooth plate (see Fig. 8). The gaps in the dispersion curves are called the stop bands. It is interesting to note that as the corrugation depth increases, the extent of the stop bands also increases. Experimentally, it is also observed that the stop band zones increase with the corrugation depth, see Figs. 6 and 7—it gives a qualitative agreement between the experimental observations and theoretical predictions. For a quantitative comparison between the experimental and theoretical results, the nondimensional frequency (Ω) and the normalized phase velocity c_L/c_S corresponding to the stop bands and pass bands shown in Table II, are calculated and listed in Table III.

As shown in Table III, c_L/c_S is 0.993 and 1.175 for the 30° and 25° angles of incidence, respectively. When the pass band and stop band frequencies are transformed from kHz (or MHz) to a nondimensional frequency (Ω) using Eq. (6), then the stop band of 370–390 kHz for Specimen No. 1 is changed to 4.51–4.75, as shown in Table III. When these stop bands (dashed lines) and pass bands (continuous lines) are plotted on the dispersion curves of Fig. 10, then sometimes good matching and discrepancies between the theoretical curves and experimental stop and pass bands are observed. Since experiments are carried out for two different incident angles that correspond to two different c_L/c_S values (0.993 and 1.175), we get two horizontal lines corresponding to these two normalized velocities, as shown in each plot of Fig. 10.

In Fig. 10(a), experimental stop bands (dashed lines) and pass bands (continuous lines) match very well with the theoretical dispersion curves. Note that the continuous lines either coincide or are located very close to the triangles (antisymmetric modes) or circles (symmetric modes), while the dashed lines are seen in the regions where neither circles nor triangles are present. However, the matching between the experimental data (horizontal continuous lines at $c_L/c_S = 0.993$ and 1.175) and the theoretical values (triangles and circles) are not as good in Figs. 10(b) and 10(c). The only matching that can be highlighted here is that, in Fig. 10(c) in the nondimensional frequency range from 2.5 to 3.5, both theoretical and experimental values show stop bands.

From Fig. 10 it can be concluded that for small corrugation depth (when the ε/D ratio is less than or equal to 0.05) the assumption of sinusoidal corrugation geometry is acceptable even when the actual geometry is not sinusoidal but periodic; however, for large corrugation depth ($\varepsilon/D > 0.1$) the sinusoidal corrugation assumption does not work very well when the actual corrugation geometry is not sinusoidal.

III. CONCLUSION

The elastic wave propagation in homogeneous plates with periodic corrugated boundaries is experimentally investigated in this paper. Guided waves in three plates with three different degrees of corrugation are studied. Different stop bands and pass bands are observed for the three plates. The extent of stop bands is found to increase with the depth of corrugation. Experimental data generated by nonsinusoidal corrugated plates are compared with the theoretical predictions for sinusoidal corrugated plates. For a small corrugation depth, the theoretical and experimental data match reasonably well. However, for a large corrugation depth, the matching is not as good—indicating that, for large degree of corrugation, the exact geometry of the plate boundary needs to be incorporated in the model. Only two parameters—the wavelength of periodicity and the depth of corrugation—are enough for correctly predicting the pass band and stop band regions in plates with a small degree of corrugation, but these two parameters are not enough for modeling wave propagation in plates with a large degree of corrugation.

ACKNOWLEDGMENTS

This research was partially supported from a research grant from the Air Force Research Laboratory, AFRL/MLLP, through CNDE (Center for Nondestructive Evaluation) of the Iowa State University and a grant from the National Science Foundation under Contract No. CMS-9901221. The authors would also like to acknowledge the technical help of Mr. R. Reibel and Dr. S. Sathish of UDRI/MLLP while carrying out the experiments.

Banerjee, S., and Kundu, T. (**2004**). "Elastic wave propagation in symmetrically periodic sinusoidal waveguide," Proc. SPIE **5394**, 89–98.

Banerjee, S., and Kundu, T. (**2006**). "Symmetric and antisymmetric Rayleigh-Lamb modes in sinusoidally corrugated waveguides: An analytical approach," Int. J. Solids Struct. (in press).

Boström, A. (**1983**). "Passbands and stopbands for an electromagnetic waveguide with a periodically varying cross section," IEEE Trans. Microwave Theory Tech. **31**, 752–756.

Boström, A. (**1989**). "Propagating, damped, and leaky surface waves on the corrugated traction-free boundary of an elastic half-space," J. Acoust. Soc. Am. **85**, 1549–1555.

Brillouin, L. (**1946**). *Wave Propagation in Periodic Structures* (Dover, New York).

Declercq, N. F., Degrieck, J., Briers, R., and Leroy, O. (**2005**). "Diffraction of homogeneous and inhomogeneous plane waves on a doubly corrugated liquid/solid interface," Ultrasonics **43**, 605–618.

El-Bahrawy, A. (**1994a**). "Stopbands and passbands for symmetric Rayleigh-Lamb modes in a plate with corrugated surfaces," J. Sound Vib. **170**, 145–160.

El-Bahrawy, A. (**1994b**). "Point force excitation of surface waves along the doubly corrugated traction-free boundary of an elastic half-space," J. Acoust. Soc. Am. **96**, 3167–3176.

Fokkema, J. H. (**1980**). "Reflection and transmission of elastic waves by the spatially periodic interface between two solids (Theory of integral-equation method)," Wave Motion **2**, 375–393.

Glass, N. E., and Maradudin, A. A. (**1983**). "Leaky surface-elastic waves on both flat and strongly corrugated surfaces for isotropic, nondissipative media," J. Appl. Phys. **54**, 796–805.

Kundu, T. (editor) (**2004**). *Ultrasonic Nondestructive Evaluation: Engineering and Biological Material Characterization* (CRC Press, Boca Raton, FL), Chap. 1, pp. 1–142.

Mead, D. J. (**1970**). "Free wave propagation in periodically supported, infinite beams," J. Sound Vib. **11**, 181–197.

Mead, D. J. (**1975**). "Wave propagation and normal modes in periodic systems: 1. Monocoupled systems," J. Sound Vib. **40**, 1–18.

Mead, D. J. (**1976**). "Loss factors and resonant frequencies of periodic damped sandwiched plates," ASME J. Eng. Ind. **98**, 75–80.

Mead, D. J. (**1986**). "A new method of analyzing wave propagation in periodic structures: Applications to periodic Timoshenko beams and stiffened plates," J. Sound Vib. **104**, 9–27.

Mead, D. J., and Markus, S. (**1983**). "Coupled flexural-longitudinal wave motion in a periodic beam," J. Sound Vib. **90**, 1–24.

Mead, D. J., and Bardell, N. S. (**1987**). "Free vibration of thin cylindrical shell with periodic circumferential stiffeners," J. Sound Vib. **115**, 499–521.

Mead, D. J., and Yaman, Y. (**1991**). "The harmonic response of rectangular sandwich plates with multiple stiffening: A flexural wave analysis," J. Sound Vib. **145**, 409–428.

Mester, S. S., and Benaroya, H. (**1995**). "Periodic and near-periodic structures," J. Sound Vib. **2**, 69–95.

Nayfeh, A. H., and Kandil, O. A. (**1978**). "Propagation waves in cylindrical hard-walled ducts with generally weak undulations," AIAA J. **16**, 1041–1045.

Orris, R. M., and Petyt, M. (**1974**). "A finite element study of harmonic wave propagation in periodic structures," J. Sound Vib. **33**, 223–236.

Russene, M., and Baz, A. (**2000**). "Control of wave propagation in periodic

composite rods using shape memory inserts," J. Vibr. Acoust. **122**, 151–159.
Sandström, S. E. (**1986**). "Stop bands in a corrugated parallel-plate waveguide," J. Acoust. Soc. Am. **79**, 1293–1298.
Vasseur, J. O., Deymier, P. A., Frantziskonis, G., Hong, G., Djafari-Rouhani, B., and Dobrzynski, L. (**1998**). "Experimental evidence for the existence of absolute acoustic band gaps in two-dimensional periodic composite media," J. Phys.: Condens. Matter **10**, 6051–6064.

Optimization of multilayered panels with poroelastic components for an acoustical transmission objective

O. Tanneau, J. B. Casimir,[a] and P. Lamary
Institut Supérieur de Mécanique de Paris, 3 rue Fernand Hainaut 93407 Saint-Ouen, France

(Received 12 July 2005; revised 12 May 2006; accepted 24 June 2006)

A method for optimizing acoustical linings is described and applied to multilayered panels including solid, fluid, and porous components. This optimization is based on an analytical simulation of the insulation properties and a genetic algorithm. The objective function is defined by taking into account both the acoustical frequency response over a 1/3 octave spectrum and the total mass of the panel. The optimization process gives rise to an optimal choice for the number of layers as well as for the nature and the thickness of each layer that maximizes the transmission loss. A practical example of such an optimization is described.
[DOI: 10.1121/1.2228663]

PACS number(s): 43.20.Gp [DF] Pages: 1227–1238

I. INTRODUCTION

Multilayered panels are widely used in acoustic engineering and noise control. Manufacturers of such panels are confronted with the problem of material selection and layer combinations. What is the best material combination in order to increase insulation properties of multilayered panels? Moreover, the question of minimizing the mass of the panels often arises in aeronautic applications. These questions could be expressed and combined to give rise to an optimization problem (Germain, 1999). The main object of this paper consists of describing such a problem and providing a potential solution.

In the field of engineering design, objective functions of optimization problems are often based on catalog selection. In this case, there is no way of calculating derivatives of objective functions and nongradient methods are a very good alternative to tackle these kind of problems. Moreover, these methods are more likely to reach a global optimum for continuous parameter problems (Anderson, 2000). Among these nongradient optimization methods, one could cite: Simulated Annealing (SA), Genetic Algorithms (GA), Random Search (RS) or Tabu Search (TS) (Kirkpatrick *et al.*, 1983; Goldberg, 1989; Jang *et al.*, 1997; Glover, 1989). Such optimization problems often involve conflicting objectives such as mass, cost and mechanical resistance. This kind of problem is called "multiobjective" and its solutions are not unique (Steuer, 1986). In this context, evolutionary algorithms are often used to find simple solutions to these discrete multiobjective optimization problems (Fonseca, 1995). This approach is currently one of the most active research directions. Recently, Xu *et al.* (2004) have used such algorithm for the optimization of flat-walled multilayered anechoic linings.

This paper is divided into three distinct parts. First, we describe a general method that allows one to evaluate the various types of performance offered by acoustical panels. The so-called "Transfer Matrix Method" is used to obtain a suitable indicator for these performances. This method has been widely used to deal with acoustic engineering problems in which porous materials are involved: surface impedance of multilayered panels (Allard *et al.*, 1987; Lauriks *et al.*, 1990), transmission loss in insulation system including a combination of solid, porous and fluid media (Bolton *et al.*, 1996; Lauriks *et al.*, 1992). In others contexts of wave propagation problems such as geophysics and general layered media, the Transfer Matrix Method is currently used (Vashishth *et al.*, 2004; Brekhovskikh, 1960). The chosen indicator is the "Transmission Loss Factor," denoted TL and defined over a given frequency range. This indicator is used to calculate the objective function for the optimization problem. In the second part, the optimization problem is defined and a genetic algorithm is presented. This problem involves both acoustical properties and the mass density of a panel. The genetic algorithm allows one to obtain a Pareto-optimal solution. The last section describes an application of the method. An optimal solution is obtained for a given TL spectrum objective and a given mass objective.

II. THE TRANSFER MATRIX METHOD

A. General formulation

Consider a panel made up of N layers. Each layer i is characterized by a constitutive fluid, solid or poro-elastic material and a constant thickness e_i. The lateral dimensions of the layers are assumed to be infinite. The multilayered panel is placed in an ambient fluid and is submitted to an incident planar acoustical wave on its first layer. The outcoming waves are evaluated on the final layer. The surrounding fluid is considered to be semi-infinite layers situated before the first panel's layer and behind the final layer. These semi-infinite layers are described as layer 1 and layer n, therefore $N=n-2$ (see Fig. 1).

The number m_i of transmitted waves in layer i depends on the nature of the constitutive material. For a fluid layer merely a longitudinal wave is transmitted. Two waves are transmitted in a solid layer: a longitudinal and a transverse wave which are, respectively, related to compression strain

[a]Electronic mail: jean-baptiste.casimir@supmeca.fr

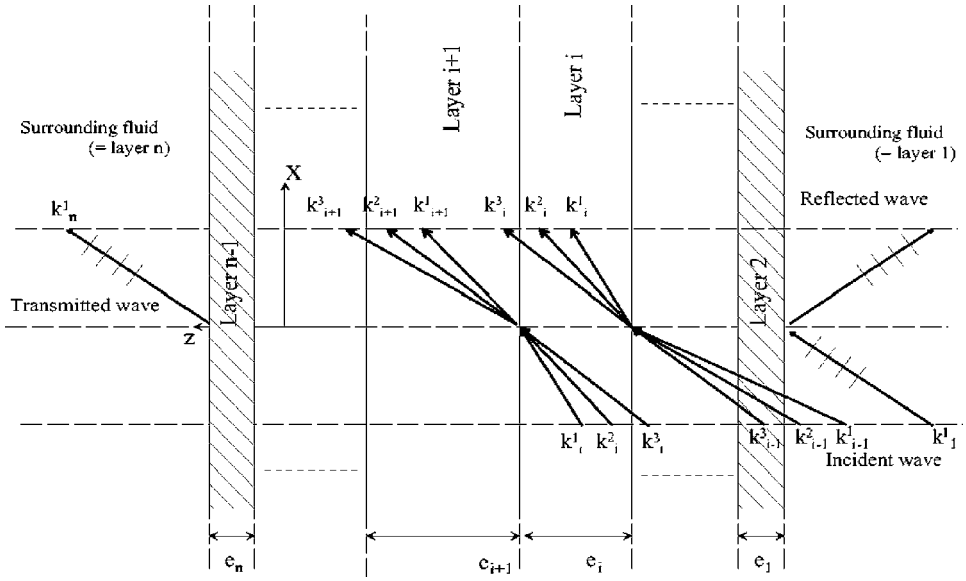


FIG. 1. Multilayered panel.

and shear strain. According to the Biot-Allard theory (Allard, 1993), in the case of a poroelastic material which is the combination of a solid and a fluid phase, three waves are transmitted: two longitudinal waves and a transverse Biot wave. The wave $j(1 \leq j \leq m_i)$ in the layer i is described by an incidence angle θ_i^j, a wave number k_i^j, and complex parameters ${}^1\varphi_i^j$, ${}^2\varphi_i^j$. These parameters are related to the forward and backward propagating components of the wave. The displacement potential Φ_i^j at a given position $M(x,y,z)$ in the layer i is given by

$$\forall y, \quad \Phi_i^j(M)_{M\in i} = ({}^1\varphi_i^j e^{-jk_i^j z\cos\theta_i^j} + {}^2\varphi_i^j e^{jk_i^j z\cos\theta_i^j}) e^{-jk_i^j x\sin\theta_i^j}. \tag{1}$$

A matricial form of expression (1) is given

$$\Phi_i^j(x,z) = [F_i^j(x,z)] \cdot \begin{Bmatrix} {}^1\varphi_i^j \\ {}^2\varphi_i^j \end{Bmatrix}. \tag{2}$$

Relations between incidence angles of waves m and n propagating in two adjacent layers are given by the well-known Snell-Descartes equation,

$$k_i^m \sin\theta_i^m = k_{i+1}^n \sin\theta_{i+1}^n. \tag{3}$$

Continuity conditions have to be written for each interface between the layers. The resulting relations involve displacements and stress components that are related to displacement potentials. Relationships between displacement potentials and stress/displacement components depend on the nature of layer i, a general matricial form for these relationships is given by expression,

$$\begin{Bmatrix} \mathbf{u_i}(x,z) \\ \bar{\bar{\sigma}}_i(x,z) \end{Bmatrix} = [H_i(x,z)] \cdot \{\phi_\mathbf{i}\} \tag{4}$$

with:

$$\{\phi_\mathbf{i}\} = \begin{pmatrix} {}^1\varphi_i^1 \\ {}^2\varphi_i^1 \\ \vdots \\ {}^1\varphi_i^{m_i} \\ {}^2\varphi_i^{m_i} \end{pmatrix}.$$

Components of the vectors $\{\phi_\mathbf{i}\}$ are amplitudes of displacement potentials of the m_i waves transmitted in layer i, $\mathbf{u_i}$ and $\bar{\bar{\sigma}}_i$ are, respectively, the displacement vector and the stress tensor in layer i.

The z position of the interface between the layers i and $i+1$ is denoted z_i^{i+1}. Continuity relations on this interface concern the displacement vector and stress tensor. A general matricial form for these relations is given by expression

$$[I_i] \cdot \begin{Bmatrix} \mathbf{u_i}(x,z_i^{i+1}) \\ \bar{\bar{\sigma}}_i(x,z_i^{i+1}) \end{Bmatrix} = [I_{i+1}] \cdot \begin{Bmatrix} \mathbf{u_{i+1}}(x,z_i^{i+1}) \\ \bar{\bar{\sigma}}_{i+1}(x,z_i^{i+1}) \end{Bmatrix}. \tag{5}$$

The components of the matrices $[I_i]$ depends on the presence or absence of transmitted wave components.

Substitution of expression (4) into Eq. (5) gives continuity relations involving displacement potentials

$$[I_i] \cdot [H_i(x,z^{i+1})] \cdot \{\phi_\mathbf{i}\} = [I_{i+1}] \cdot [H_{i+1}(x,z_i^{i+1})] \cdot \{\phi_\mathbf{i+1}\}. \tag{6}$$

As a result of the Snell-Descartes relation (3), the x dependance for each component of the matrices $[H_i(x,z_i^{i+1})]$ involves an identical exponential function that could be removed from Eq. (6). Hence, this equation is written,

$$[A_i(z_i^{i+1}) - A_{i+1}(z_i^{i+1})] \cdot \begin{Bmatrix} \phi_\mathbf{i} \\ \phi_\mathbf{i+1} \end{Bmatrix} = \{0\}, \tag{7}$$

where

$$[A_i(z_i^{i+1})] = [I_i] \cdot [G_i(z_i^{i+1})].$$

$[G_i(z_i^{i+1})]$ is the x independant part of $[H_i(x,z_i^{i+1})]$.

Equations (7) are written for each interface and are gathered in the matricial expression

$$\begin{pmatrix} A_1(z_1^2) & -A_2(z_1^2) & 0 & \cdots & 0 \\ 0 & A_2(z_2^3) & -A_3(z_2^3) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & A_{n-1}(z_{n-1}^n) & -A_n(z_{n-1}^n) \end{pmatrix} \cdot \begin{Bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_n \end{Bmatrix} = \{0\}. \quad (8)$$

B. Boundary conditions

Boundary conditions are given in the first and the last semi-infinite layers. The multilayered panel is placed in a fluid medium so that conditions on external interfaces are defined by considering the surrounding fluids as semi-infinite fluid domains. Inside the first semi-infinite layer $m_1=1$ and the incident wave is imposed, it is given by the potential ${}^1\varphi_1^1$. Inside the last semi-infinite layer $m_n=1$ and the backward wave is such that ${}^2\varphi_n^1=0$. Introduction of these boundary conditions in the system (8) leads to a modified system of linear equations whose unknowns are ${}^2\varphi_1^1, \phi_2, \ldots, \phi_{n-1}, {}^1\varphi_n^1$.

C. Resolution

For any given circular frequency ω, the modified system of linear equations shall be solved. Potentials are obtained throughout the panel. The transparency factor τ is defined for the whole panel according to

$$\tau = \frac{\rho_n |{}^1\varphi_n^1|^2}{\rho_0 |{}^1\varphi_1^1|^2}, \quad (9)$$

where ρ_n and ρ_0 are the density of the first and the last semi-infinite layers.

The resolution is achieved for a set of circular frequencies ω over a given range and thus a (TL) graph is obtained. TL is defined according to

$$\mathrm{TL} = 10 \log \frac{1}{\tau}. \quad (10)$$

In the case of diffuse noise, incidence angles are uniformely distributed over a given range $[0, \theta_f]$ and the transparency factor (Mulholland *et al.*, 1968) is defined according to

$$\tau = \frac{\int_0^{\theta_f} \tau(\theta) \sin\theta \cos\theta d\theta}{\int_0^{\theta_f} \sin\theta \cos\theta d\theta}. \quad (11)$$

III. CONSTITUTIVE RELATIONSHIPS IN THE LAYERS

A. Objective

The aim of this section is to define the stress and displacement components that are involved in continuity relations. In order to build the system of linear Eq. (8), the coupling variables have to be written in terms of the displacement potential according to expression (4). The number of such variables depends on the nature of the constitutive material. Fluid, solid, and poroelastic materials are fully examined.

B. Acoustical medium

A general fluid medium is commonly referred to an "acoustical medium." In the framework of linear acoustics and in the absence of fluid viscosity effects, continuity relations involve normal displacements and pressure field p. Therefore these components have to be written in terms of the unknown displacement potential φ.

In a Cartesian coordinate system whose z axis is perpendicular to the layers, one obtains

$$u_z = \frac{\partial \varphi}{\partial z}$$

$$p = -K\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_z}{\partial z}\right) = -K\left(\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial z^2}\right), \quad (12)$$

where K is the bulk modulus of the fluid.

The dimension along the y axis is not considered because only planar waves are taken into account. This propagation problem is bidimensional.

Introduction of expression (1) in system (12) gives the matricial expression (4),

$$\begin{Bmatrix} u_z \\ p \end{Bmatrix} = [{}^F H_i(x,z)] \cdot \{\phi_i\} \quad (13)$$

and therefore

$$[{}^F G_i(z)] = \begin{pmatrix} -jk\cos\theta e^{-jkz\cos\theta} & jk\cos\theta e^{jkz\cos\theta} \\ k^2\rho c_0^2 e^{-jkz\cos\theta} & k^2\rho c_0^2 e^{jkz\cos\theta} \end{pmatrix}, \quad (14)$$

where ρ is the density and $c_0 = \sqrt{K/\rho}$ is the propagation velocity.

C. Elastic medium

The elastic medium is assumed isotropic and constitutive relationships are obtained with small perturbation assumptions and linear elasticity theory. Displacement field $\mathbf{u}$ is described by a scalar potential φ and a vectorial potential $\boldsymbol{\Psi}$ according to

$$\mathbf{u} = \boldsymbol{\nabla}\varphi + \boldsymbol{\nabla} \times \boldsymbol{\Psi}. \quad (15)$$

Potentials of planar wave solutions are given by

$$\varphi = ({}^1\varphi e^{-jk_L z\cos\theta} + {}^2\varphi e^{jk_L z\cos\theta}) e^{-jk_L x\sin\theta},$$

$$\boldsymbol{\Psi} = \mathbf{v}({}^1\psi e^{-jk_T z\cos\gamma} + {}^2\psi e^{jk_T z\cos\gamma}) e^{-jk_T x\sin\gamma}, \quad (16)$$

where $\mathbf{v}$ is an unitary vector, φ and $\boldsymbol{\Psi}$ are potentials, respectively, associated with a longitudinal wave and a transversal wave. ${}^1\varphi$, ${}^2\varphi$, ${}^1\psi$, and ${}^2\psi$ are scalar parameters.

Continuity relations concern displacement and stress vectors at the interface between two layers. The components

involved in these relations have to be written in terms of the potentials φ and $\mathbf{\Psi}$. In a Cartesian coordinate system, one obtains

$$\begin{cases} u_x = \dfrac{\partial \varphi}{\partial x} - \dfrac{\partial \Psi_y}{\partial z} \\ u_y = \dfrac{\partial \Psi_x}{\partial z} - \dfrac{\partial \Psi_z}{\partial x} \\ u_z = \dfrac{\partial \varphi}{\partial z} + \dfrac{\partial \Psi_y}{\partial x} \end{cases} \quad \begin{cases} t_x = 2\mu \dfrac{\partial^2 \varphi}{\partial x\, \partial z} - \mu \dfrac{\partial^2 \Psi_y}{\partial z^2} + \mu \dfrac{\partial^2 \Psi_y}{\partial x^2} \\ t_y = \mu \dfrac{\partial \Psi_x}{\partial z^2} - \mu \dfrac{\partial \Psi_z}{\partial x\, \partial z} \\ t_z = (2\mu + \lambda) \dfrac{\partial^2 \varphi}{\partial z^2} + \lambda \dfrac{\partial^2 \varphi}{\partial x^2} + 2\mu \dfrac{\partial^2 \Psi_y}{\partial x\, \partial z} \end{cases} \tag{17}$$

where λ and μ are Lamé coefficients.

Expressions (17) are obtained for planar waves parallel to the y axis and the layer is assumed to be infinite in this direction. Therefore, there is no variation along the y axis and partial derivatives relative to the variable y are zero. These assumptions imply that the stress vector along the y axis vanishes, $t_y = 0$. The second and fifth relations show that $t_y = \mu(\partial u_y / \partial z)$, the component u_y is constant along the z axis. Continuity relations only concern displacement components u_x, u_z and stress components t_x, t_z.

Introduction of expression (16) in system (17) gives the matricial expression (4),

$$\begin{Bmatrix} t_z \\ u_z \\ t_x \\ u_x \end{Bmatrix} = [{}^S H_i(x,z)] \begin{Bmatrix} {}^1\varphi \\ {}^2\varphi \\ {}^1\Psi_y \\ {}^2\Psi_y \end{Bmatrix} = [{}^S H_i(x,z)]\{\phi_{\mathbf{i}}\} \tag{18}$$

and therefore

$$[{}^S G_i(z)] = \begin{bmatrix} k_L^2(\lambda + 2\mu\cos^2\theta)e^{-jk_L z\cos\theta} & k_L^2(\lambda + 2\mu\cos^2\theta)e^{jk_L z\cos\theta} & \cdots & k_T^2\mu\sin 2\gamma e^{-jk_T z\cos\gamma} & -k_T^2\mu\sin 2\gamma e^{jk_T z\cos\gamma} \\ -jk_L\cos\theta e^{-jk_L z\cos\theta} & jk_L\cos\theta e^{jk_L z\cos\theta} & & -jk_T\sin\gamma e^{-jk_T z\cos\gamma} & -jk_T\sin\gamma e^{jk_T z\cos\gamma} \\ k_L^2\mu\sin 2\theta e^{-jk_L z\cos\theta} & -k_L^2\mu\sin 2\theta e^{jk_L z\cos\theta} & & -k_T^2\mu\cos 2\gamma e^{-jk_T z\cos\gamma} & -k_T^2\mu\cos 2\gamma e^{jk_T z\cos\gamma} \\ -jk_L\sin\theta e^{-jk_L z\cos\theta} & -jk_L\sin\theta e^{jk_L z\cos\theta} & & jk_T\cos\gamma e^{-jk_T z\cos\gamma} & -jk_T\cos\gamma e^{jk_T z\cos\gamma} \end{bmatrix}. \tag{19}$$

D. Poroelastic medium

Acoustical insulation necessitates special materials like foam, glass-wool, and porous medium. Such kinds of materials are considered as two-phase media. Both a solid and a fluid phase are intermingled throughout the material. Two approaches are used for modeling such materials: equivalent fluid models and Biot's model. Equivalent fluid models (Delany *et al.*, 1970) are based on calculation of homogenized properties from the behavior of elementary cells. The equations presented in Sec. IV B must be used in modifying certain parmameters. Champoux-Allard's model (Allard, 1993) suggests a mass density and a bulk modulus as a function of circular frequency ω, respectively, given by

$$\rho(\omega) = \rho_0 \alpha_\infty \left(1 + \frac{\sigma_R \phi_P}{j\omega\alpha_\infty\rho_0}\sqrt{1 + \frac{4j\alpha_\infty^2\eta\rho_0\omega}{\sigma_R\Lambda^2\phi_P^2}}\right), \tag{20}$$

and

$$K(\omega) = \frac{\gamma P_0}{\gamma - (\gamma - 1)\dfrac{1}{1 + \dfrac{8\eta}{j\Lambda'^2 N_{\mathrm{Pr}}\omega\rho_0}\sqrt{1 + j\rho_0\dfrac{\omega N_{\mathrm{Pr}}\Lambda'^2}{16\eta}}}}, \tag{21}$$

where poroelastic properties are ϕ_P, porosity; σ_R, resistivity relative to air flow; α_∞, tortuosity; Λ, viscosity length; Λ', thermal length; and fluid properties are ρ_0, mass density; η, viscosity; N_{Pr}, Prandtl's number; $\gamma = c_P/c_V$, specific heat ratio.

For materials whose solid phase vibrations must be taken into account, equivalent fluid models cannot be used and Biot's models are required (Biot, 1956). Constitutive equations for such models are given hereafter.

Through a Fourier transform, the equilibrium equations are given by

$$\nabla \cdot \bar{\bar{\sigma}}^S + \omega^2(\tilde{\rho}_{11}\mathbf{u^S} + \tilde{\rho}_{12}\mathbf{u^F}) = \mathbf{0},$$

$$\nabla \cdot \bar{\bar{\sigma}}^F + \omega^2(\tilde{\rho}_{12}\mathbf{u^S} + \tilde{\rho}_{22}\mathbf{u^F}) = \mathbf{0}, \tag{22}$$

where $\tilde{\rho}_{11}$, $\tilde{\rho}_{12}$, and $\tilde{\rho}_{22}$ are parameters depending on the nature and the geometry of the poroelastic medium and the density of the fluid phase (Allard, 1993).

Strain-stress relations describe the behavior of each phase according to (Johnson, 1986)

$$\bar{\bar{\sigma}}^S = 2\mu\bar{\bar{\epsilon}}^S + (P - 2\mu)(\bar{\bar{\epsilon}}^S{:}\bar{\bar{I}})\bar{\bar{I}} + Q(\bar{\bar{\epsilon}}^F{:}\bar{\bar{I}})\bar{\bar{I}},$$

$$\bar{\bar{\sigma}}^F = Q(\bar{\bar{\epsilon}}^S{:}\bar{\bar{I}})\bar{\bar{I}} + R(\bar{\bar{\epsilon}}^F{:}\bar{\bar{I}})\bar{\bar{I}}, \tag{23}$$

where μ is the shear modulus and P, Q, R are the bulk modulii of the poroelastic material. In the case of foams, in the classical expressions of these moduli, the incompressibility of air is given by (21) (Biot and Willis, 1957).

Wave equations are obtained from Eqs. (22) and (23) written for each phase, one obtains

$$(P-\mu)\boldsymbol{\nabla\nabla}\,.\,\mathbf{u^S}+\mu\boldsymbol{\Delta}\mathbf{u^S}+Q\boldsymbol{\nabla\nabla}\,.\,\mathbf{u^F}+\omega^2(\tilde{\rho}_{11}\mathbf{u^S}+\tilde{\rho}_{12}\mathbf{u^F})=\mathbf{0},$$

$$Q\boldsymbol{\nabla\nabla}\,.\,\mathbf{u^S}+R\boldsymbol{\nabla\nabla}\,.\,\mathbf{u^F}+\omega^2(\tilde{\rho}_{12}\mathbf{u^S}+\tilde{\rho}_{22}\mathbf{u^F})=\mathbf{0}. \tag{24}$$

Scalar potentials φ^S, φ^F and vectorial potentials $\boldsymbol{\Psi^S}$, $\boldsymbol{\Psi^F}$ are introduced to describe displacements of the two phases according to

$$\mathbf{u^S}=\boldsymbol{\nabla}\varphi^S+\boldsymbol{\nabla}\times\boldsymbol{\Psi^S},$$

$$\mathbf{u^F}=\boldsymbol{\nabla}\varphi^F+\boldsymbol{\nabla}\times\boldsymbol{\Psi^F}. \tag{25}$$

From Eqs. (24) and definitions (25), one obtains equations satisfied by potentials φ^S, φ^F and $\boldsymbol{\Psi^S}$, $\boldsymbol{\Psi^R}$ and it may be shown (Allard, 1993) that displacements for each phase $\mathbf{u^S}$ and $\mathbf{u^F}$ can be expressed according to

$$\mathbf{u^S}=\boldsymbol{\nabla}\pi_1+\boldsymbol{\nabla}\pi_2+\boldsymbol{\nabla}\times\boldsymbol{\Psi},$$

$$\mathbf{u^F}=r_1\boldsymbol{\nabla}\pi_1+r_2\boldsymbol{\nabla}\pi_2+r_3\boldsymbol{\nabla}\times\boldsymbol{\Psi}, \tag{26}$$

where r_i are given by

$$r_i=\frac{Pk_i^2-\omega^2\tilde{\rho}_{11}}{\omega^2\tilde{\rho}_{12}-Qk_i^2}.$$

π_1, π_2, and $\boldsymbol{\Psi}$ are displacement potentials, respectively, associated with two longitudinal and a transversal wave.

As for both elastic and acoustical media, continuity relations concern displacement and stress vectors at the interface between two layers. For the fluid phase displacement u_z^F and pressure p^F the solid-phase displacement components u_x^S, u_z^S and stress components t_x and t_z are taken into account.

Development of displacement expressions (26) and stress expressions (23) in the Cartesian coordinate system leads to the matricial expression (4), one obtains

$$\begin{Bmatrix} p^F \\ t_z^S \\ t_x^S \\ u_z^F \\ u_z^S \\ u_x^S \end{Bmatrix}=[{}^PH_i(x,z)]\begin{Bmatrix} {}^1\pi_1 \\ {}^2\pi_1 \\ {}^1\pi_2 \\ {}^2\pi_2 \\ {}^1\Psi_y \\ {}^2\Psi_y \end{Bmatrix}=[{}^PH_i(x,z)]\,.\,\{\boldsymbol{\phi_i}\} \tag{27}$$

and therefore

$$[{}^PG_i(z)]$$

$$=\begin{bmatrix} (Rr_1+Q)k_1^2e^{-jk_1z\cos\theta_1} & (Rr_1+Q)k_1^2e^{jk_1z\cos\theta_1} \\ -[2\mu\sin^2\theta_1-(P+r_1Q)]k_1^2e^{-jk_1z\cos\theta_1} & -[2\mu\sin^2\theta_1-(P+r_1Q)]k_1^2e^{jk_1z\cos\theta_1} \\ \mu k_1^2\sin 2\theta_1e^{-jk_1z\cos\theta_1} & -\mu k_1^2\sin 2\theta_1e^{jk_1z\cos\theta_1} \\ -jk_1r_1\cos\theta_1e^{-jk_1z\cos\theta_1} & jk_1r_1\cos\theta_1e^{jk_1z\cos\theta_1} \\ -jk_1\cos\theta_1e^{-jk_1z\cos\theta_1} & jk_1\cos\theta_1e^{jk_1z\cos\theta_1} \\ -jk_1\sin\theta_1e^{-jk_1z\cos\theta_1} & -jk_1\sin\theta_1e^{jk_1z\cos\theta_1} \end{bmatrix}$$

$$\cdots\begin{bmatrix} (Rr_2+Q)k_2^2e^{-jk_2z\cos\theta_2} & (Rr_2+Q)k_2^2e^{jk_2z\cos\theta_2} \\ -[2\mu\sin^2\theta_2-(P+r_2Q)]k_2^2e^{-jk_2z\cos\theta_2} & -[2\mu\sin^2\theta_2-(P+r_2Q)]k_2^2e^{jk_2z\cos\theta_2} \\ \mu k_2^2\sin 2\theta_2e^{-jk_2z\cos\theta_2} & -\mu k_2^2\sin 2\theta_2e^{jk_2z\cos\theta_2} \\ -jk_2r_2\cos\theta_2e^{-jk_2z\cos\theta_2} & jk_2r_2\cos\theta_2e^{jk_2z\cos\theta_2} \\ -jk_2\cos\theta_2e^{-jk_2z\cos\theta_2} & jk_2\cos\theta_2e^{jk_2z\cos\theta_2} \\ -jk_2\sin\theta_2e^{-jk_2z\cos\theta_2} & -jk_2\sin\theta_2e^{jk_2z\cos\theta_2} \end{bmatrix}$$

$$\cdots\begin{bmatrix} 0 & 0 \\ \mu k_3^2\sin 2\gamma e^{-jk_3z\cos\gamma} & \mu k_3^2\sin 2\gamma e^{jk_3z\cos\gamma} \\ -\mu k_3^2\cos 2\gamma e^{-jk_3z\cos\gamma} & -\mu k_3^2\cos 2\gamma e^{jk_3z\cos\gamma} \\ -jk_3r_3\sin\gamma e^{-jk_3z\cos\gamma} & -jk_3r_3\sin\gamma e^{jk_3z\cos\gamma} \\ -jk_3\sin\gamma e^{-jk_3z\cos\gamma} & -jk_3\sin\gamma e^{jk_3z\cos\gamma} \\ jk_3\cos\gamma e^{-jk_3z\cos\gamma} & -jk_3\cos\gamma e^{jk_3z\cos\gamma} \end{bmatrix}. \tag{28}$$

IV. CONTINUITY RELATIONS AT THE INTERFACES

A. Objective

The objective of this section is to examine continuity relations for all possible interfaces. These relations depend on the nature of the materials of adjacent layers. Six combinations involving the three materials described above have to be taken into account.

B. Continuity relations

- Fluid/Fluid interface (F/F): Continuity relations between fluid layers involve both pressure and normal displacement. One obtains

$$u_{z_i}(z_i^{i+1}) = u_{z_{i+1}}(z_i^{i+1}),$$

$$p_i(z_i^{i+1}) = p_{i+1}(z_i^{i+1}). \tag{29}$$

- Solid/Solid interface (S/S): Continuity relations between solid layers involve stress and displacement components. One obtains

$$\mathbf{u}_i(z_i^{i+1}) = \mathbf{u}_{i+1}(z_i^{i+1}),$$

$$\mathbf{t}_i(z_i^{i+1}) = \mathbf{t}_{i+1}(z_i^{i+1}). \tag{30}$$

- Fluid/Solid interface (F/S): Continuity relations between a fluid layer and a solid one only involve normal stress and normal displacement components. One obtains

$$u_{z_i}(z_i^{i+1}) = u_{z_{i+1}}(z_i^{i+1}),$$

$$-p_i(z_i^{i+1}) = \sigma_{zz_{i+1}}(z_i^{i+1}),$$

$$0 = \sigma_{xz_{i+1}}(z_i^{i+1}). \tag{31}$$

- Poroelastic/Poroelastic interface (P/P): Continuity relations between poroelastic layers involve displacement components of the solid phase, conservation of the fluid flow, total normal stress components, fluid pressure $p_i = \sigma_i^F / \phi_P$, and shear stress in the solid phase. One has

$$\begin{cases} \mathbf{u}_i^{\mathbf{S}}(z_i^{i+1}) = \mathbf{u}_{i+1}^{\mathbf{S}}(z_i^{i+1}), \\ \Phi_{P_i}[u_{z_i}^F(z_i^{i+1}) - u_{z_i}^S(z_i^{i+1})] = \Phi_{P_{i+1}}[u_{z_{i+1}}^F(z_i^{i+1}) - u_{z_{i+1}}^S(z_i^{i+1})], \\ \sigma_{zz_i}^S(z_i^{i+1}) + \sigma_{zz_i}^F(z_i^{i+1}) = \sigma_{zz_{i+1}}^S(z_i^{i+1}) + \sigma_{zz_{i+1}}^F(z_i^{i+1}), \\ \dfrac{\sigma_{zz_i}^F(z_i^{i+1})}{\Phi_{P_i}} = \dfrac{\sigma_{zz_{i+1}}^F(z_i^{i+1})}{\Phi_{P_{i+1}}}, \\ \sigma_{xz_i}^S(z_i^{i+1}) = \sigma_{xz_{i+1}}^S(z_i^{i+1}). \end{cases} \tag{32}$$

- Poroelastic/Fluid interface (P/F): Continuity relations between a poroelastic layers and a fluid layer involve conservation of the fluid flow and normal stress components in each phase, shear stress in the solid phase vanishes. Thus we obtain

$$(1-\Phi_{P_i})u_{z_i}^S(z_i^{i+1}) + \Phi_{P_i}u_{z_i}^F(z_i^{i+1}) = u_{z_{i+1}}^F(z_i^{i+1}),$$

$$\frac{\sigma_{zz_i}^F(z_i^{i+1})}{\Phi_{P_i}} = -p_{i+1},$$

$$\frac{\sigma_{zz_i}^S(z_i^{i+1})}{1-\Phi_{P_i}} = -p_{i+1},$$

$$\sigma_{xz_i}^S(z_i^{i+1}) = 0. \tag{33}$$

- Poroelastic/Solid interface (P/S): Continuity relations between a poroelastic layer and a solid layer involve displacement components of the solid phase and continuity of fluid/solid normal displacement. One obtains

$$\begin{cases} \mathbf{u}_i^{\mathbf{s}}(z_i^{i+1}) = \mathbf{u}_{i+1}^{\mathbf{s}}(z_i^{i+1}), \\ u_{z_i}^F(z_i^{i+1}) = u_{z_{i+1}}^S(z_i^{i+1}), \\ \sigma_{zz_i}^F(z_i^{i+1}) + \sigma_{zz_i}^S(z_i^{i+1}) = \sigma_{zz_{i+1}}^S(z_i^{i+1}), \\ \sigma_{xz_i}^S(z_i^{i+1}) + \sigma_{xz_{i+1}}^S(z_i^{i+1}). \end{cases} \tag{34}$$

C. Matricial relations of continuity

Expressions (29)–(34) allow one to obtain continuity relations (7) for each kind of interface. Matricial expressions are summarized in Table I.

V. OPTIMIZATION

A. Definition of the problem

As mentioned in the Introduction, an optimization problem with regard to the question of multilayered panel design often arises. Insulation properties of such panels are very dependent on the material selected, layer thicknesses and layer combination. Acoustical engineering seeks to find the configuration, for example, that minimizes the acoustical transmissibility of the panel. In addition, minimizing the mass of the panel could be a second objective that designers have to take into account and the total thickness of the panel is often a constraint to satisfy. Another limiting condition is the restricted number of available materials and the maximal number of layers. Costs and environmental factors often limit the choice to a list of predefined materials. This kind of problem is called a constrained multiobjective optimization problem, it is defined by the following data:

- The insulation properties and mass of the whole panel define n-parameter objective functions.
- Materials, layer sequences and layer thicknesses define the n parameters of the objective functions, these parameters are often called design variables.
- The total thickness of the panel, the maximal number of layers and a restricted list of materials define the constraints of the problem. A solution that satisfies these constraints is termed a "feasible solution."

The multiobjective optimization problem consists in finding n parameters that minimize the objective functions and that satisfy the constraints. In general, for this kind of problem, there is no set of n parameters that minimize both the two

TABLE I. Matricial expressions of continuity relations.

	$[I_i]$	$[G_i]\cdot\{\varphi_i\}$	$[I_{i+1}]$	$[G_{i+1}]\cdot\{\varphi_{i+1}\}$
F/F	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	$\begin{Bmatrix}u_z\\ p\end{Bmatrix}$	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	$\begin{Bmatrix}u_z\\ p\end{Bmatrix}$
S/S	$\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$	$\begin{Bmatrix}t_z\\ u_z\\ t_x\\ u_x\end{Bmatrix}$	$\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$	$\begin{Bmatrix}t_z\\ u_z\\ t_x\\ u_x\end{Bmatrix}$
F/S	$\begin{bmatrix}-1 & 0\\ 0 & 1\\ 0 & 0\end{bmatrix}$	$\begin{Bmatrix}u_z\\ p\end{Bmatrix}$	$\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}$	$\begin{Bmatrix}t_z\\ u_z\\ t_x\\ u_x\end{Bmatrix}$
P/P	$\begin{bmatrix}\frac{1}{\phi_{P_i}} & 0 & 0 & 0 & 0 & 0\\ 1 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & \phi_{P_i} & -\phi_{P_i} & 0\\ 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 1\end{bmatrix}$	$\begin{Bmatrix}p^F\\ t_z^S\\ t_x^S\\ u_z^F\\ u_z^S\\ u_x^S\end{Bmatrix}$	$\begin{bmatrix}\frac{1}{\phi_{P_i}} & 0 & 0 & 0 & 0 & 0\\ 1 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & \phi_{P_i} & -\phi_{P_i} & 0\\ 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 1\end{bmatrix}$	$\begin{Bmatrix}p^F\\ t_z^S\\ t_x^S\\ u_z^F\\ u_z^S\\ u_x^S\end{Bmatrix}$
P/F	$\begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & \phi_{P_i} & 1-\phi_{P_i} & 0\end{bmatrix}$	$\begin{Bmatrix}p^F\\ t_z^S\\ t_x^S\\ u_z^F\\ u_z^S\\ u_x^S\end{Bmatrix}$	$\begin{bmatrix}\phi_{P_i} & 0\\ 1-\phi_{P_i} & 0\\ 0 & 0\\ 0 & 1\end{bmatrix}$	$\begin{Bmatrix}p^F\\ t_z^S\\ t_x^S\\ u_z^F\\ u_z^S\\ u_x^S\end{Bmatrix}$
P/S	$\begin{bmatrix}1 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 1\end{bmatrix}$	$\begin{Bmatrix}p^F\\ t_z^S\\ t_x^S\\ u_z^F\\ u_z^S\\ u_x^S\end{Bmatrix}$	$\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$	$\begin{Bmatrix}p^F\\ t_z^S\\ t_x^S\\ u_z^F\\ u_z^S\\ u_x^S\end{Bmatrix}$

objective functions but there exists some sets that are not dominated in a Pareto sense. These sets are called Pareto optimal solutions of the multiobjective optimization problem and are generally not unique. Pareto optimal solutions are such that there is no other set of n parameters that are better in all design objectives. Therefore, the leading problem consists of obtaining one of these optimal solutions or all of them.

B. Objective functions

Objective functions of a multiobjective optimization problem define the criteria that are to be minimized or maximized. The TL spectrum, that is, the transmission loss function evaluated on a given frequency range, is used to evaluate the performances of a feasible solution. However, instead of maximizing TL, the problem often consists of finding a layer configuration that leads to an objective TL spectrum chosen in advance. Therefore, the problem consists in minimizing the difference between the TL spectrum and the objective spectrum $\mathrm{TL}^{\mathrm{obj}}$.

First, the frequency range of interest is split into n bandwidths, I_i. On each bandwidth, an objective transmission loss $\mathrm{TL}_i^{\mathrm{obj}}$ is defined. The TL is evaluated for three frequencies in each bandwidth and for five incidence angles, the obtained values give rise to a mean value R_i that is compared to the objective value $\mathrm{TL}_i^{\mathrm{obj}}$. The objective function is given by a scalar value E that is obtained according to

$$E = \max_i E_i \tag{35}$$

with

$$E_i = R_i^{\mathrm{obj}} - R_i \quad \text{if } R_i \leq R_i^{\mathrm{obj}}.$$

$$E_i = 0 \quad \text{if } R_i > R_i^{\mathrm{obj}}.$$

The second objective function is the mass of the whole panel, as for TL a more general approach consists of finding a solution that minimizes the difference with an objective mass M^{obj}. The objective mass could be zero if the mass is to be minimized. Therefore, the second objective function G is obtained according to

$$G = \log\frac{M}{M^{\text{obj}}} \quad \text{if } M \geq M^{\text{obj}},$$

$$G = 0 \quad \text{if } M < M^{\text{obj}}. \tag{36}$$

Pareto optimal solutions for such multiobjective optimization problems are by no means unique. Additional information needs to be introduced to select one. The way this information is introduced leads to three different approaches:

- *A priori* methods;
- Progressive methods;
- *A posteriori* methods.

The first consists of introducing preferences at the beginning of the search process. The simplest method is based on a weighted combination of the objective functions into one (Steuer, 1986). Therefore, the resulting function is used as the objective function of a single objective optimization problem. The obtained solution is an element of a Pareto optimal set that is a function of the weighting used.

Progressive methods consist in introducing preference criteria during the optimization process (Benayoun *et al.*, 1971). The automatization of these approaches is very difficult and depends on the nature of the problem.

Finally, *a posteriori* methods consist of selecting an optimum solution among all the Pareto optimal solutions calculated by the iterative process. The main difficulty is to obtain all the feasible optimal solutions instead of one. To this end, specialized multiobjective genetic algorithms have been developed. They are supposed to converge to the Pareto optimal set or a Pareto optimal subset (Schaeffer, 1985).

For the problem described in this paper, an *a priori* method has been used with an objective function given by the following combination:

$$F = E + 10 \,.\, G. \tag{37}$$

C. Design variables

Design variables are the objective function's parameters. For the present problem, these variables have to define the material properties, the layer sequence, and the thickness of each layer. Moreover, they must satisfy the problems's constraints. As mentioned above, three kinds of constraints are considered:

- A restricted list of M available materials;
- A maximal number L of layers;
- A maximal thickness $T_{\max}$ for the whole panel.

Variables that characterize material properties have to satisfy the first constraint. This implies that these variables are discrete. The maximal number is a finite integer, therefore parameters that define layer sequences are discrete as well. Finally, the only continuous parameters are those that define the thickness of each layer. In order to formulate a discrete variable optimization problem, thicknesses could be expressed as discrete variables if they are multiples of a given elementary thickness t.

One means to satisfy both the first and second constraints would be to define L integer variables that identify the constitutive material for each layer in a given order and an integer variable that indicates how many layers have to be considered. Let m_i be the integer that identifies the material for layer i in the predefined list and l the total number of layers. The first two constraints are given by

$$1 \leq m_i \leq M, \quad i \in 1, \ldots, L, \tag{38}$$

$$1 \leq l \leq L. \tag{39}$$

Let n_i be the integer variable that characterizes the thickness t_i of the layer i according to

$$t_i = n_i \times t. \tag{40}$$

The third constraint is given by

$$t\sum_{i=1}^{l} n_i \leq T_{\max}. \tag{41}$$

Therefore, the parameters of the objective function F are: $m_1, \ldots, m_L$, $t_1, \ldots, t_L$, and l.

D. Genetic algorithm

1. Principle

Continuous variable optimization problems are often formulated and solved using traditional gradient based methods. Discrete variable problems are sometimes converted into continuous variable problems but it may not be warranted for many problems. In these cases, stochastic search techniques are very good alternatives especially when objective functions should be rapidly evaluated.

For the problem described above, the Transfer Matrix Method allows one to characterize acoustical performances of multi-layered panels with a low calculus cost time. Moreover, the problem has been formulated with discrete design variables. This leads us to opt for a Genetic Algorithm (GA) approach.

GAs are evolutionary algorithms based on Darwin's survival of the fittest theory. The principle of such algorithms is quite simple. The n design parameters are coded in a chromosome. An initial randomized population of such chromosomes is built and each individual is evaluated with the objective function. The best individuals are selected to give rise to the next generation. Individuals of the next generation are obtained from crossing and mutation procedures between chromosomes of previous selected individuals. The iterative procedure is performed over several generations and the best individual from the resulting population is often a good approximation of the optimization problem's solution.

2. Chromosome coding

Each individual in the population is described by a chromosome that codes the design variables that describe a potential panel solution. These design variables have been defined in Sec. V C and are as follows:

- Material identification of the layers given by L integers: $m_1, \ldots, m_L$;

- Thickness of the layers given by L integers: $t_1, \ldots, t_L$;
- Number of effective layers given by an integer l.

The information is collected in the individual's chromosome according to the following representation:

$$n \quad m_1 \quad n_1 \quad m_2 \quad n_2 \quad \ldots \quad m_l \quad n_l \quad \ldots \quad m_L \quad n_L.$$

L layers are described but only the first l layers are taken into account in the calculation of the objective function. The last $l-n$ pairs of integers are only used in crossing and mutation processes. The thickness integer n_i is a number between 1 and 1000 that defines the thickness t_i according to relation (40). The parameter t is given at the beginning of the calculation process.

3. Parameters of the algorithm

The major elements of a GA that uses a population of p individuals are here described:

- Randomized initialization of the p individuals, the resulting population is denoted P_0;
- Evaluation of the p individuals of P_0 with the objective functions;
- for i from 1 to G:
  - — selection of p_b individuals in P_{i-1};
  - — reproduction of p_c children from the p_b individuals;
  - — mutation of a fraction of the p_c children;
  - — insertion of the p_c children in the population P_i;
  - — evaluation of the p individuals of P_i;
- The solution is the best individual in the population P_G.

Parameter p has to be large enough to obtain an efficient convergence of the algorithm. It may be 50 or 100. Parameter G is the number of generations; it has to lead to a final population that contains the solution of the optimization problem. It may be increased if convergence is not observed. Typically, the number of generations depends on the problem and may be several hundreds.

The evaluation process consists in mapping the objective function over the whole population. The obtained values are indicators of how individuals have performed in the problem domain, however a fitness function is used to transform these values into a measure of relative fitness. The fitness function f that has been used to evaluate relative fitness of an individual x_k is

$$f(x_k) = \frac{F(x_k)}{\sum_{k=1}^{p} F(x_k)}. \tag{42}$$

The selection process consists of choosing individuals for reproduction on the basis of their relative fitness. A Roulette wheel selection method is used. This method consists of the association of a real number interval to each individual. The intervals are not overlapping and their size is proportional to the relative fitness of the associated individual. A random number is generated and the individual associated with the interval that contains the number is selected for reproduction. This selection process is repeated p_b times. The number p_b depends on the size of the next generation and the reproduction process.

For the described algorithm, the size of the next genera-

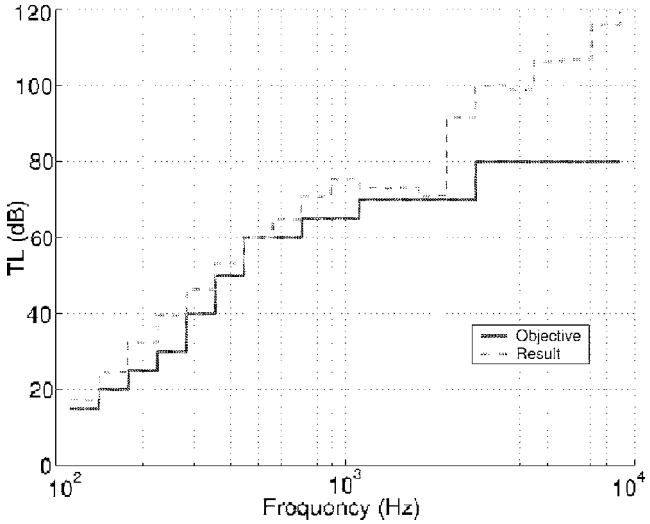


FIG. 2. Objective and obtained solution TL spectrum.

TABLE II. Available material list.

Property	ϕ_P	σ_R (N s/m^4)	α_∞	Λ (μm)	Λ' (μm)	ρ (kg/m^3)	E (Pa)	ν	η	c_o (m/s)
Air	⋯	⋯	⋯	⋯	⋯	1.225	⋯	⋯	⋯	342
Steel	⋯	⋯	⋯	⋯	⋯	7800	210.10^9	0.3	0.001	⋯
Aluminium	⋯	⋯	⋯	⋯	⋯	2700	70.10^9	0.3	0.001	⋯
Heavy mass	⋯	⋯	⋯	⋯	⋯	2100	77.10^6	0.35	0.21	⋯
Loaded rubber	⋯	⋯	⋯	⋯	⋯	2100	77.10^6	0.35	0.21	⋯
Rubber	⋯	⋯	⋯	⋯	⋯	1000	100.10^6	0.45	0.2	⋯
Glass wool 1	0.98	35000	1	60	150	9.6	10.10^3	0	0.2	⋯
Glass wool 2	0.98	55000	1	44	130	16	50.10^3	0	0.2	⋯
Foam 1	0.81	550000	1.5	10.5	31.5	211	344.10^3	0.33	0.15	⋯
Foam 2	0.96	480000	1.38	16.1	48.3	11	234.10^3	0.33	0.09	⋯
Foam 3	0.97	17000	1.02	240	490	10	505.10^3	0.3	0.07	⋯
Foam 4	0.95	12000	1	50	60	40	90.10^3	0.3	0.1	⋯

TABLE III. Combination solution 1.

Layer	Material	Thickness
1	Loaded rubber	4 mm
2	Foam 3	32.6 mm
3	Air	0.8 mm
4	Foam 3	22.2 mm
5	Foam 2	7 mm
6	Heavy mass	3.8 mm
7	Glass wool 1	9.6 mm

tion is $0.9\times p$. The reproduction process is such that two parents give rise to two children, therefore the number p_b is $0.9\times N$.

The reproduction consists in producing new chromosomes and is based on a multicrossover process. l positions in chromosome coding are selected in a random way. Information coded into a pair of chromosomes between two successive positions is exchanged, therefore two new chromosomes are produced. A crossover rate of 0.7 is used. This implies that only 70% of the selected parents are combined.

The next step is a mutation process of some children. The mutation consists of introducing randomized information into a chromosome. The mutation rate gives the probability that random information within a child's chromosome will be introduced, and its value is 0.5%.

The last step is insertion of children into the old population. This insertion is based on replacing the least fit individuals by the new generation.

E. Applications

1. Optimization of a multilayered panel

The described GA based on the Transfer Matrix Method has been implemented in a MATLAB environment. The program thus obtained has been used to obtain optimal layer combinations characterized by an objective loss transmission spectrum and an objective mass. An example of these results is presented in this section.

The objective loss transmission spectrum is defined over $(0, 10\,000$ Hz), and is represented in Fig. 2 a solid line.

The mass objective is set to 5 kg m^{-2}. A list of 11 available materials is given in Table II.

The genetic algorithm has been applied for a population of 100 individuals. Generations have been performed until the best performances converged. This has been observed after 200 generations.

The loss transmission spectrum of the converged solution is given in Fig. 2 with a dashed line. Its level is higher than the objective. The combined solution is given in Table III.

The combination is classic and corresponds to a high performance panel, and consists of a double walled system. The internal medium is a foam with high absorption properties combined with a thin air layer (<1 mm) that reduces acoustical transmissibility in the solid phase. Figure 3 shows the influence of an air gap, and the transmissibility objective

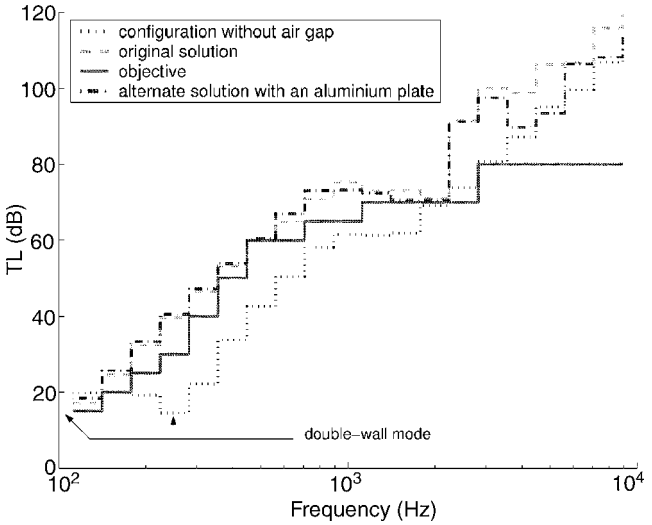


FIG. 3. TL spectra for modified solutions.

is clearly reached. In this case, the substitution of the rubber layer with a rigid one (aluminum) does not affect the performances of the panel.

The solution thus obtained is one of the Pareto optimal solutions. As mentioned above, this kind of solution is not unique and depends on the choice of the objective function as given by expression (37). Here, the solution obtained satisfies the TL spectrum objective, but the mass parameter is not minimized. The density of this Pareto-optimal solution is approximately 17 kg m^{-2}, and is the best that may be obtained for the given spectrum objective. In this example, the chosen objective function gives greater weight to the TL spectrum objective. One can modify the objective function to give greater weight to the density. More generally, an objective function could be defined with expression (43),

$$F = E + kG, \tag{43}$$

where k is a constant parameter.

The greater the parameter k, the greater the weight for the mass objective in the search for a Pareto solution. For example, another simulation was performed with $k=50$.

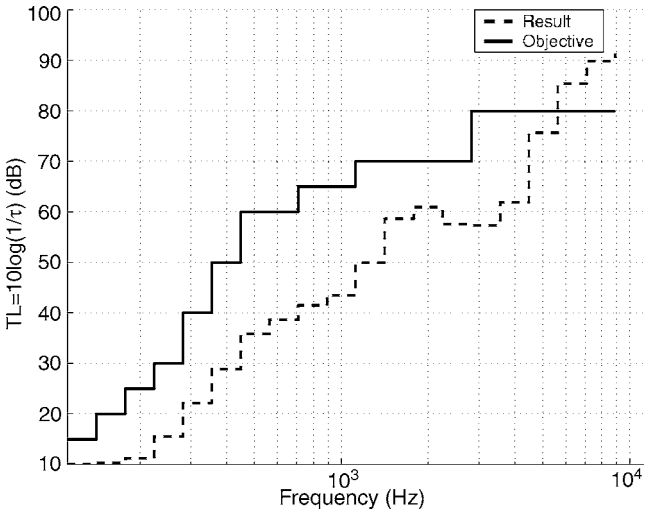


FIG. 4. TL spectrum with a modified objective function ($k=50$).

TABLE IV. Combination solution 2.

Layer	Material	Thickness
1	Loaded rubber	2.6 mm
2	Foam 3	17.4 mm
3	Air	0.1 mm
4	Foam 3	34.4 mm
5	Foam 2	17.4 mm
6	Foam 1	8.1 mm

TABLE V. Sandwich panel combination.

Layer	Material	Thickness
1	Light composite plate	(specified)
2	Foam 1	5.0 mm
3	Foam 2	10.0 mm
4	Air	<1 mm
5	Foam 1	9.0 mm
6	Dense composite plate	(specified)

Mass and TL spectrum objectives were identical to those in the previous simulation but priority was given to the density. The TL spectrum for the stabilized solution is given in Fig. 4 and the combined solution is given in Table IV.

Clearly, the TL spectrum objective is not satisfied but the density is minimized. The solution obtained is a lightweight combination of a rubber sheet and foam layers with an air gap that has a density of 5 kg m^{-2}. The TL spectrum solution is not minimized but is the best that may be obtained for the given density objective. It is an optimal solution in a Pareto sense.

2. Examples of industrial design

The present method has been used for designing acoustical panels in aircraft construction. Two examples of such applications are described hereafter. The multilayered light panels are part of an airplane fuselage. Some of the layers correspond to specified parameters, therefore the chromosome coding is modified to take into account these fixed layers.

In the first example, external layers are specified for a sandwich panel. The optimization problem is related to the choice of internal layers for a 25 mm total thickness. The mass objective for internal layers is 2.5 kg m^{-2}. The TL spectrum objective and the obtained solution are given in Fig. 5, and the combined solution is given in Table V.

Therefore, a 1 m $\times$ 1 m multilayered panel very similar to that proposed by the optimal solution has been manufactured. A special grid has been used to preserve the air gap. The acoustical transparency of this panel has been tested and the measured TL spectrum is compared with the objective in Fig. 6.

One can notice that the TL spectrum is in close agreement with the prescribed objective. For frequencies greater than 2000 Hz, the processed TL spectrum is overestimated. This discrepancy often arises for high TL levels (>60 dB), it is usually explained by a loss of precision in the experimental setup and especially as a result of imperfect boundary conditions.

The second example involves the design of a thicker double-walled panel (60 mm). The results are given in Fig. 7 and the combined solution is given in Table VI. The obtained results closely agree with the prescribed objective.

These two latter simulations have been conducted in an industrial context and provide a good idea of the effectiveness of the method. In both situations, the optimization algorithm allows a reduction in the number of physical tests that are required for the development of these insulation products. It provides an initial solution that can be afterwards adapted to the manufacturing constraints. The design engineer can improve the given solution with the use of other types of simulation software and a few experimental tests.

The method's performances are attractive as a result of the compromise between the quick analytical approach, the precision in Biot's models, and the practical constraints related to a restricted library of realistic materials.

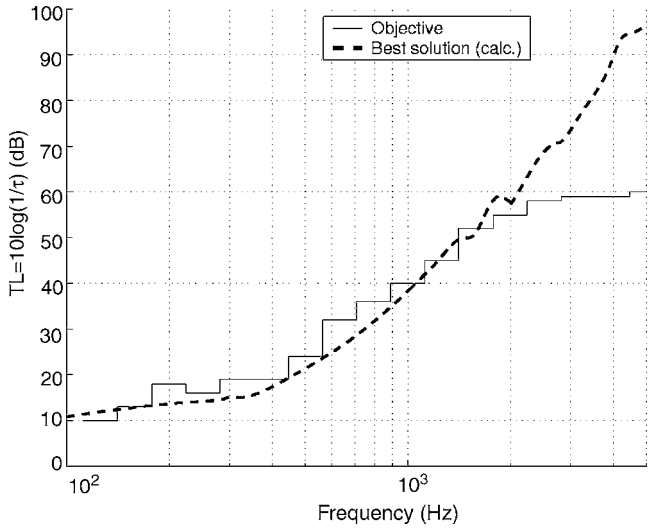


FIG. 5. Objective and obtained solution TL spectrum (sandwich panel).

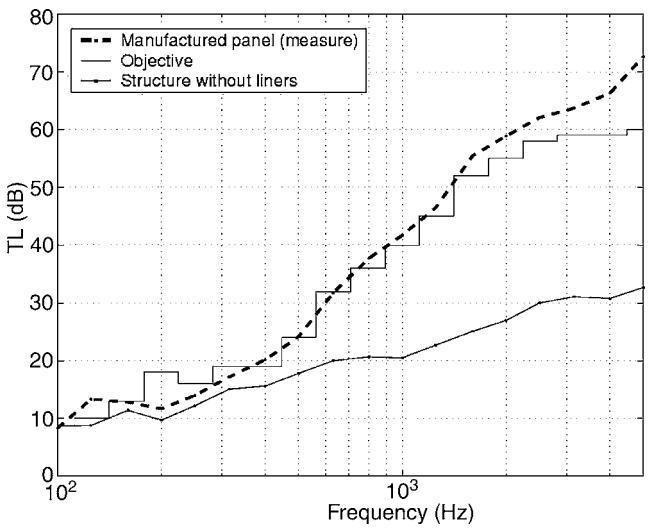


FIG. 6. Objective and manufactured panel TL spectra (sandwich panel).

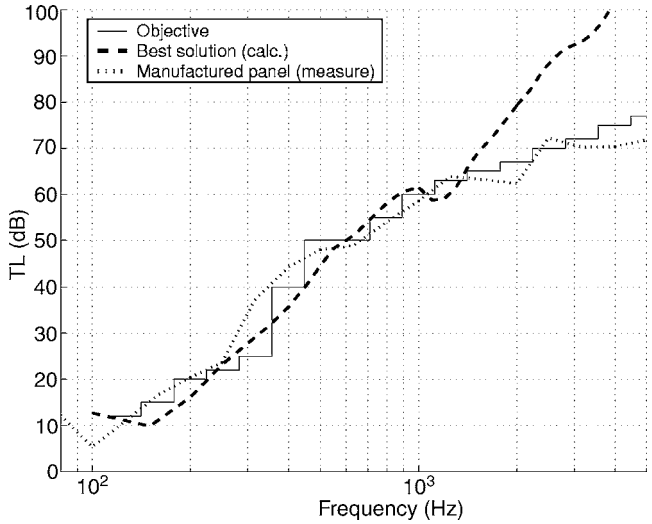


FIG. 7. Objective and manufactured panel TL spectra (thick sandwich panel).

VI. CONCLUSION

Optimization of multilayered panels is a complex problem that implies multiobjective criteria. The design variables may be discrete or continuous and genetic algorithms are efficient approaches to obtain optimal solutions for such problems. The major limitation for the use of genetic algorithms is the time cost. The nature of this kind of algorithm implies that the objective function could be evaluated rapidly. In this context, the Transfer Matrix Method is an efficient tool. In this paper, a very simple genetic algorithm is combined with the Transfer Matrix Method and is used to solve a real application. The solution is not novel but this shows that such problems could be easily treated with these tools. In order to obtain a set of Pareto optimal solutions, specialized multiobjective genetic algorithms (Schaeffer, 1985) may well be as efficient as the algorithm used herein.

TABLE VI. Sandwich panel combination.

Layer	Material	Thickness
1	Aluminum	(specified 1.2 mm)
2	Glass wool	19.0 mm
3	Foam 1	9.0 mm
4	Glass wool	8.0 mm
5	Foam 2	18.0 mm
6	Lightweight plate	(specified 5.0 mm)

ACKNOWLEDGMENTS

The authors wish to thank the PAULSTRA company for its financial support and Michel Pompei (PAULSTRA) for his helpful collaboration.

Allard, J. F., Champoux, Y., and Depollier, C. (**1987**). "Modelization of layered sound absorbing materials with transfer matrices," J. Acoust. Soc. Am. **82**, 1792–1796.

Allard, J. F. (**1993**). *Propagation of Sound in Porous Media* (Elsevier Applied Science, New York).

Anderson, J. (**2000**). "A survey of multiobjective optimization in engineering design," Technical Report No. LiTH-IKP-R-1097, Department of Mechanical Engineering, Linkping University.

Benayoun, R., de Montgolfier, J., Terny, J., and Laritchev, O. (**1971**). "Linear programming with multiple objective functions: Step Method (STEM)," Math. Program. **1**, 366–375.

Biot, M. A. (**1956**). "The theory of propagation of elastic waves in a fluid saturated porous solid," J. Acoust. Soc. Am. **28**, 168–191.

Biot, M. A. and Willis, D. G. (**1957**). "The elastic coefficients of the theory of consolidation," J. Appl. Mech. **24**, 594–601.

Bolton, J. S., Shiau, N. M., and Kang, Y. J. (**1996**). "Sound transmission through multi-panel structures lined with elastic porous materials," J. Sound Vib. **191**, 312–347.

Brekhovskikh, L. M. (**1960**). *Waves in Layered Media* (Academic, New York).

Delany, M. E. and Bazley, N. E. (**1970**). "Acoustic properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116.

Fonseca, C. and Fleming, P. (**1995**). "An overview of evolutionary algorithms in multiobjective optimization," Evol. Comput. **3**, 1–18.

Germain, C. (**1999**). "Optimisation vibro-acoustique de panneaux intérieurs de portes automobiles," vibroacoustic optimization of car door panels, thesis in French, Université de Valenciennes.

Glover, F. (**1989**). "Tabu Search - Part I," ORSA J. Comput. **1**, 190–206.

Goldberg, D. E. (**1989**). *Genetic Algorithms for Search, Optimization and Machine Learning* (Addison-Wesley, Reading).

Jang, J. R., Sun, C., and Mizutani, E. (**1997**). *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence* (Prentice-Hall, Englewood Cliffs).

Johnson, D. L. (**1986**). "Recent developments in the acoustic properties of porous media," in *Proceedings of the International School of Physics Enrico Fermi, Course XCIII* (D. Sette North-Holland, Amsterdam), pp. 255–290.

Kirkpatrick, S., Gelatt, C. D., Jr., and Vecchi, M. P. (**1983**). "Optimization by simulated annealing," Science **220**, 671–680.

Lauriks, W., Cops, A., Allard, J. F., Depollier, C., and Rebillard, P. (**1990**). "Modelization at oblique incidence of layered porous materials with impervious screens," J. Acoust. Soc. Am. **87**, 1200–1206.

Lauriks, W., Mees, P., and Allard, J. F. (**1992**). "The acoustic transmission through layered systems," J. Sound Vib. **155**, 125–132.

Mulholland, K. A., Price, A. J., and Parbook, H. D. (**1968**). "TL of multiple panels in a random incidence field," J. Acoust. Soc. Am. **43**, 1432–1435.

Schaeffer, J. (**1985**). "Multiple objective optimization with vector evaluated genetic algorithms" in 1st international Conference on Genetic Algorithms, Pittsburg.

Steuer, R. (**1986**). *Multiple Criteria Optimization: Theory, Computation and Application* (Wiley, New York).

Vashishth, A. K. and Khurana, P. (**2004**). "Waves in stratified anisotropic poroelastic media: A transfer matrix approach," J. Sound Vib. **277**, 239–275.

Xu, J., Nannariello, J., and Fricke, F. R. (**2004**). "Optimizing flat-walled multilayered anechoic lining using evolutionary algorithms," Appl. Acoust. **65**, 1009–1026.

A boundary element method for porous media

Olivier Tanneau
Vibro Acoustic Group, Institute of Mechanics of Paris, Saint-Ouen, France

Pierre Lamary
Vibro Acoustic Laboratory, State University of Campinas, Campinas, São Paulo, Brazil

Yvon Chevalier
Vibro Acoustic Group, Institute of Mechanics of Paris, Saint-Ouen, France

(Received 18 September 2005; revised 10 April 2006; accepted 6 June 2006)

This paper deals with a new method for solving coupled acoustical problems such as those arising from the study of insulation panels with porous linings. The main idea is to use the boundary element method (BEM) for modeling porous media in order to simplify problem solving in mid-frequency range. This approach reduces the entire problem to only unknowns on the boundaries. Developments have been oriented in a bi-dimensional formulation with a constant element discretization scheme. A multi-region assembly of porous subdomains and a coupling procedure with structural finite elements complete this approach. Comparisons between the finite element method (FEM) and the present solution (mixed FEM and BEM) show a close agreement. Moreover, the analysis of a multi-layer system through this mixed numerical method results in faster processing time and less memory usage than a conventional full finite element method. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2221407]

PACS number(s): 43.20.Gp, 43.20.Jr, 43.20.Tb [LLT] Pages: 1239–1251

I. INTRODUCTION

The theory of Biot[1,2] for porous materials is of major interest concerning acoustical purposes such as absorption or insulation problems. Over the last decade, various research studies[3–5] have used the finite element method (FEM) for numerical analysis of continuous porous media in different contexts such as single- or multi-layer samples in impedance tubes or the transmission loss of sandwich panels. This approach leads to very efficient results but also to great computer processing time and memory usage. This induced cost results from two reasons: firstly the relatively high number of unknowns per node (because of the intrinsic bi-phase nature of porous media) and, secondly, the high number of elements necessary to obtain convergence in mid-frequency range. The number of nodes per wavelength has to be set to at least six and often more. Ill-conditioning of the assembly matrix may also be performed by means of such a method. Recent works[6] based upon *n*-order Legendre polynomia (hierarchical finite elements) demonstrate that high-order finite elements are required in most practical cases.

Consequently, modeling complex industrial systems such as insulation panels is a strenuous task, especially if the frequency range of interest is wide. The present work is an attempt to propose an alternative to the FEM approach using the boundary element method (BEM) for porous domains.

The BEM is a numerical method working with discrete unknowns positioned only on the boundary of the domain. As a result, there is no need for interior meshing and the number of unknowns increases more slowly with frequency than for a complete FEM mesh. Based upon a boundary integral formulation equivalent to the local partial differential equation of equilibrium, the method is potentially applicable to a wide range of physical problems and the reader can find general explanations in references such as those by Banergee[7] or Brebbia.[8]

Some authors[7] investigated the bi-phase coupled problem of poroelasticity or thermoelasticity in the early 1980s. More recently, the BEM has been applied to porous media in geomechanics where the possibility that the BEM may model semi-infinite geometry holds major interest.[9–11] Generally based upon simplified versions of the Biot model dedicated to soil consolidation, this work needs to be adapted to absorbing materials such as glass wool or foams.

First, this paper presents the theoretical background for the study. Next, two situations are investigated: a simple porous domain model and the simulation of the transmission of a small double-wall partition. Comparisons are made with the use of a classical finite element method in order to validate the present work.

II. THEORETICAL ASPECTS

A. Governing equations

The Biot model,[1] in its frequency domain version, is the starting point for the present work. The Johnson-Champoux-Allard[2] model links measurable data and coefficients of the equations. In this theory using a *u-p* formalism,[3] a displacement field u^S and a stress field $\hat{\sigma}^S$ describe the solid phase of the porous medium. On the other hand, a pressure field p and an associated flux q are the unknowns of the fluid phase. The equilibrium of the two phases is thus expressed

$$\hat{\sigma}^S_{ij,j} + \omega^2 \tilde{\rho}_e u^S_i + \tilde{\gamma} p_{,i} = f^S_i, \tag{1}$$

$$q_{i,i} + \kappa k^2 p - \tilde{\gamma} u^S_{i,i} = f^p \tag{2}$$

The constitutive laws are $\hat{\sigma}^S = \lambda \varepsilon^S_{kk} \delta_{ij} + 2\mu \varepsilon^S_{ij}$ and $q_i = \kappa p_{,i}$. The material coefficients are defined by the following:

(i) An equivalent mass of the solid phase $\tilde{\rho}_e$

This complex mass is related to

$$\tilde{\rho}_e = \tilde{\rho}_{11} - \frac{\tilde{\rho}_{12}^2}{\tilde{\rho}_{22}}$$

and

$$\tilde{\rho}_{11} = \rho_1 + \rho_a + \frac{1}{j\omega}\tilde{b},$$

$$\tilde{\rho}_{12} = -\rho_a - \frac{1}{j\omega}\tilde{b},$$

$$\tilde{\rho}_{22} = \phi_P \rho_o + \rho_a + \frac{1}{j\omega}\tilde{b},$$

where

$$\rho_a = \phi_P \rho_o (\alpha_\infty - 1),$$

and, with the use of the Johnson model

$$\tilde{b} = \sigma_R \phi_P^2 \sqrt{1 + \frac{4j\alpha_\infty^2 \eta \rho_o \omega}{\sigma_R \Lambda^2 \phi_P^2}}.$$

The main entries for the porous materials are as follows: ρ_1 is the density of the skeleton (in vacuo), ϕ_P is the porosity, σ_R is the resistivity (in $\mathrm{Ns/m^4}$), α_∞ is the tortuosity, and Λ is the viscous length (in μm).

The following complementary parameters involve air properties: dynamic viscosity $\eta = 1.84e^{-5}$ Pa.s and density $\rho_o = 1.225\ \mathrm{Kg.m^{-3}}$

(ii) An equivalent wave number of the fluid phase k is

$$k^2 = \omega^2 \frac{\tilde{\rho}_{22}}{R}.$$

$R = \phi_P K(\omega)$ is the equivalent bulk modulus of the fluid phase (with the hypothesis of a perfect rigid skeleton).

The Champoux-Allard model yields

$$K(\omega) = \frac{\gamma P_o}{\gamma - (\gamma - 1)\left(1 + \frac{8\eta}{j\Lambda'^2 N_{\mathrm{Pr}} \omega \rho_o} \sqrt{1 + j\rho_o \omega N_{\mathrm{Pr}} \Lambda'^2 / 16\eta}\right)^{-1}}.$$

Here, a new porous property is introduced: Λ'. It is a characteristic dimension (in m) influencing high-frequency thermal effects. Other air properties are specific heat ratio $\gamma = 1.4$, atmospheric pressure $P_o = 101320$ Pa, and Prandt number $N_{\mathrm{Pr}} = 0.71$.

(iii) A coupling factor $\tilde{\gamma}$ is

$$\tilde{\gamma} = \phi_P \left(\frac{\tilde{\rho}_{12}}{\tilde{\rho}_{22}} - \frac{Q}{R}\right).$$

Here Q is a coupling modulus given by

$$Q = (1 - \phi_P) K(\omega).$$

Note that R and Q are here given under the hypothesis that the bulk modulus of the fluid phase is negligible in regard to the homogeneous bulk modulus of the skeleton, itself negligible in regard to the bulk modulus of the material of the solid phase. Further expressions, for the general case, can be found in Ref. 12.

Moreover, we shall introduce the following

(iv) A scale factor κ is

$$\kappa = \frac{\phi_P^2}{\omega^2 \tilde{\rho}_{22}}.$$

This factor scales the fluid equation to the solid one having the same coupling coefficient γ.

(v) The Lame coefficients are λ and μ. These modules define the constitutive law of the solid phase *in vacuo* (no coupling with the fluid phase). In a frequency analysis, dynamic values are used and a simple model introducing a constant damping ratio η is used: $\lambda* = \lambda(1 + j\eta)$ and $\mu* = \mu(1 + j\eta)$. In the following sections, λ and μ are to be interpreted in their dynamic sense $*$.

B. Integral identities

Assume two solution states $(u^S, \hat{\sigma}^S, p, q, f^S, f^p)$ and $(u^{S*}, \hat{\sigma}^{S*}, p^*, q^*, f^{S*}, f^{p*})$ verifying Eqs. (1) and (2). In an arbitrary domain V bounded by S where both states are defined, consider the cross product:

$$A = \int_V (u_i^{S*} f_i^S - u_i^S f_i^{S*} + p^* f^p - p f^{p*}) dV. \tag{3}$$

The equilibrium equations (1) and (2) allow

$$A = \int_V (u_i^{S*} \hat{\sigma}^S_{ij,j} - u_i^S \hat{\sigma}^{S*}_{ij,j}) dV + \int_V (p^* q_{i,i} - p q^*_{i,i}) dV + \int_V \tilde{\gamma} ((u_i^{S*} p)_{,i} - (u_i^S p^*)_{,i}) dV. \tag{4}$$

An integration by parts and the divergence theorem lead to

$$\int_V u_i^{S*} \hat{\sigma}^S_{ij,j} dV = \int_S u_i^{S*} \hat{\sigma}^S_{ij} n_j dS - \int_V u^{S*}_{i,j} \hat{\sigma}^S_{ij} dV,$$

$$\int_V p^* q_{i,i} dV = \int_S p^* q_i n_i dS - \int_V p^*_{,i} q_i dV.$$

By observing the symmetry in the constitutive laws, we conclude: $u^{S*}_{i,j} \hat{\sigma}^S_{ij} - u^S_{i,j} \hat{\sigma}^{S*}_{ij} = 0$ and $p^*_{,i} q_i - p_{,i} q_i^* = 0$. The equation (4) then becomes

$$\int_V (u_i^{S*} f_i^S - u_i^S f_i^{S*} + p^* f^p - p f^{p*}) dV = \int_S (u_i^{S*} \hat{\sigma}^S_{ij} - u_i^S \hat{\sigma}^{S*}_{ij} + p^* q_j - p q_j^* + \tilde{\gamma} u_j^{S*} p - \tilde{\gamma} u_j^S p^*) n_j dS. \tag{5}$$

These integrals are an extension of the Maxwell-Betti identity. A compressed form can be written as

$$\int_V (u_i^* f_i - u_i f_i^*) dV = \int_S (u_i^* t_i - u_i t_i^* + u_i^* \gamma_{ij} u_j) dS, \tag{6}$$

where the generalized vectors u and t are

$$\mathbf{u} = \begin{Bmatrix} u_x^S \\ u_y^S \\ p \end{Bmatrix}, \quad \underline{\underline{\sigma}} = \begin{bmatrix} \hat{\sigma}_{xx}^S & \hat{\sigma}_{xy}^S \\ \hat{\sigma}_{yx}^S & \hat{\sigma}_{yy}^S \\ q_x & q_y \end{bmatrix}, \quad \mathbf{f} = \begin{Bmatrix} f_x^S \\ f_y^S \\ f^p \end{Bmatrix}$$

and

$$t_i = \sigma_{ij} n_j, \quad \underline{\underline{\gamma}} = \tilde{\gamma} \begin{bmatrix} 0 & 0 & n_x \\ 0 & 0 & n_y \\ -n_x & -n_y & 0 \end{bmatrix}.$$

The change in the unknowns $t_i' = t_i + \gamma_{ij} u_j$ and $t'^* = t^*$, is a way to write the reciprocity identity in a more classical form:

$$\int_V (u_i^* f_i - u_i f_i^*) dV = \int_S (u_i^* t_i' - u_i t_i'^*) dS. \tag{7}$$

C. Boundary integral formulation

Considering now that the state * is a case of an infinite porous domain loaded by a singular source,

$$\mathbf{f}^* = -\delta(\mathbf{y} - \mathbf{x}) \mathbf{e}^k,$$

where $\mathbf{e}^k$ is a unity vector. The index k can assume three values, each corresponding to a possible load (one for each scalar equation). As a definition of the dirac we have

$$u_k(\mathbf{x}) = \int_V (u_i(\mathbf{y}) \delta(\mathbf{y} - \mathbf{x}) e_i^k) dV$$

for $\mathbf{x} \in V, \mathbf{x} \notin S$.

For each load k, the solution of Eqs. (1) and (2) are expressed as $u_i^{k*}(\mathbf{x},\mathbf{y})$ and $t_i^{k*}(\mathbf{x},\mathbf{y})$. They are also called Green's function tensors. By using the reciprocity identity (7), the two-dimensional boundary integral representations of Eqs. (1) and (2) can be written as

$$u_k(\mathbf{x}) = -\int_V u_i^{k*}(\mathbf{x},\mathbf{y}) f_i(\mathbf{y}) dV + \int_S (u_i^{k*}(\mathbf{x},\mathbf{y}) t_i'(\mathbf{y}) - u_i(\mathbf{y}) t_i'^{k*}(\mathbf{x},\mathbf{y})) dS \tag{8}$$

If $\mathbf{x} \to S$, the limit of the equation (8) on a smooth surface (not valid for corners and sharp edges) tends to

$$\frac{1}{2} u_k(\mathbf{x}) = -\int_V u_i^{k*}(\mathbf{x},\mathbf{y}) f_i(\mathbf{y}) dV + \int_S (u_i^{k*}(\mathbf{x},\mathbf{y}) t_i'(\mathbf{y}) - u_i(\mathbf{y}) t_i'^{k*}(\mathbf{x},\mathbf{y})) dS. \tag{9}$$

The volume integral allows an interior loading that is not necessary here. So, the problem is now described by only unknowns on the boundary of the domain.

D. Constant discretization scheme

The simplest discretization of such a surface integral is a flat boundary element scheme with constant distribution of variables over the elements. If we divide the boundary S into n such elements, then Eq. (9) can be written for the fth element on the boundary as

$$\frac{1}{2} u_k(\mathbf{x}_e) = \sum_{f=1,n} t_i'(\mathbf{x}_f) \int_{\mathbf{x} \in S_f} u_i^{k*}(\mathbf{x}_e,\mathbf{x}) dS - u_i(\mathbf{x}_f) \int_{\mathbf{x} \in S_f} t_i'^{k*}(\mathbf{x}_e,\mathbf{x}) dS. \tag{10}$$

The integrals are the influence coefficients of any element over the others and they are denoted G_{ki}^{ef} and H_{ki}^{ef}. Each of these kernels is obtained using Gauss integration. A classical logarithmic Gaussian point distribution is used for the singular kernel G_{ij}^{ee} and H_{ij}^{ee} is set to $\frac{1}{2}\delta_{ij}$. The equation (10) reduces to

$$\frac{1}{2} u_k(\mathbf{x}_e) = \sum_{f=1,n} t_i'(\mathbf{x}_f) G_{ki}^{ef} - u_i(\mathbf{x}_f) H_{ki}^{ef} \tag{11}$$

and is expressed in matricial form:

$$[H] \cdot \{T\} - [G] \cdot \{U\} = \{0\}, \tag{12}$$

where

$$\{U\} = \begin{Bmatrix} \mathbf{u}(\mathbf{x}_1) \\ \mathbf{u}(\mathbf{x}_2) \\ \vdots \\ \mathbf{u}(\mathbf{x}_n) \end{Bmatrix}, \quad \{T\} = \begin{Bmatrix} \mathbf{t}'(\mathbf{x}_1) \\ \mathbf{t}'(\mathbf{x}_2) \\ \vdots \\ \mathbf{t}'(\mathbf{x}_n) \end{Bmatrix}.$$

G and H are composed of all the elementary matrices:

$$[G] = \begin{bmatrix} G^{11} & G^{12} & \cdots & G^{1n} \\ G^{21} & G^{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ G^{n1} & \cdots & \cdots & G^{nn} \end{bmatrix},$$

$$[H] = \begin{bmatrix} H^{11} & H^{12} & \cdots & H^{1n} \\ H^{21} & H^{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ H^{n1} & \cdots & \cdots & H^{nn} \end{bmatrix} + \frac{1}{2}[I].$$

The known values of $\{U\}$ and $\{T\}$ are rearranged in a vector $\{Y\}$ while the unknowns are in $\{X\}$ so that (12) can be rewritten as

$$[A]\{X\} = [B]\{Y\} \tag{13}$$

and solved for $\{X\}$. Postprocessing of the boundary solution $\{X\}$ can be performed using Eq. (8), in order to examine the interior solution.

E. Green's tensor

$u_{ik}^*(\mathbf{x},\mathbf{y})$ and $t_{ik}^*(\mathbf{x},\mathbf{y})$ are the solutions in point $\mathbf{y}$ of the equations (1) and (2) in free space with a singular load located in point $\mathbf{x}$. Two different kinds of loads have to be investigated: the two source terms of the solid equilibrium

equation in directions x and y and the source term of the fluid equation corresponding to a compressive isotropic flux load.

1. Solid loads

In this case, $f^S=-\delta(\mathbf{y}-\mathbf{x})\mathbf{e}^k$ and $f^p=0$. A decoupling of the equations can be obtained by the introduction of the potentials A_L, A_T, and A_P:

$$\mathbf{u}^S=\nabla\nabla\cdot(A_L\mathbf{e}^k)-\nabla\wedge\nabla\wedge(A_T\mathbf{e}^k),$$
$$p=\nabla\cdot(A_P\mathbf{e}^k), \tag{14}$$

Using Green's solution τ for Poisson's equation, Dirac's function can also be decomposed into

$$\Delta(\tau\mathbf{e}^k)=\nabla\nabla\cdot(\tau\mathbf{e}^k)-\nabla\wedge\nabla\wedge(\tau\mathbf{e}^k)=-\delta(\mathbf{y}-\mathbf{x})\mathbf{e}^k. \tag{15}$$

With these new potentials and loads, the equations (1) and (2) lead to

$$(\lambda+2\mu)\Delta A_L+\omega^2\tilde{\rho}_c A_L+\tilde{\gamma}A_P=\tau, \tag{16}$$

$$\kappa\Delta A_P+\kappa k^2 A_P-\tilde{\gamma}\Delta A_L=0, \tag{17}$$

$$\mu\Delta A_T+\omega^2\tilde{\rho}_e A_T=\tau. \tag{18}$$

The third is an uncoupled classical wave equation and describes the shear wave associated with the solid phase. The other two are related to the two compression waves (each is linked to one of the two phases) and are coupled. This two equations can be rewritten in matrix form:

$$\tilde{M}_1\cdot\Delta\begin{Bmatrix}A_L\\A_P\end{Bmatrix}+\tilde{M}_2\cdot\begin{Bmatrix}A_L\\A_P\end{Bmatrix}=\begin{Bmatrix}\tau\\0\end{Bmatrix} \tag{19}$$

or

$$\Delta\begin{Bmatrix}A_L\\A_P\end{Bmatrix}+\tilde{M}\cdot\begin{Bmatrix}A_L\\A_P\end{Bmatrix}=\tau\mathbf{g}, \tag{20}$$

where

$$\tilde{M}_1=\begin{bmatrix}\lambda+2\mu & 0\\ -\tilde{\gamma} & \kappa\end{bmatrix},\quad \tilde{M}_2=\begin{bmatrix}\omega^2\tilde{\rho}_e & \tilde{\gamma}\\ 0 & \kappa k^2\end{bmatrix},$$

and $\tilde{M}=\tilde{M}_1^{-1}\cdot\tilde{M}_2$ and $\tau\mathbf{g}=\tilde{M}_1^{-1}\cdot\{{}^{\tau}_{0}\}$.

The solutions are sought within an eigenvector basis of the matrix $\tilde{M}$:

$$\begin{Bmatrix}A_L\\A_P\end{Bmatrix}=\alpha_1\begin{Bmatrix}1\\r_1'\end{Bmatrix}+\alpha_2\begin{Bmatrix}1\\r_2'\end{Bmatrix}, \tag{21}$$

$$\mathbf{g}=g_1\begin{Bmatrix}1\\r_1'\end{Bmatrix}+g_2\begin{Bmatrix}1\\r_2'\end{Bmatrix}, \tag{22}$$

where r_1' and r_2' are given by

$$r_i'=\frac{(\lambda+2\mu)k_i^2-\omega^2\tilde{\rho}_e}{\tilde{\gamma}}.$$

The problem may be summarized in wave equations of the form

$$\Delta\alpha_i+k_i^2\alpha_i=g_i\tau(r), \tag{23}$$

where k_1^2 and k_2^2 are the eigenvalues of $\tilde{M}$. The shear wave equation is included in this form with $k_3^2=\omega^2\tilde{\rho}_e/\mu$ and $g_3=1/\mu$. The bi-dimensional solution α_i of this equation is

$$\alpha_i(r)=\frac{g_i}{2\pi k_i^2}(\ln r+K_0(jk_ir)), \tag{24}$$

where K_0 is the modified Bessel function of zero order. Combining this solution (24) with the equations (14) and (21) allows us to obtain the Green solutions (for $k=1,2$):

$$u_{ik}^S=A\delta_{ik}-Br_{,i}r_{,k} \tag{25}$$

and

$$p_k=Cr_{,k}, \tag{26}$$

where

$$A=\frac{1}{2\pi}\left[-\sum_{m=1,2}g_m\frac{K_1(z_m)}{z_m}+g_3\left(K_0(z_3)+\frac{K_1(z_3)}{z_3}\right)\right],$$

$$B=\frac{1}{2\pi}\left[-\sum_{m=1,2}g_mK_2(z_m)+g_3K_2(z_3)\right],$$

$$C=\frac{j}{2\pi}\sum_{m=1,2}\frac{r_m'g_m}{k_m}K_1(z_m),$$

and

$$r_{,i}=\frac{y_i-x_i}{r},\quad z_i=jk_ir.$$

These terms form the two first columns of the tensor u_{ik}^*. The tensor t_{ik}^* may be derived from this solution and is expressed as

$$t_{ik}^S=\lambda\left(A'-B'-\frac{B}{r}\right)r_{,k}n_i+\mu\left(\left(A'-\frac{B}{r}\right)(r_{,j}n_j\delta_{ik}+r_{,i}n_k)-\frac{2B}{r}r_{,k}n_i+2\left(-B'+\frac{2B}{r}\right)r_{,i}r_{,k}r_{,j}n_j\right) \tag{27}$$

and

$$q_k=-\kappa\left(\frac{C}{r}n_k+\left(C'-\frac{C}{r}\right)r_{,k}r_{,i}n_i\right) \tag{28}$$

The sign $'$ indicates an r derivative.

2. Fluid load

This case is less complex than the solid one. Only a simple Lamé potential decomposition has to be performed in order to decouple the system. The fluid load is defined here by $\mathbf{f}^S=\mathbf{0}$ and $f^p=-\delta(\mathbf{y}-\mathbf{x})$. Equations (1) and (2), expressed with the Lamé potentials are

$$(\lambda+2\mu)\Delta\varphi+\omega^2\tilde{\rho}_e\varphi+\tilde{\gamma}p=0, \tag{29}$$

$$\mu\Delta\boldsymbol{\psi}+\omega^2\tilde{\rho}_e\boldsymbol{\psi}=0, \tag{30}$$

$$\kappa\Delta p+\kappa k^2 p-\tilde{\gamma}\Delta\varphi=-\delta(\mathbf{y}-\mathbf{x}), \tag{31}$$

where

$$\mathbf{u}^S=\nabla\varphi+\nabla\wedge\boldsymbol{\psi}. \tag{32}$$

Observing that the shear wave equation is not loaded and is uncoupled from the other equations, with $\boldsymbol{\psi}=\mathbf{0}$, thus only p and φ have to be considered. The same method as for the solid load for decoupling the equations is used and leads to

$$\Delta\begin{Bmatrix}\varphi\\ p\end{Bmatrix}+\tilde{M}\cdot\begin{Bmatrix}\varphi\\ p\end{Bmatrix}=\mathbf{h}\begin{Bmatrix}0\\ -\delta\end{Bmatrix}, \tag{33}$$

where

$$\begin{Bmatrix}\varphi\\ p\end{Bmatrix}=\alpha_1\begin{Bmatrix}1\\ r_1'\end{Bmatrix}+\alpha_2\begin{Bmatrix}1\\ r_2'\end{Bmatrix} \tag{34}$$

and

$$\mathbf{h}=h_1\begin{Bmatrix}1\\ r_1'\end{Bmatrix}+h_2\begin{Bmatrix}1\\ r_2'\end{Bmatrix}.$$

The elementary decoupled bi-dimensional equations thus become

$$\alpha_i''+\frac{2}{r}\alpha_i'+k_i^2\alpha_i=-h_i\delta(\mathbf{y}-\mathbf{x}),$$

where the solution is

$$\alpha_i(r)=\frac{h_i}{2\pi}K_0(jk_ir). \tag{35}$$

Applying (35) to (36), leads to Green's solution for the p unknown,

$$p=\frac{1}{2\pi}\sum_{m=1,2}r_m'h_mK_0(z_m), \tag{36}$$

and the displacement $\mathbf{u}^S=\nabla\alpha_1+\nabla\alpha_2$ is

$$u_i^S=Ar_{,i}, \tag{37}$$

where

$$A=\frac{-j}{2\pi}\sum_{m=1,2}k_mh_mK_1(z_m).$$

These Green's functions form the last column of the tensor $u_i^{k*}(\mathbf{y}-\mathbf{x})$.

Stresses and flux are then

$$t_i^S=\left[(\lambda+2\mu)\frac{A}{r}+\lambda A'\right]n_i+2\mu\left(-\frac{A}{r}+A'\right)r_{,i}r_{,j}n_j \tag{38}$$

and

$$q=-\frac{j\kappa r_{,i}n_i}{2\pi}\sum_{m=1,2}r_m'h_mK_1(z_m). \tag{39}$$

F. Multi-region assembly

The boundary discretization described above leads to a much smaller system than that of the whole body discretization as in the finite element method. On the other hand, the system matrices generated by the boundary element method are fully populated. To reduce overall computer costs, a multi-region assembly approach has been introduced.

In a constant element scheme, a basic approach is possible. Assume two domains having a common boundary S^c as depicted in Fig. 1.

On this boundary, the continuity hypotheses are

$$\begin{aligned}&U_1^c=U_2^c,\\ &T_1^c+T_2^c=0.\end{aligned} \tag{40}$$

The system of equations (12) can be transformed as follows:

$$[H]^{-1}[G]\{U\}=\{T\}. \tag{41}$$

With the introduction of $[A]=[H]^{-1}[G]$ and the reordering,

$$\begin{bmatrix}A_{11}&A_{1c}\\ A_{c1}&A_{cc}\end{bmatrix}\begin{Bmatrix}U_1\\ U_c\end{Bmatrix}=\begin{Bmatrix}T_1\\ T_c\end{Bmatrix},$$

$$\begin{bmatrix}A_{cc}'&A_{2c}\\ A_{c2}&A_{22}\end{bmatrix}\begin{Bmatrix}U_c\\ U_2\end{Bmatrix}=\begin{Bmatrix}T_c\\ T_2\end{Bmatrix}.$$

Using the conditions of continuity (40), assembling generates the global system:

$$\begin{bmatrix}A_{11}&A_{1c}&0\\ A_{c1}&A_{cc}+A_{cc}'&A_{c2}\\ 0&A_{2c}&A_{22}\end{bmatrix}\begin{Bmatrix}U_1\\ U_c\\ U_2\end{Bmatrix}=\begin{Bmatrix}T_1\\ 0\\ T_2\end{Bmatrix}. \tag{42}$$

The calculus of the matrices $[A]$ has to be performed very carefully. An acceptable numerical evaluation of this matrix is obtained by a direct algorithm without any explicit inversion of $[H]$.

G. Coupling with an elasto-dynamic domain

The main aim of the present study is to construct models of multi-layer insulation panels including thick layers of porous lining and structural plates. Most of the cases of interest are double-walled sets.

So, it is necessary to couple the porous boundary elements to elasto-dynamic parts. Just as in the former section,

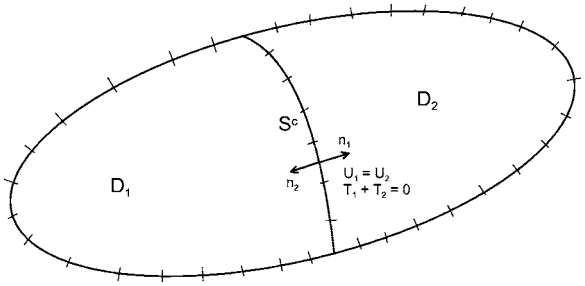


FIG. 1. Geometry of a two-subregion problem.

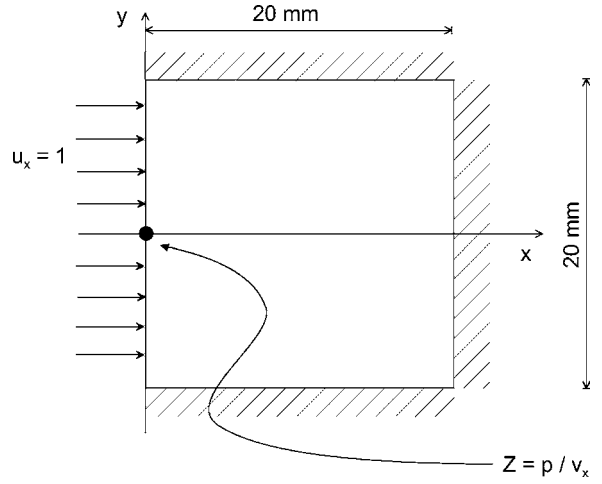


FIG. 2. Impedance test case.

two domains are connected to each other by a common border line. But here, one is porous and the other has an elastic behavior. The elastodynamic domain is assumed to be discretized with a classical finite element method. The corresponding numerical system is

$$\begin{bmatrix} K_{xx} & K_{xy} & \cdots \\ K_{yx} & K_{yy} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{Bmatrix} u_x \\ u_y \\ \vdots \end{Bmatrix} = \begin{Bmatrix} F_x \\ F_y \\ \vdots \end{Bmatrix}. \tag{43}$$

Only degrees of freedom over the boundary are here introduced, the others are in the dots $\cdots$.

The porous system can be written

$$-[G]\begin{Bmatrix} u_x^S \\ u_y^S \\ p \\ \vdots \end{Bmatrix} + [H]\begin{Bmatrix} t_x' \\ t_y' \\ q' \\ \vdots \end{Bmatrix} = \{0\}. \tag{44}$$

The continuity relations between porous (number 1) and solid (number 2) media are

(i) impermeability where $(u_i^{F1} - u_i^{S1})n_i = 0$ u^F is the displacement of the (fluid phase),
(ii) solid displacement continuity $u_i^{S1} = u_i^{S2}$, and
(iii) total stress continuity $\sigma_{ij}^{t1} n_j = \sigma_{ij}^{t2} n_j$.

These coupling conditions have to be represented in term of t_i' and q'. The total load $t_i^t = \sigma_{ij}^t n_j$ is related to $\hat{t}_i (= \hat{\sigma}_{ij}^S n_j)$ and t_i' by

$$t_i^t = \hat{t}_i - p n_i, \quad t_i' = \hat{t}_i + \tilde{\gamma} p n_i$$

and

$$t_i^t = t_i' - (1 + \tilde{\gamma}) n_i p. \tag{45}$$

The modified flux $q' = q - \tilde{\gamma} u_n^S$ can be linked to displacements. The following definitions are reintroduced:

$$q = \kappa p_{,i} n_i$$

and

$$\phi_P p_{,i} = \omega^2 \tilde{\rho}_{22} u_i^F + \omega^2 \tilde{\rho}_{12} u_i^S.$$

This last equation is the fluid equation for the Biot model when displacement terminology is used.

Recalling the definition of γ

$$\tilde{\gamma} = \phi_P \left(\frac{\tilde{\rho}_{12}}{\tilde{\rho}_{22}} - \frac{Q}{R} \right),$$

and setting $Q/R = (1 - \phi_P)/\phi_P$, a very simple relation appears:

$$q' = (\phi_P u_i^F + (1 - \phi_P) u_i^S) n_i. \tag{46}$$

Note that the hypotheses in Sec. II A for the expression of Q and R lead also to simplify the coupling conditions in the original $u-p$ formalism.

With the impermeability relation, the modified flux q' is equal to the normal solid displacement:

$$q' = u_i^S n_i. \tag{47}$$

The system (44) here becomes

$$-[G]\begin{Bmatrix} u_x^S \\ u_y^S \\ p \\ \vdots \end{Bmatrix} + [H]\begin{Bmatrix} t_x^t \\ t_y^t \\ 0 \\ \vdots \end{Bmatrix} + [H][C]\begin{Bmatrix} u_x^S \\ u_y^S \\ p \\ \vdots \end{Bmatrix} = \{0\}, \tag{48}$$

where

$$[C] = \begin{bmatrix} 0 & 0 & (1 + \tilde{\gamma}) n_x & 0 \\ 0 & 0 & (1 + \tilde{\gamma}) n_y & 0 \\ n_x & n_y & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The third term depends on the primary unknowns u^S and p and is directly linked to the properties of the coupling between the two phases of the porous and the solid.

To assemble the systems (43) and (48), the same approach as for porous to porous assembly procedure is used and the boundary element system is rewritten as

TABLE I. Material parameters.

Material	ϕ_P	σ_R (Nm^{-4} s)	α_∞	Λ (m)	Λ' (m)	ρ (kg m^{-3})	E	ν	η
Foam A	0.93	55 000	3.2	30×10^{-6}	320×10^{-6}	30	504 KPa	0.4	0.1
Foam B	0.97	17 000	1.02	240×10^{-6}	490×10^{-6}	10	505 KPa	0.3	0.07
Plate	…	…	…	…	…	1514	58 MPa	0.3	0.01

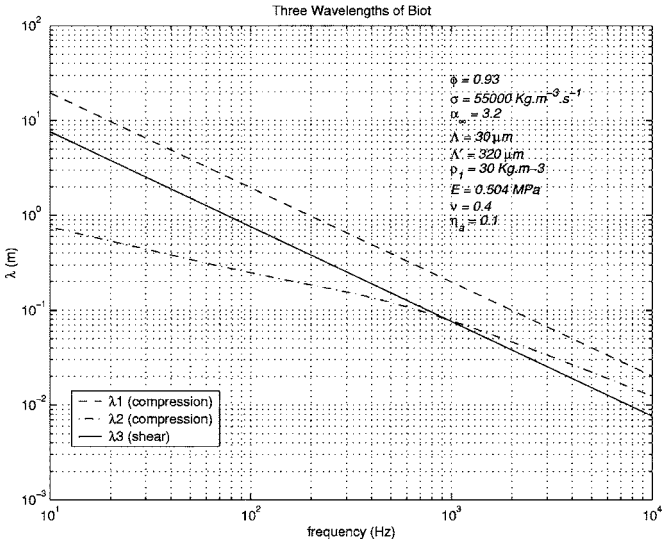


FIG. 3. Three wavelengths of Biot for the foam A.

$$\begin{bmatrix} K'_{xx} & K'_{xy} & K'_{xp} & \cdots \\ K'_{yx} & K'_{yy} & K'_{yp} & \cdots \\ K'_{px} & K'_{py} & K'_{pp} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{Bmatrix} u_x^S \\ u_y^S \\ p \\ \vdots \end{Bmatrix} = \begin{Bmatrix} F_x^t \\ F_y^t \\ 0 \\ \vdots \end{Bmatrix}, \tag{49}$$

where

$$[K'] = [S]([H]^{-1}[G] - [C]).$$

The matrix $[S]$ is a diagonal matrix serving to scale the boundary system with finite elements because the right side members are forces for finite elements and stress loads for the original boundary element system. With constant elements, $F_i^t = \int_{S_f} t_i^t \, dS = S_f t_i^t$. The continuity in stress is then satisfied by $F_i^t + F_i = 0$. Thus the assembly procedure leads to

$$\begin{bmatrix} K'_{xx} + K_{xx} & K'_{xy} + K_{xy} & K'_{xp} & \cdots \\ K'_{yx} + K_{yx} & K'_{yy} + K_{yy} & K'_{yp} & \cdots \\ K'_{px} & K'_{py} & K'_{pp} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{Bmatrix} u_x^S \\ u_y^S \\ p \\ \vdots \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ \vdots \end{Bmatrix}. \tag{50}$$

If the nodes of the two meshes were not coincident, projection operators should have been introduced. In the present work, a direct assembly procedure was chosen. We have, however, an approximation due to the shape functions chosen over each domain: piecewise linear or quadratic for the conventional structural finite elements and piecewise constant for the boundary elements.

III. EXAMPLES OF APPLICATION

A. Surface impedance calculation

The first test case involves a single porous domain analysis. A square porous sample is glued (see Fig. 2) in a rigid cavity of the same size. One side of the cavity is open to a classical air domain. The acoustical velocity amplitude is assumed to be constant and equal to unity on this boundary.

Introducing u^a and p^a, respectively the acoustical displacement and pressure, the continuity relations between porous and fluid materials are

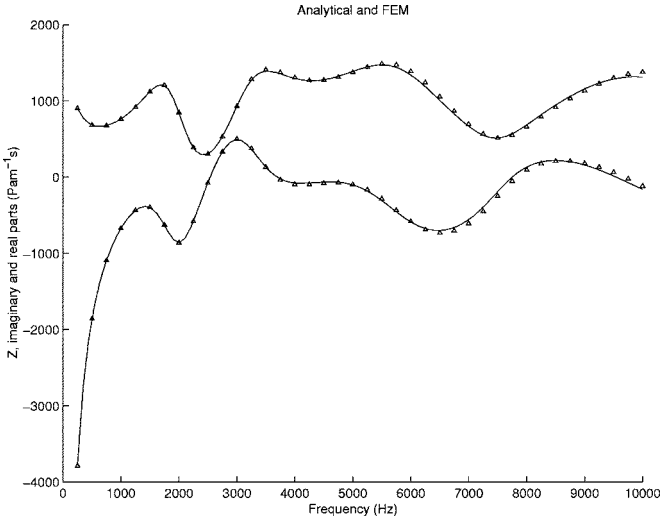


FIG. 4. Comparisons, infinite layer case: analytic (—) and FEM (△) (mesh 20×1).

(i) total stress continuity: $t_i^t = \sigma_{ij}^t n_i = -p_a n_i$,
(ii) pressure continuity: $p = p_a$, and
(iii) flux continuity: $\phi_P u_n^F + (1 - \phi_P) u_n^S = u_n^a = (1/\rho_o \omega^2) \times (\partial p^a / \partial n)$.

If u^a has to be imposed, using (46), we obtain $q' = u_n^a$. Nether $\hat{t}'$ nor u_n^S is directly known but a relation between p and $\hat{t}'$ exists and has to be introduced. Using $t_i^t = -pn_i$ and Eq. (45) leads to $\hat{t}_i' = \tilde{\gamma} p n_i$. The system (44) is thus modified as follows:

$$-[G]\begin{Bmatrix} u_x^S \\ u_y^S \\ p \\ \vdots \end{Bmatrix} + [H]\begin{Bmatrix} 0 \\ 0 \\ u_n^a \\ \vdots \end{Bmatrix} + [H][C]\begin{Bmatrix} u_x^S \\ u_y^S \\ p \\ \vdots \end{Bmatrix} = \{0\}, \tag{51}$$

where

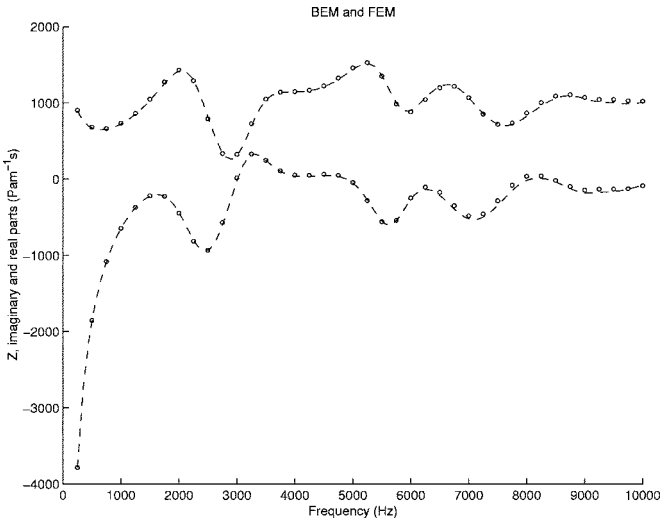


FIG. 5. Comparisons, clamped sample case: BEM (···) (mesh 21×21) and FEM (○) (mesh 20×20).

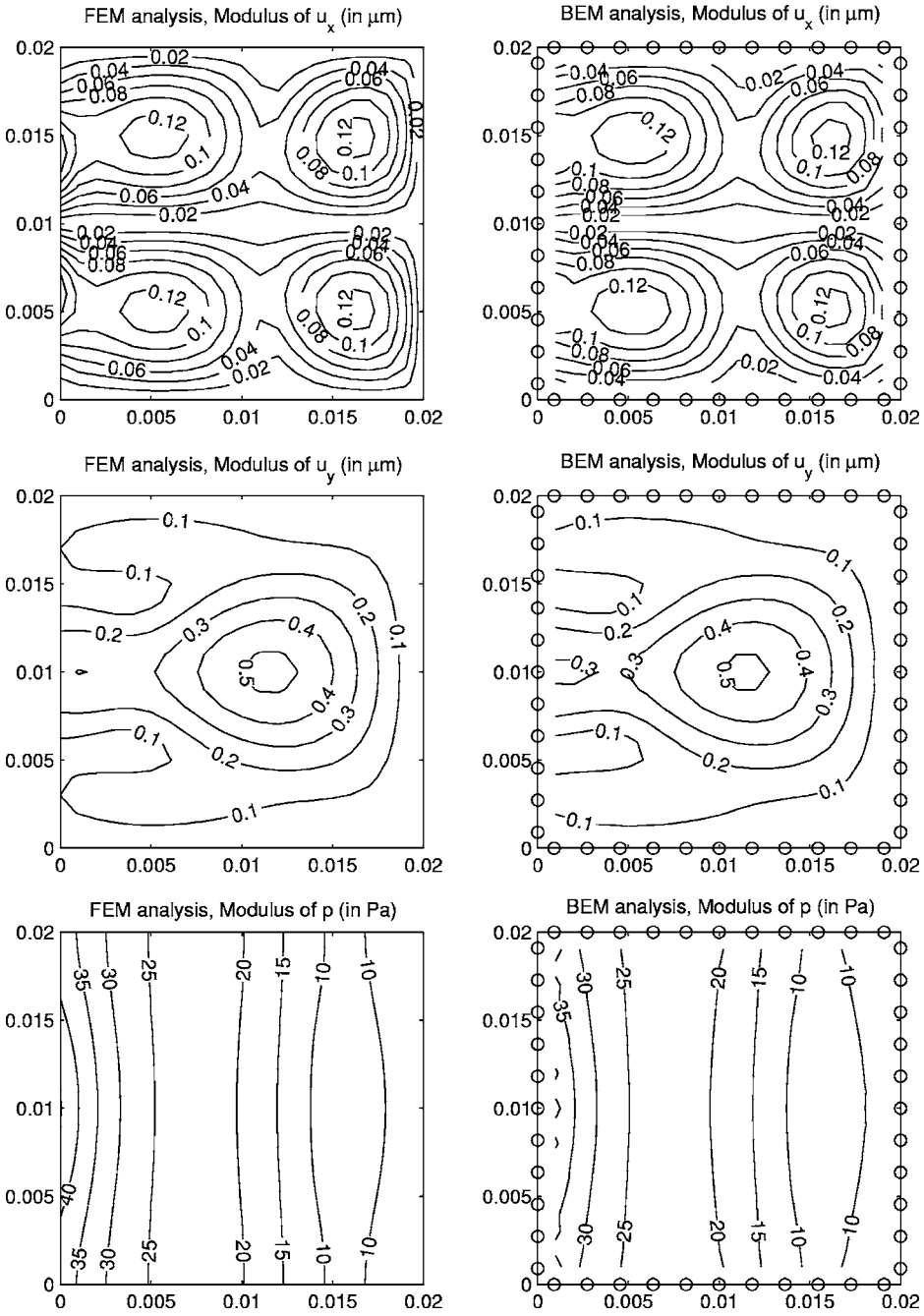


FIG. 6. Comparisons of BEM and FEM solution fields at 5000 Hz. Left column: finite element solution (mesh 20×20 elements), right column: boundary element solution (mesh 11×11 elements); positions of collocation points are plotted for the BEM situation.

$$[C]=\begin{bmatrix} 0 & 0 & \tilde{\gamma}n_x & 0 \\ 0 & 0 & \tilde{\gamma}n_y & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The impedance $Z(\omega)=p^a(\omega)/v^a(\omega)$ is computed at the center of the boundary and plotted.

This case is relatively close to the impedance tube test. Using lateral clamped and rigid boundaries allows us to appreciate the influence of this kind of mounting but overall creates a bi-dimensional internal behavior in the porous sample.

The characteristics of the porous material are taken from Ref. 13, given in Table I (foam A), and the three Biot wavelengths are plotted on Fig. 3. This reference provides results for the perfect situation involving an infinite porous layer of uniform thickness and glued to a rigid wall. This analytical solution is also reproduced in the present study (Fig. 4).

The aim of this test case is to validate the porous boundary method presented above. A comparison with other numerical methods such as the finite element method seems to be particularly appropriate here. First, the finite element method applied to porous media has been validated in previous studies[14–16] either by experimental/simulation or in an analytical/numeric perspective. Moreover, an entirely numerical study allows much more complex situations (multidimensional) than the one-dimensional case computable analytically. Finally, the finite element and boundary element approaches are theoretically equivalent.

A bidimensional finite element tool including porous el-

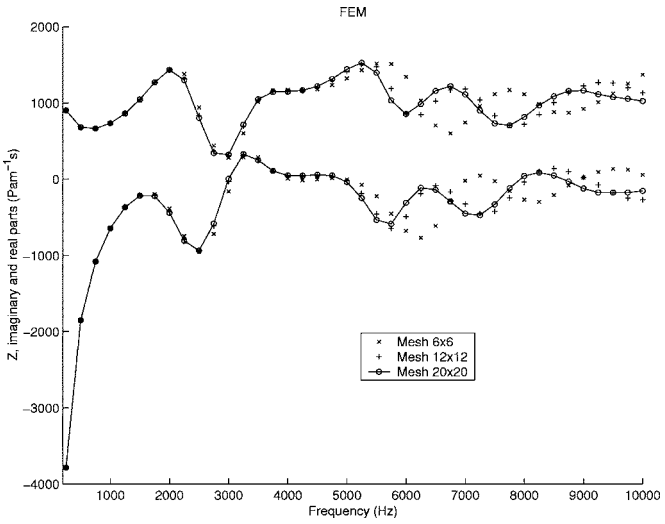


FIG. 7. Solutions for unconverged meshes: finite element method.

ements has been developed in Ref. 17. The basis of this reference is a $u-p$ formalism.[3] Starting from the equilibrium equations (1) and (2), the weak variational form is obtained as the sum of two integrals:

$$\int_S (\delta\varepsilon_{ij}^S \hat{\sigma}_{ij}^S - \omega^2 \tilde{\rho}_e \delta u_i^S u_i^S - \tilde{\gamma} \delta u_i^S p_{,i}) dS = \int_L \delta u_i^S \hat{\sigma}_{ij}^S n_j dL, \tag{52}$$

$$\int_S (\kappa \delta p_{,i} p_{,i} - \kappa k^2 \delta p p - \tilde{\gamma} \delta p_{,i} u_i^S) dS = \int_L \kappa \delta p (p_{,i} + \tilde{\beta} u_i^S) n_i dL, \tag{53}$$

where

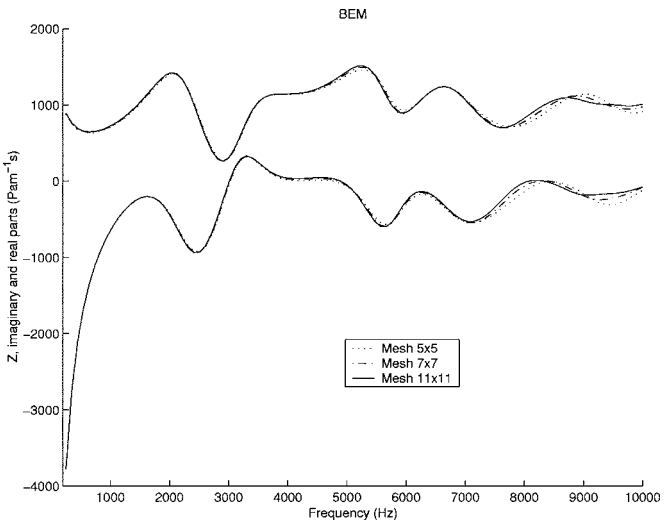


FIG. 8. Solutions for unconverged meshes: boundary element method.

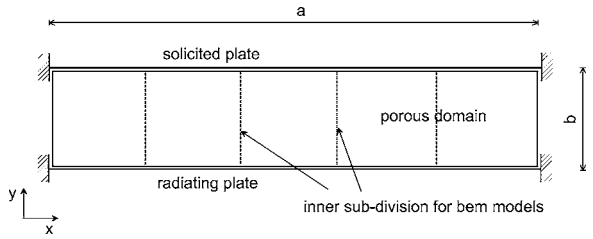


FIG. 9. Transmission test case.

$$\tilde{\beta} = -\frac{\omega^2 \tilde{\rho}_{22}}{\phi_P^2} \tilde{\gamma}. \tag{54}$$

A standard Lagrangian Q4 reference element based on this formulation is introduced. The boundary condition is easy to apply and concerns only the second member. Thus,

$$\int_L \delta u_i^S \hat{\sigma}_{ij}^S n_j dL = 0$$

and

$$\int_L \kappa \delta p (p_{,i} + \tilde{\beta} u_i^S) n_i dL = \int_L \delta p q' dL = \int_L \delta p u_n^a dL.$$

In Figs. 4 and 5, two comparisons are drawn. The first shows that the finite element in a one-dimensional situation (all u_x components are blocked, u_y are free except along the right-hand edge side) gives the same result as the theoretical solution. This preliminary result confirms the validity of the finite element tool.

As no interior clamping is readily possible in the boundary element method the preceding 1D case cannot be treated. The second concerns the present boundary element method and illustrates that both numerical methods lead to the same solution in a relatively complex situation involving a maximum of terms in the equations.

In Fig. 6, various solution fields are plotted for the frequency of 5000 Hz. The fields plotted for the boundary element approach are obtained through a postprocessing using

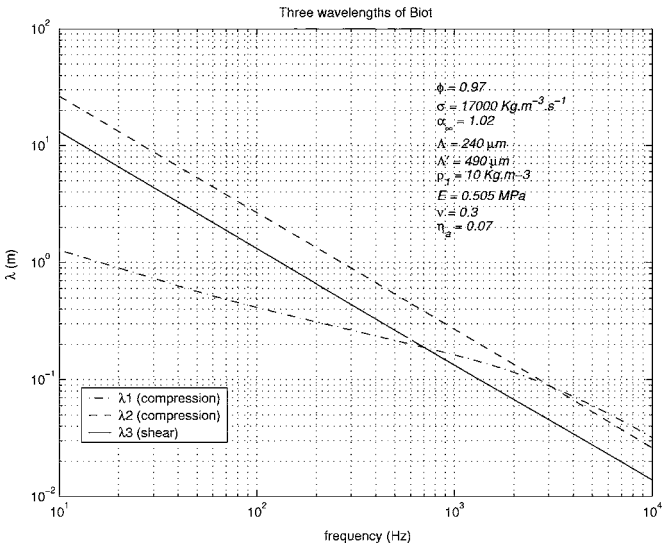


FIG. 10. Biot wavelengths of the foam B.

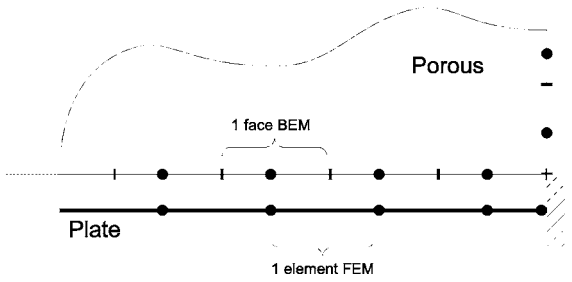


FIG. 11. Principle of meshing for assembly of porous (BEM) and structural (FEM) parts.

Eq. (8) on an interior triangular mesh. This figures illustrates the preceding remarks and validates the assertion that the solution is the same over the whole domain.

Here, all numerical results were presented with converged meshes. In Figs. 7 and 8, unconverged results are plotted. This study shows that the boundary element method presents a relatively good convergence rate as regards the mesh size, although the present implementation only uses a constant discretization scheme. Indeed, the boundary elements mesh of only 5×5 elements here gives a better approximation of the solution than the 12×12 finite element mesh and is very close to the solutions obtained with 20×20 finite element or 11×11 boundary element meshes.

This observation is important because the boundary element method could present an improvement *vis-a-vis* the finite element approach only if the convergence rate is such that a same level of approximation can be obtained at lower numerical cost.

Here the boundary element mesh 11×11 generates a numerical system matrix of size $N=132$ including $N^2=17\,424$ nonzero terms. The 20×20 linear finite element mesh leads to a sparse matrix of size 2646 with 133 956 nonzeros. Linear equation solvers work well with matrices having high sparsity patterns and the numerical cost (in memory or CPU usage) depends on the size but also on the density of the system. However, in the present case the boundary element method maintains an important advantage with respect to the finite element approach. The square geometry helps, of course, but this assertion should be verified for more complex geometry having a good V/S ratio (for each BEM subdivision of the whole geometry more exactly).

B. Transmission loss of a sandwich panel

This case study is a more complex example where multidomain and multi-physical capabilities are explored. Transmission loss (TL) is a measure of the acoustical insulation property of a panel. The formal definition is

$$\mathrm{TL}=10\log\frac{W_i}{W_t}, \tag{55}$$

where W_i and W_t are respectively the power loading on one side and the radiated power from the other side of the tested panel.

The configuration studied here is a sandwich panel made of two thin composite plates (thickness=0.7 mm) and a foam core (see Fig. 9). The material characteristics are given in Table I (foam B) and the wavelengths are plotted in Fig. 10. The global width b and thickness a can take different values. The plates are clamped on their bounds. We use linear two-node Kirschoff elements. The continuity equations discussed in Sec. II G are applied between both exterior plates and the foam. The nodes of the different meshes are coincident (see Fig. 11). On the lateral boundaries, the u_x^S degrees of freedom of the foam are set to zero.

Both finite element and boundary element approaches are used and compared. The same procedure determining the TL is respected: first the load is a diffuse-field approximated by a sum of 32 plane waves of varying incidence from −78 to +77 deg by increments of 5 deg. The radiated power is evaluated by a method similar to classical experimental procedure using an intensity probe. This power is obtained by integration over a grid parallel to the panel of the normal intensity vector (as drawn in Fig. 12):

$$W_t=\int_{\mathrm{grid}}\mathrm{Re}(pv_n^*)dL, \tag{56}$$

where the acoustical normal velocity and pressure are simply $p=(p_A+p_B)/2$ and $v_n=-(p_B-p_A)/j\omega\rho_o d$. p_A and p_B are computed by the use of the Rayleigh integral:

$$p(M)=\frac{j\omega^2\rho_o}{2}\int_L u_n(P)H_o^2(kr)dL, \tag{57}$$

where r is the distance between point P on the radiating plate and point $M=A$ or $=B$ on the measure grid. u_n is the normal displacement of the plate and H_o^2 is the Hankel function of second kind and order zero.

In Fig. 13, TL curves are plotted for a thickness of 10 mm and widths of 150, 300, and 600 mm. The porous boundary element subdomains are square blocks of 5×5 elements. The frequency range covers a band from 100 to 5000 Hz with 60 points having a logarithmic pattern.

The infinite case is also plotted by the use of a classical transfer matrix method.[18,19] The curves tend logically to this limit, showing here the transition from a bidimensional to a one-dimensional behavior. The width of the plates influences their dynamic behavior. The 150 mm case yields a low-frequency response where the first mode (220 Hz approxi-

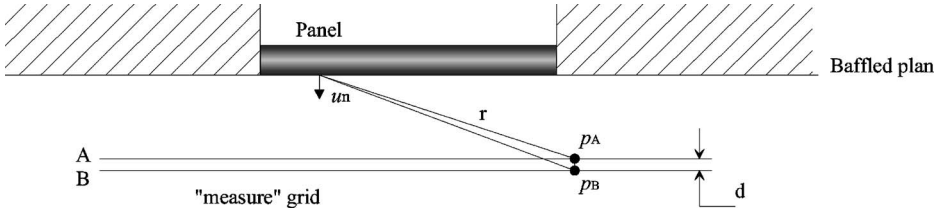


FIG. 12. Radiated power estimation by intensimetry.

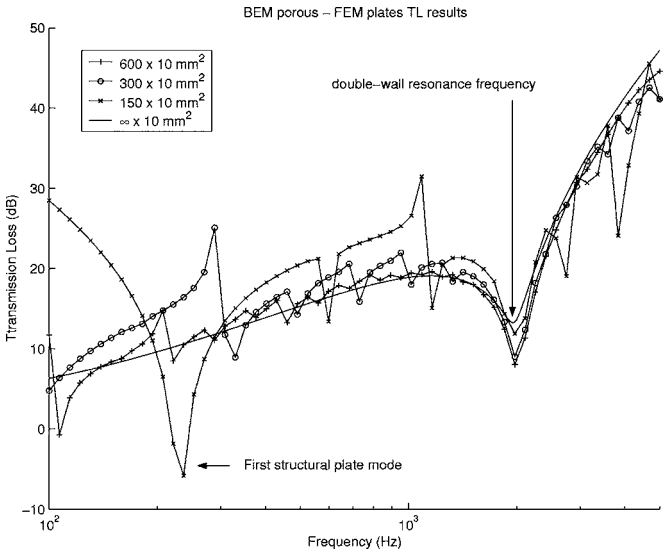


FIG. 13. TL of double-wall panel for different sizes using the boundary element method coupled to structural finite elements and a transfer matrix method for the semi-infinite case.

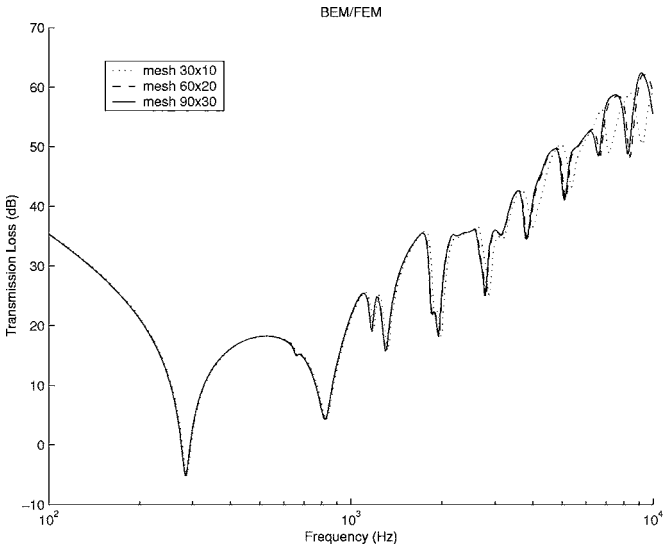


FIG. 15. TL of the 150×50 mm^2 panel for different meshes (identified by the number of elements in x and y direction): mixed BEM/FEM method.

mately) is included in the frequency range of interest. The stiffness controlled zone is clearly visible below. The 600 mm is, on the contrary, a high-frequency situation where structural modes are very close together. These modes have limited influence and the TL curve is globally very close to that of the infinite panel. Only the double-wall mode ("breathing mode") does not move from its position. This resonance frequency (2000 Hz approximately) depends only on the surface mass of the plates and on the thickness of the foam.

A much thicker double-wall is studied in Figs. 14 and 15. The width is 150 mm and thickness is 50 mm. Complete finite element and mixed boundary and finite element analysis are compared for several meshes. For each method three meshes of elements of the same size in x and y directions are tested. Concerning the boundary element method, the porous layer is divided into six subdomains.

Note that at 10 kHz, the shortest wave (the shear wave) has a length of 14 mm and the shortest longitudinal wave has a length of 26 mm. A number of six elements per wavelength, which is generally a good starting point for mesh dimensioning (but sometimes insufficient, especially for poro-elasticity[15]) results in 22 and 11 elements respectively in the thickness direction of this panel.

Both methods give similar results for meshes of the same size. The intermediate mesh (60×20 elements) is acceptable for both methods.

The unconverged mesh (30×10 elements) introduces the same error: an overestimation of modal frequencies at high frequency. Such is typical of structural finite elements in the case of insufficient mesh. With a square element mesh, the plates' finite elements determine the convergence rate.

Curves obtained with the most refined meshes (90×30) are plotted in Fig. 16 and show good agreement.

To appreciate the convergence of porous lining discreti-

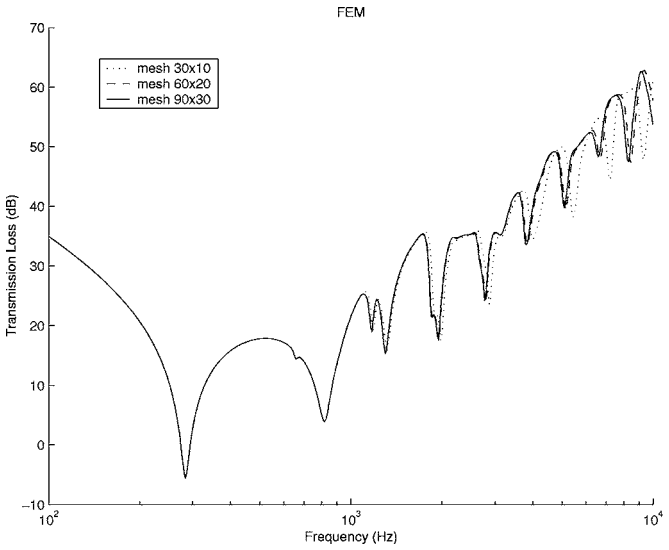


FIG. 14. TL of the 150×50 mm^2 panel for different meshes (identified by the number of elements in x and y direction): pure finite element analysis.

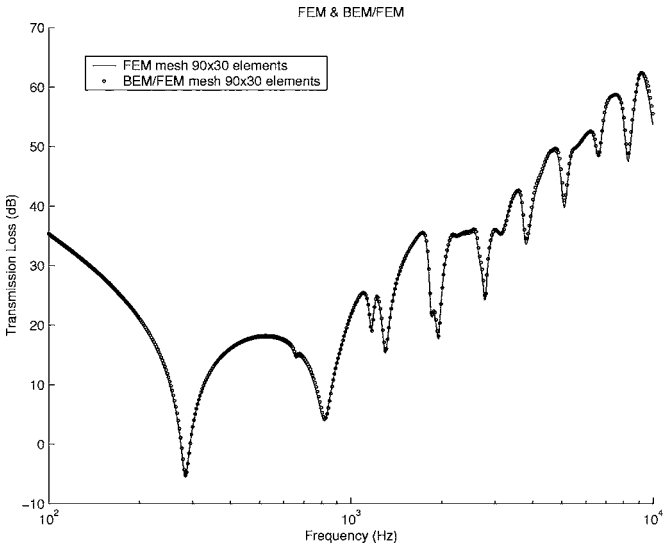


FIG. 16. TL of the 150×50 mm^2 panel for both methods.

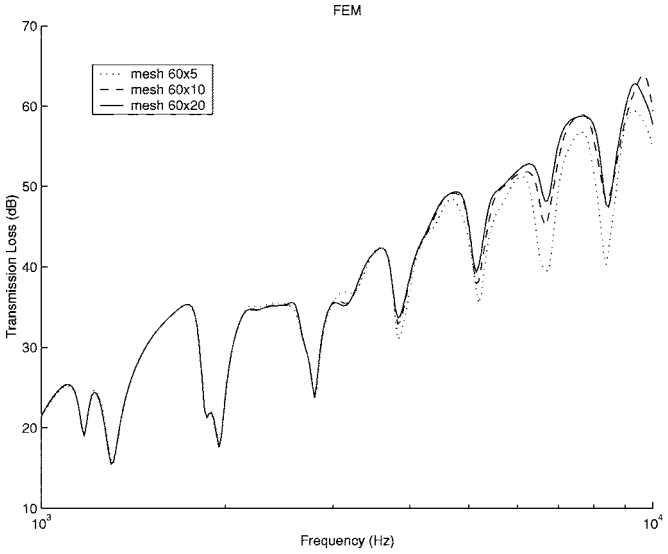


FIG. 17. TL of 150×50 panel, high-frequency curve for mesh of different size in the thickness direction: finite element method.

zation, Figs. 17 and 18 yield the evolution of the solutions at high frequencies for meshes where the number of elements is set to 60 in the y direction. The boundary element method allows here an important reduction in the number of elements in the x direction for equivalent solution. Only five porous boundary elements in the thickness leads to results similar to those involving ten finite elements. This observation conforms to the conclusions of the first study case involving a single porous sample.

Retaining these latter equivalent models, the size of the numerical assembly system can be introduced: the mesh 60×5 in six subdomains leads to a matrix of size 597 (number of degrees of freedom) containing 45 037 nonzero values. The complete finite element mesh 60×10 generates a matrix of size 2135 and containing 55 205 nonzero values. The new approach entails a major reduction in the number of degrees of freedom but also a non-negligible reduction in the number of matrix terms. A gain thus obtained will correspond to increased speed and lower memory use during calculation for most of the available numerical solvers. As a remark, the block structure of the BEM/FEM matrix gave a LU factorization (using an appropriate renumbering pretreatment) conserving the sparsity: only an amplification of factor 1.9 between the number of nonzeros of the initial matrix and the L and U matrices has been observed.

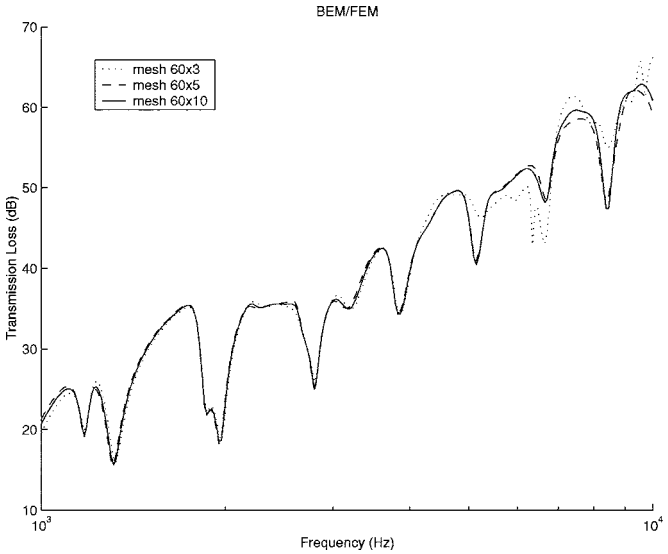


FIG. 18. TL of 150×50 panel, high-frequency curve for mesh of different size in the thickness direction: mixed boundary and finite element method.

The present approach could certainly be improved by an incompatible coupling technique between porous domains and structures.

IV. CONCLUSIONS

A new method has been proposed and tested over a wide range of vibro-acoustic problems including porous materials. The solutions obtained were verified by comparison to more conventional approaches.

The method leads to a converged solution with fewer numerical resources than a simple linear finite element approach in spite of the use of a constant discretization scheme over the boundary elements.

The introduction to the method of analytical Green's solutions explains this relatively good behavior of the boundary element approach. Improvements could be added to this present work in the domains of higher order boundary elements, variational statement for the boundary integrals, and coupling techniques.

We did a straightforward implementation (sufficient for testing and for the seeking of simplicity) but including the necessary stages for most vibro-acoustic calculation strategies: a discretization procedure based upon an integral $u-p$ formulation of the Biot model, a capability of subdomain partition, and coupling with a solid part. Of course, each of these stages needs to integrate better numerical solutions for further deployment.

Then we will be able to solve more complex problems. The bi-dimensional approximation gives interesting perspectives for an easy modeling of nonhomogeneous materials mixing various porous regions, holes or solid inclusions, taking advantages of the multi-domain strategy and the very simple meshing steps required.

The extension of this method to 3D cases should be considered with a very similar theoretical formulation. Beyond material studies with relatively small samples, the vibro-acoustical analysis of trimmed structures could benefit from such an approach. FEM voluminous porous parts should be replaced by BEM subregions with probably a gain in numerical cost but also an easiest mesh generation (for complex geometries such as automotive dash-board, helicopter roof,…).

ACKNOWLEDGMENTS

We wish to acknowledge the support of Paulstra Industries and the State of São Paulo Research Foundation (FAPESP).

[1]M. Biot, "The theory of propagation of elastic waves in a fluid satured porous solid," J. Acoust. Soc. Am. **28**, 168–191 (1956).
[2]J. Allard, *Propagation of Sound in Porous Media* (Elsevier Applied Sci-

ence, Essex, UK, 1993).
[3]N. Atalla, R. Panneton, and P. Debergue, "A mixed displacement-pressure formulation for poroelastic materials," J. Acoust. Soc. Am. **104**, 1444–1452 (1998).
[4]Y. Kang and J. Bolton, "A finite element model for sound transmission through foam-lined double panel structures," J. Acoust. Soc. Am. **99**, 2755–2765 (1996).
[5]J. Coyette, "The use of finite element and boundary element models for predicting the vibro-acoustic behavior of layered structures," Adv. Eng. Software **30**, 133–139 (1999).
[6]N.-E. Hörlin, "3d hieracrchical hp-fem applied to elasto-acoustic modelling of layered porous media," J. Sound Vib. **285**, 341–363 (2004).
[7]P. Banerjee, *The Boundary Element Methods in Engineering*, 2nd ed. (McGraw-Hill, London, 1994).
[8]C. A. Brebbia, *The Boundary Element Method for Engineers*, 2nd ed. (Pentech, London, 1980).
[9]M. C. Cavalcanti and J. C. F. Telles, "Biot's consolidation theory—application of bem with time independent fundamental solutions for poroelastic saturated media," Eng. Anal. Boundary Elem. **27**, 145–157 (2003).
[10]K. H. Park and P. K. Banerjee, "Two- and three-dimensional soil consolidation by bem via particular integral," Comput. Methods Appl. Mech. Eng. **191**, 3233–3255 (2002).
[11]C. Boutin, "Dynamique des milieux poreux saturés déformables-fonctions de green-perméamtre dynamique," Ph.D. thesis, Un. scientifique, technologique et médicale de Grenoble, Vaux-en-Velin, 1987.
[12]M. Biot and D. Willis, "The elastic coefficients of the theory of consolidation," J. Appl. Mech. **24**, 594–601 (1957).
[13]J. F. Allard, C. Depollier, and W. Lauriks, "Measurement and prediction of surface impedance at oblique incidence of a plastic foam of high flow resistivity," J. Sound Vib. **132**(1), 51–60 (1989).
[14]R. Panneton and N. Atalla, "An efficient finite element scheme for solving the three-dimensional poroelasticity problem in acoustics," J. Acoust. Soc. Am. **101**(6), 3284–3298 (1997).
[15]N. Dauchez and S. Sahraoui, "Convergence of poroelastic finite elements based on biot displacement formulation," J. Acoust. Soc. Am. **109**(1), 33–40 (2001).
[16]F. Sgard, N. Atalla, and J. Nicolas, "A numerical model for the low frequency diffuse field sound transmission loss of double-wall sound barriers with elastic porous linings," J. Acoust. Soc. Am. **108**(6), 2865–2872 (2000).
[17]O. Tanneau, "Modélisation de panneaux aéronautiques," Ph.D. thesis, Université Pierre et Marie Curie, Paris, 2004.
[18]J. Bolton, N.-M. Shiau, and Y. Kang, "Sound transmission through multipanel structures lined with elastic porous materials," J. Sound Vib. **191**(3), 312–347 (1996).
[19]C. Germain, "Optimisation vibroacoustique de panneaux intérieurs de portes automobiles," Ph.D. thesis, Un. de Valenciennes, 1999.

Series expansion for the sound field of rotating sources

Michael Carley[a)]
Department of Mechanical Engineering, University of Bath, Bath BA2 7AY, England

(Received 10 March 2006; revised 19 May 2006; accepted 6 June 2006)

A method is derived for the fast, exact prediction of acoustic fields around rotating sources by using a series expansion which generalizes a previously published method for a circular piston. The technique gives exact predictions for the field outside the sphere containing the rotor in a computational time two orders of magnitude less than that required for direct numerical evaluation of the acoustic integrals. Its use is demonstrated by application to two sample problems characteristic of real aircraft propellers. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2221410]

PACS number(s): 43.20.Rz, 43.20.Tb [SFW] Pages: 1252–1256

I. INTRODUCTION

This paper presents an efficient, exact method for the prediction of noise from rotating sources, such as aircraft propellers. The problem to be considered is the evaluation of the integral for the sound radiated by a sinusoidally varying source distribution on a disk. A rotor of radius a has a geometry which can be decomposed into Fourier components in azimuth θ. For a given azimuthal order n the source on the disk is $s(r_1,\theta_1)=\epsilon s_n(r_1)\exp jn\theta_1$, where the coordinate system is that shown in Fig. 1, the subscript 1 refers to a variable of integration and $\epsilon \ll 1$.

We begin with the special case of $s_n \equiv 1$. In this case, the radiated acoustic field can be written as a Rayleigh integral[1]

$$p(\mathbf{x},\omega) = -\epsilon M_t^2 n^2 e^{jn\theta} I_n(a,r,z,\omega), \tag{1}$$

$$I_n(a,r,z,\omega) = \int_0^a \int_0^{2\pi} \frac{e^{j(kR'+n\theta_1)}}{4\pi R'} r_1 d\theta_1 dr_1,$$

$$R' = [r^2 + r_1^2 - 2rr_1 \cos\theta_1 + z^2]^{1/2}, \tag{2}$$

where ω is the source frequency, c is the speed of sound, and $k=\omega/c$. For a source which rotates at angular velocity Ω, $\omega=n\Omega$ and $M_t=\Omega a/c=ka/n$ is the tip rotational Mach number.

A number of methods have been devised for the evaluation of I_n. In particular, previous work has recognized its relationship to the standard Rayleigh integral with $n\equiv 0$. Oberhettinger's numerical approach for transient radiation,[2] which reduces I_0 to a one-dimensional integral, has been extended to the case of $n\neq 0$ (Refs. 1 and 3–5) and used to study the physical structure of rotating sound fields. In this paper, a similar approach is used and the exact series expansion of Mast and Yu[6] is applied to the problem of rotating sources to yield a very efficient, exact, general method of computing the noise from a rotor. Mast and Yu's approach to the problem of radiation from a piston is the most recent in a series of papers which use series expansions to evaluate the field. The earliest such work in English appears to be that of Spence,[7] although Mast and Yu cite two papers in German, those of Backhaus (1930) and Stenzel (1935). Further developments of the series expansion approach have included Carter and Williams' expansion,[8] the work of Wittman and Yaghjian[9] who introduced methods from electromagnetism and, in particular, the results of Hasegawa, Inoue and Matsuzawa[10] which formed the basis of Mast and Yu's recent work.

II. THEORY

In a recent paper,[6] a series expansion was developed for the calculation of harmonic radiation from a circular piston. The method of this paper is now used to develop a similar expansion for a disk-shaped source with azimuthal variation in source strength. The result presented, Eq. (6), gives the acoustic field around a rotor at distances greater than the rotor radius. It has not, so far, proven possible to derive an equivalent expansion for the region within the sphere enclosing the source.

A. Exact series

The expansion for a rotor noise field can be derived from the integral of Eq. (2) using the summation theorem for Bessel functions[11]

$$\frac{e^{jkR'}}{R'} = \frac{j\pi}{2(\rho S)^{1/2}} \sum_{m=0}^{\infty} (2m+1) J_{m+1/2}(k\rho) H_{m+1/2}^{(1)}(kS) \times P_m(\cos\theta_1 \sin\beta), \tag{3}$$

where

$$R' = [\rho^2 + S^2 - 2S\rho\cos\theta_1 \sin\beta]^{1/2},$$

$$r^2 + r_1^2 + z^2 = \rho^2 + S^2, \quad \sin\beta = r_1 r/\rho S,$$

J_ν is the Bessel function of the first kind and order ν, $H_\nu^{(1)}$ is the Hankel function of order ν, and P_m is the Legendre polynomial of order m.

In order to perform the integration over θ_1, we require the addition theorem for Legendre polynomials[11]

a)Electronic mail: m.j.carley@bath.ac.uk

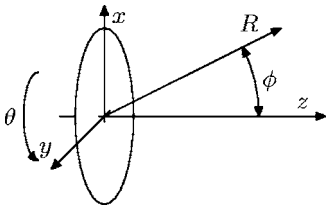


FIG. 1. Coordinate system for the disk source.

$$P_m(\cos\theta_1 \sin\beta) = P_m(\cos\beta)P_m(0) + 2\sum_{q=1}^{\infty} \frac{\Gamma(m-q+1)}{\Gamma(m+q+1)} P_m^q(\cos\beta) \times P_m^q(0)\cos q\theta_1,$$

where P_m^q is the associated Legendre function which is zero for $q>m$. Inserting this into

$$\int_0^{2\pi} P_m(\cos\theta_1 \sin\beta) e^{jn\theta_1} d\theta_1,$$

gives, for $n\neq 0$:

$$\int_0^{2\pi} P_m(\cos\theta_1 \sin\beta) e^{jn\theta_1} d\theta_1 = 2\pi \frac{\Gamma(m-n+1)}{\Gamma(m+n+1)} P_m^n(\cos\beta) P_m^n(0), \tag{4}$$

which, upon insertion into Eqs. (2) and (3), gives

$$I_n = j\pi^{3/2} 2^{n-2} \sum_{m=0}^{\infty} \frac{(4m+2n+1)\Gamma(2m+1)}{\Gamma(2m+2n+1)\Gamma(m+1)\Gamma(1/2-m)} \times \int_0^a \frac{1}{(\rho S)^{1/2}} J_{n+2m+1/2}(k\rho) H_{n+2m+1/2}^{(1)}(kS) \times P_{n+2m}^n(\cos\beta) r_1 dr_1, \tag{5}$$

where use has been made of the fact that $P_{n+2m+1}^n(0)\equiv 0$.

A series expansion can now be derived by making a suitable choice of coordinates. The choice of Mast and Yu,[6] $\rho=r_1$, $S=R=(r^2+z^2)^{1/2}$ with $\sin\beta=\sin\phi$ gives an expansion valid for $R>a$. With these coordinates, the integral can be evaluated[12] and

$$I_n = (\pi a^2)^{1/2} \frac{(-ka)^{n+1/2}}{(kR)^{1/2}} \sum_{m=0}^{\infty} A_m H_{n+2m+1/2}^{(1)}(kR) \times \mathrm{unP}_{n+2m}^n(\cos\phi) \times {}_1F_2\left[\frac{n+2m+2}{2}; \frac{n+2m+4}{2}, n+2m+3/2; -\left(\frac{ka}{2}\right)^2\right] \left(\frac{ka}{2}\right)^{2m+1/2},$$

$$A_m = (-1)^m \frac{2^{2m-1}}{n+2m+2} \frac{(2m-1)!!}{(2n+2m)!!(2n+4m-1)!!}, \tag{6}$$

where use has been made of the functional relations for the gamma function[11] and ${}_1F_2(\cdot)$ is a generalized hypergeometric function[13]

$${}_1F_2(a;b,c;x) = \sum_{n=0}^{\infty} B_n x^n,$$

$$B_n = \frac{(a)_n}{(b)_n (c)_n} \frac{1}{n!}, \tag{7}$$

where $(a)_n=\Gamma(a+n)/\Gamma(a)$ is Pochammer's symbol.[11]

The series expansion in Eq. (6) is the main result of this paper. It gives an exact expansion for the acoustic field around a rotating source in terms of three special functions. It is worth noting that the special functions are "uncoupled:" for given k and n, if the polar angle ϕ is fixed, only the Hankel functions need be computed more than once per term so that the field can be calculated very quickly for a large number of points on a given radius.

B. General source distributions

The expansion of Eq. (6) can be extended to cover the general case where $s_n(r_1)$ varies with radius. The integral is rewritten

$$K_n = \int_0^a s_n(r_1) \int_0^{2\pi} \frac{e^{j(kR+n\theta_1)}}{4\pi R} r_1 d\theta_1 dr_1, \tag{8}$$

which upon integration by parts becomes

$$K_n = s_n(a) I_n(a,r,z,\omega) - \int_0^a \frac{ds_n}{dr_1} I_n(r_1,r,z,\omega) dr_1. \tag{9}$$

On the assumption that $s_n(a)=0$, i.e., that the source vanishes at a rotor blade tip, Eq. (9) can be integrated termwise to give

$$K_n = -\pi^{1/2} \frac{(-k)^{n+1/2}}{(kR)^{1/2}} \sum_{m=0}^{\infty} A_m H_{n+2m+1/2}^{(1)}(kR) P_{n+2m}^n(\cos\phi) \times \left(\frac{k}{2}\right)^{2m+1/2} \times \int_0^a \frac{ds_n}{dr_1} r_1^{n+2m+2} {}_1F_2\left[-\left(\frac{kr_1}{2}\right)^2\right] dr_1, \tag{10}$$

where the parameters of the hypergeometric function have been dropped for brevity.

In computing the acoustic field of a real source, the integrals of Eq. (10) would have to be calculated numerically, although this would only have to be done once, since the integrals depend only on the source and are the same no matter what the observer position. In order to find an analytical formula, we use a source distribution which has been applied in asymptotic studies[14] $s_n=(a-r_1)^{\nu}$. Then the integral is readily evaluated

$$\int_0^a {}_1F_2[-(\alpha x)^2](a-x)^{\nu} x^n dx = a^{n+\nu+1} \sum_{q=0}^{\infty} B_q \frac{(2q+n)!}{(\nu+1)_{2q+n+1}} \times (-\alpha a)^{2q}, \tag{11}$$

and the acoustic field integral is

$$K_n = j\pi^{1/2}\frac{(-k)^n}{R^{1/2}}\sum_{m=0}^{\infty} A_m H^{(1)}_{n+2m+1/2}(kR)P^n_{n+2m}(\cos\phi) \times\left(\frac{k}{2}\right)^{2m+1/2} \times \int_0^a \nu(a-r_1)^{\nu-1} r_1^{n+2m+2} \times {}_1F_2\left[-\left(\frac{kr_1}{2}\right)^2\right] dr_1. \quad (12)$$

III. RESULTS

Results are presented to demonstrate the efficiency and accuracy of the series expansion. For comparison, I_n and K_n are also computed by direct two-dimensional integration using Gaussian quadrature. The number of terms in the quadrature was adjusted to the minimum which gave a converged result. In evaluating the expansion, the Hankel functions were computed using the finite series[11]

$$H^{(1)}_{n-1/2}(kR) = j^{-n}\left(\frac{2}{\pi kR}\right)^{1/2} e^{jkR} \times\sum_{q=0}^{n-1}(-1)^q\frac{(n+q-1)!}{q!(n-q-1)!}\frac{1}{(j2kr)^q}, \quad (13)$$

the associated Legendre functions were computed using the implementation in the GNU Scientific Library[15] and the generalized hypergeometric function and its integral were evaluated by direct summation to machine precision (tolerance 10^{-15}). In summing the series, the convergence criterion was that the magnitude of the last term added be less than 10^{-6}. All calculations were performed on a GNU/Linux personal computer using code written in C using the GNU C compiler and the GNU Scientific Library.[15]

A. Numerical performance

The first test was a check on the accuracy and efficiency of the series expansion compared to direct integration. Figure 2 compares the real part of K_n computed using the two methods for $n=32$, $M_t=1.0$, $ka=32$ at a polar angle $\phi=\pi/2$, parameters characteristic of a high speed propeller of diameter 2.74 m rotating at 1200 rpm, similar to those used in a

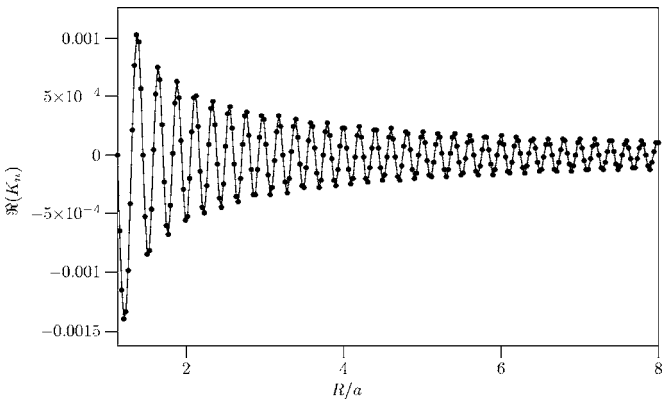


FIG. 2. Real part of acoustic integral I_n for $n=32$, $\nu=1/2$, and $\phi=\pi/2$; by direct integration and series evaluation: series solution solid; direct integration dots.

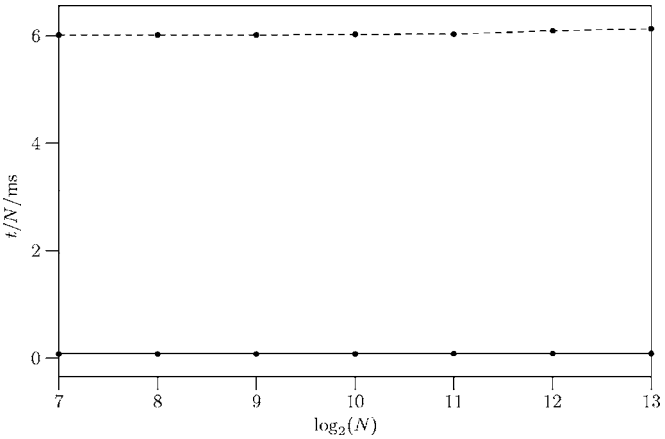


FIG. 3. Computational time per point t/N vs number of points N: solid line: series expansion; dashed line: direct integration.

NASA study on high speed propellers.[16] As a check on the integration scheme, a variation in the radial source distribution was introduced by setting $\nu=1/2$. The direct integration was performed using 128 quadrature points in radius and azimuth. The series expansion converged after nine terms and, as can be seen, its accuracy is very good. To examine the computational effort, the calculation was repeated for a piston (i.e., $n=0$) with the same value of ka and the same convergence criterion. This required 39 terms for convergence, more than four times as many as in the rotor case.

Figure 3 shows the computational effort for the calculation as computational time per field point against $\log_2(N)$ where N is the number of field points. Over a wide range of N, the computation time per point for the series expansion is two orders of magnitude less than that required for direct integration, even though the result is no less accurate. The mean time per point for the series evaluation was 0.08 ms while that for direct integration was 6.0 ms.

B. Acoustic fields

Sample results are presented for the acoustic field around a rotating source, taking parameters representative of

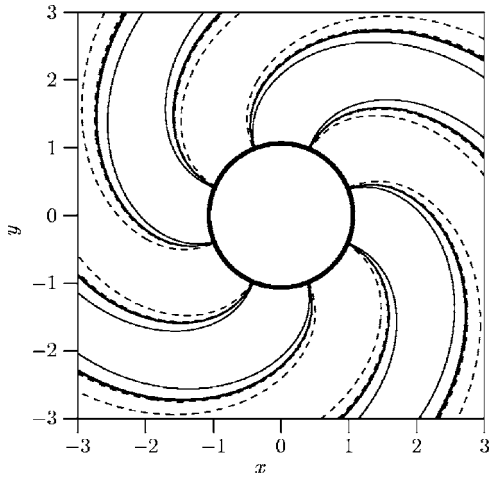


FIG. 4. Real part of acoustic pressure in the plane $z=0$ with $M_t=0.7$, $n=4$, and $\nu=1/4$. Contour levels $\pm10^{-3}$, 10^{-4}, 10^{-5}, positive levels solid, negative levels dashed.

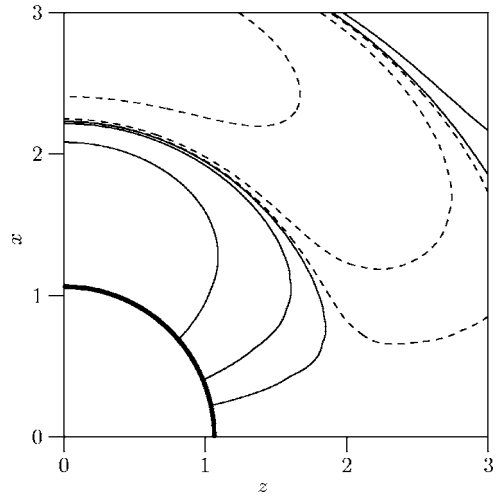


FIG. 5. Real part of acoustic pressure in the plane $y=0$ with $M_t=0.7$, $n=4$, and $\nu=1/4$. Contour levels $\pm 10^{-3}$, 10^{-4}, 10^{-5}, positive levels solid, negative levels dashed.

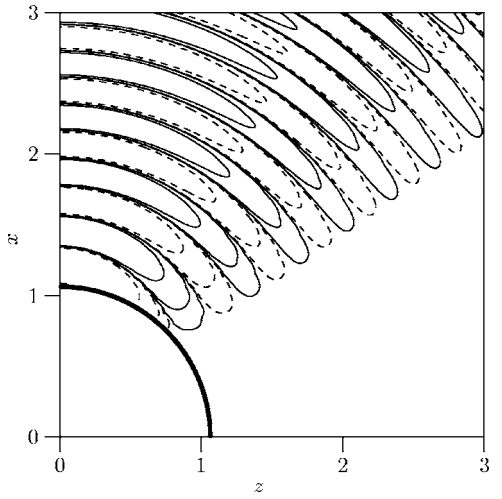


FIG. 7. Real part of acoustic pressure in the plane $y=0$ with $M_t=1.125$, $n=16$, and $\nu=1/4$. Contour levels $\pm 10^{-3}$, 10^{-4}, 10^{-5}, positive levels solid, negative levels dashed.

conventional rotors and of high speed propellers. The structure of rotating sound fields has been presented in detail in previous work[1,3–5] and so the examples shown here serve to illustrate application of the method.

The first results are calculated for $M_t=0.7$, $n=4$, $ka=2.8$, and $\nu=1/4$, equivalent to a 2.74 m propeller rotating at 830 rpm. Figure 4 shows the acoustic field $[\Re(K_n)]$ in the plane $z=0$ while Fig. 5 shows equivalent results for the plane $y=0$. A maximum of 28 terms were required in evaluating the series. Data were computed outside the sphere $R \geqslant 1.0625a$ with $a=1$, the inner radius being shown as a heavy line in each plot. The field has the structure described in previous work,[1] being composed of segments like those of an orange with the field decaying exponentially over the "peel" around the sonic radius $1/M_t=1.43$. Since the rotor lies completely inside the sonic radius, it radiates only weakly into the far field.

Figures 6 and 7 show similar results for a high speed rotor (diameter 2.74 m rotating at 1334 rpm) with $M_t=1.125$, $n=16$, $ka=18$, and $\nu=1/4$. A maximum of 23 terms was needed in evaluating the series. In this case, part of the source lies outside the sonic radius $1/M_t=0.89$ and it can radiate strongly into the far field, without losing energy in the transition to the radiation zone. This strong radiation shows up as the long "swirls" spiraling out of the source disk in Fig. 6 and as the slow decay of the field on radial lines in Fig. 7. This plot also shows the sharp demarcation between the "quiet zone" of a supersonic rotor, near the z axis, and the "loud zone" where field points are subject to a source approaching at a Mach number of unity.

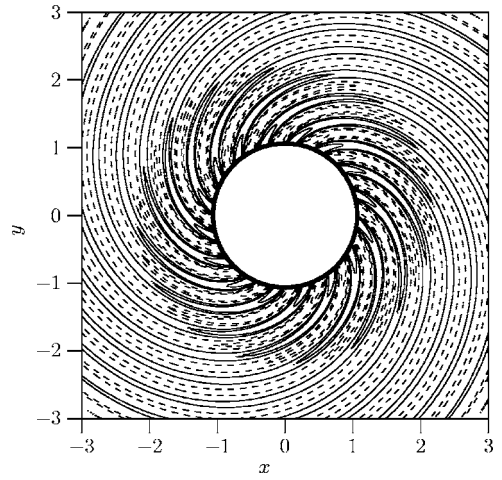


FIG. 6. Real part of acoustic pressure in the plane $z=0$ with $M_t=1.125$, $n=16$, and $\nu=1/4$. Contour levels $\pm 10^{-2}$, 5×10^{-3}, 2.5×10^{-3}, positive levels solid, negative levels dashed.

IV. CONCLUSIONS

A fast, exact method for the prediction of acoustic fields around rotating sources has been developed as a generalization of a technique for the prediction of time-harmonic fields around circular pistons.[6] The method uses an exact series expansion valid for general radial source distributions which gives a converged solution in a time two orders of magnitude smaller than that required for evaluation by direct two-dimensional integration. The accuracy and efficiency of the method have been demonstrated by computing the field around sample sources characteristic of actually existing propellers.

[1]C. J. Chapman, "The structure of rotating sound fields," Proc. R. Soc. London, Ser. A **440**, 257–271 (1993).
[2]F. Oberhettinger, "On transient solutions of the 'baffled piston' problem," J. Res. Natl. Bur. Stand., Sect. B **65**(1), 1–6 (1961).
[3]M. Carley, "Sound radiation from propellers in forward flight," J. Sound Vib. **225** (2), 353–374 (1999).
[4]M. Carley, "Propeller noise fields," J. Sound Vib. **233** (2), 255–277 (2000).
[5]M. Carley, "The structure of wobbling sound fields," J. Sound Vib. **244** (1), 1–19 (2001).
[6]T. D. Mast and F. Yu, "Simplified expansions for radiation from a baffled circular piston," J. Acoust. Soc. Am. **118** (6), 3457–3464 (2005).
[7]R. D. Spence, "The diffraction of sound by circular disks and apertures," J. Acoust. Soc. Am. **20** (1), 380–386 (1948).
[8]A. H. Carter and J. A. O. Williams, "A new expansion for the velocity potential of a piston source," J. Acoust. Soc. Am. **23** (2), 179–184 (1951).
[9]R. C. Wittman and A. D. Yaghjian, "Spherical-wave expansions of piston-radiator fields," J. Acoust. Soc. Am. **90** (3), 1647–1655 (1991).
[10]T. Hasegawa, N. Inoue, and K. Matsuzawa, "A new rigorous expansion for

the velocity potential of a circular source piston," J. Acoust. Soc. Am. **74** (3), 1044–1047 (1983).
[11]I. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 5th ed. (Academic, London, 1980).
[12]A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev, *Integrals and Series* (Gordon and Breach, New York, 1986), Vol. **2**.
[13]L. J. Slater, *Generalized Hypergeometric Functions* (Cambridge University Press, Cambridge, 1966).
[14]A. B. Parry and D. G. Crighton, "Asymptotic theory of propeller noise part I: Subsonic single rotation propeller," AIAA J. **27** (9), 1184–1190 (1989).
[15]M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth, and F. Rossi, *GNU Scientific Library Reference Manual* (Network Theory, Bristol, 2005).
[16]F. Farassat, M. H. Dunn, and P. L. Spence, "Advanced propeller noise prediction in the time domain," AIAA J. **30** (9), 2337–2340 (1992).

Air entrapment in piezo-driven inkjet printheads

Jos de Jong and Gerrit de Bruin
Physics of Fluids Group, Faculty of Science and Technology and Burgers Center of Fluid Dynamics, University of Twente, Enschede, P.O. Box 217, 7500 AE Enschede, The Netherlands

Hans Reinten, Marc van den Berg, and Herman Wijshoff
Océ Technologies B.V., P.O. Box 101, 5900 MA Venlo, The Netherlands

Michel Versluis and Detlef Lohse
Physics of Fluids Group, Faculty of Science and Technology, Burgers Center of Fluid Dynamics, MESA+Institute for Nanotechnology, Institute for Biomedical Technology (BMTI) and Institute of Mechanics, Processes and Control-Twente (IMPACT), University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

(Received 31 May 2005; revised 22 May 2006; accepted 25 May 2006)

The stability of inkjet printers is a major requirement for high-quality-printing. However, in piezo-driven inkjet printheads, air entrapment can lead to malfunctioning of the jet formation. The piezoactuator is employed to actively monitor the channel acoustics and to identify distortions at an early stage. Modifications of the response of the piezoactuator indicate entrapped air bubbles and these allow us to investigate them. When we employ the signal as a trigger for high-speed imaging, we can visualize the consequences of the entrained bubbles on the droplet formation. Two mechanisms are found to cause air entrapment: First, a distorted droplet formation caused by small particles, and, second, an accumulation of ink on the nozzle plate, which favors void formation once the meniscus is pulled back. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2216560]

PACS number(s): 43.25.Yw, 43.38.Fx, 43.35.-c [AJS] Pages: 1257–1265

I. INTRODUCTION

Today's drop-on-demand inkjet printers[1] have as crucial requirements that they are fast, can produce small droplets, and are reliable and robust against distortions. For one full-color A0 print, up to one billion droplets are required. The failure of only a few drops could jeopardize the whole result.

Though in principle inkjet printing with piezoelectric actuation[1] can fulfill all of the above requirements, sometimes, even when the droplet forming process may be very stable for literally millions of droplets, from one to the next actuation cycle there may be an occurrence giving rise to a malfunction of the droplet formation. In the modern versions droplets are jetted every 50 μs. Therefore, for the early detection of anomalies, monitoring of the printing process *while printing* becomes crucial.

In piezoelectric printheads the piezoceramic material is deformed, resulting in pressure buildup at the nozzle region, which is released through drop formation at the nozzle. Fluid acoustics are involved to guide the waveform energy toward the nozzle, and to create pressure and velocity profiles needed for the droplet jetting process.[2] For the discussion of different types of piezoelectric drop-on-demand systems and their principles we refer to Refs. 3–10.

In this paper we will focus on the malfunction of piezoelectric printheads and introduce a method to acoustically detect them at an early stage. It is known that the formation of *air bubbles* during operation can be the origin of the failure of the piezoelectric inkjet system.[11,12] These bubbles alter the acoustic field required for the droplet formation and in many cases cause the droplet formation to stop completely. The bubble only dissolves when the acoustic field is turned off, bringing the total print process to an end. Obviously, first, an early detection of the bubble is desirable, so that, second, some measures can be taken to get rid of the bubble.

In this paper we address the first point and present a measurement technique to acoustically monitor the ink channel and to detect air bubbles therein while printing. This active monitoring is performed by using the piezoelement as a sensor, measuring the pressure in the channel between the droplet-firing pulses by the same piezo. The electrical signal measured from the piezo will be referred to as the acoustical signal in this article. The acoustical signal actively monitors the droplet formation and reacts within microseconds to any change in the acoustical signal. We can therefore use this signal to trigger a high speed camera with which we image the droplet formation process. The high-speed images are correlated with the acoustical signal.

How do the bubbles enter into the ink channel? The origin of the air bubbles can be either nucleation inside the ink channel or air entrapment at the nozzle. As the absence of large pressure amplitudes in the ink channel rules out the first mechanism, we focus on air bubbles entrapped at the nozzle. Once entrapped, the bubble will grow by rectified diffusion[13–15] under the influence of the acoustic actuations, resulting in malfunctioning of the printhead. We will also address the consequences for the functioning of the printhead, in particular, the effect of the entrapped air bubble on the droplet speed. From the high speed imaging of the droplet formation and the air entrapment process, we identify two air entrapment mechanisms:

- a distorted droplet formation caused by small (dust) particles

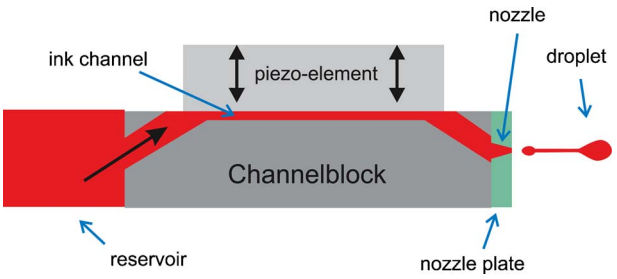


FIG. 1. (Color online) Sketch of the geometry of the printhead. The reservoir is pressure controlled, pushing ink into the rectangular ink channel of the typical length of 10 mm, width of 200 μm, and height of 150 μm. The piezoelement of length 8 mm is covered with a foil of 20 μm thickness, which is in direct contact with the ink. The nozzle plate is nickel-made with round openings of diameter 30 μm, the nozzles themselves. The jetted droplets have a diameter of typically 35 μm.

- creating an ink layer on the nozzle plate, which favors void formation once the meniscus is pulled back and the ink closes the void

The paper is organized as follows: In Sec. II we explain the printhead setup and its operation. In Sec. III we present results on the acoustical bubble detection technique. Section IV is devoted to the high-speed imaging of droplet formation, which is triggered once the acoustical signal gets distorted. In Sec. V we identify the two above mentioned ways how bubbles are entrained in the nozzle. The last section is left to discussions, conclusions, and an outlook.

II. PRINTHEAD SETUP AND OPERATION

The printheads under consideration in this paper are side-shooter printheads developed by Océ and used for professional printing in an industrial environment. The schematic setup of the printhead is depicted in Fig. 1. The channel block is a graphite block with ink channels inside. The ink reservoir is connected to the channel and can be set at a specific ambient pressure. For example, to prevent ink leakage from the nozzles when not jetting, the ambient pressure in the reservoir is lowered, typically by 8 mbar. In the rectangular ink channel, one of the four walls is formed by the piezo. When a voltage pulse is applied to the piezo, the piezo first contracts, increasing the volume of the ink channel. This results in a lower pressure in the channel. In the second part of the pulse, the piezo is expanded again, reducing the volume of the ink channel and thus pushing ink therein, resulting in a buildup of (positive) pressure ("push mode"). The pressure waves generated by the piezo travel toward the ink reservoir, where they are reflected out of phase, and toward the nozzle, where they are reflected in phase. The ink is pressed out through a 30 μm diameter electroformed nickel nozzle.

The printhead driving protocol is sketched in Fig. 2. The applied pulse is a trapezium pulse with a total length of about 15 μs, and the repetition rate of the pulse is 20 kHz (corresponding to 50 μs pulse to pulse distance), which is the firing frequency of the droplets. When the piezo is no longer actuated, it is employed to monitor the pressure waves inside

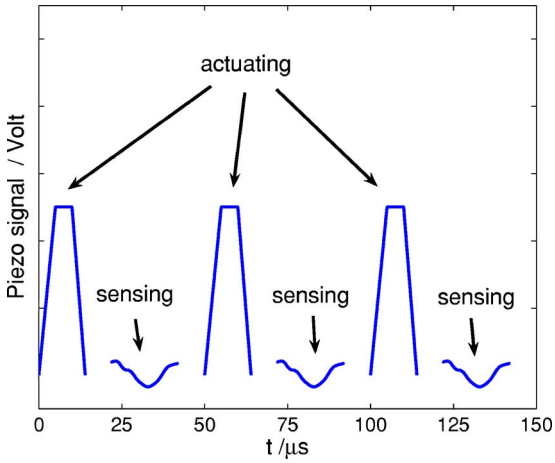


FIG. 2. (Color online) Schematic drawing of the piezo signals. Ink is pressed out of the nozzle through short pressure pulses of 15 μs width, being typically 50 μs apart (i.e., 20 kHz drop printing frequency). In between the pressure pulses the piezoelement is used as a pressure sensor.

the ink channel. The switching from actuating to sensing and vice versa takes about 5 μs and the temporal resolution of the signal is 0.1 μs.

To convey a qualitative idea on the normal operation of the printhead and the droplet formation process, we have modeled the three dimensional ink channel and the nozzle with the CFD program Flow-3D.[16] This CFD code uses a Volume of Fluid technique.[17] The piezo is simulated by a moving obstacle at one side of the ink channel. This compressible model incorporates viscosity and surface tension. It is also considered that the graphite walls of the channel are not stiff, leading to a reduction of the effective sound velocity. Details on the employed numerical model can be found in the articles of Hirt[17] and of Wijshoff.[19] In Fig. 3 the experimental (upper) and numerical (lower) droplet formation processes are compared. They show reasonable qualitative agreement. A small deviation is seen at the breakup of the droplet. This deviation is presumably caused by noise in the actual droplet formation[18] and by the finite mesh size of the model. However, we stress again that no quantitative comparison is aimed at, at this point.

III. ACOUSTICAL DETECTION OF AIR BUBBLES

As pointed out in the previous section and sketched in Fig. 2, the roughly 30 μs long period between pressure pulses is employed to monitor the channel acoustics. This is possible because the acoustic pressure waves present in the ink channel deform the piezo. The force exerted on the piezoelement is converted to an electrical signal that we call the acoustical signal,[24]

$$a(t) = \int_{\text{piezo}} \frac{dF}{dt} dA .$$

Here, $F(t)$ is the force exerted on the piezo, perpendicular to the ink channel. Its temporal change is integrated over the surface A of the piezo, which is in direct contact with the ink. The piezo is nearly as long as the ink channel (8 vs 10 mm total length). Therefore the signal integrates the forces

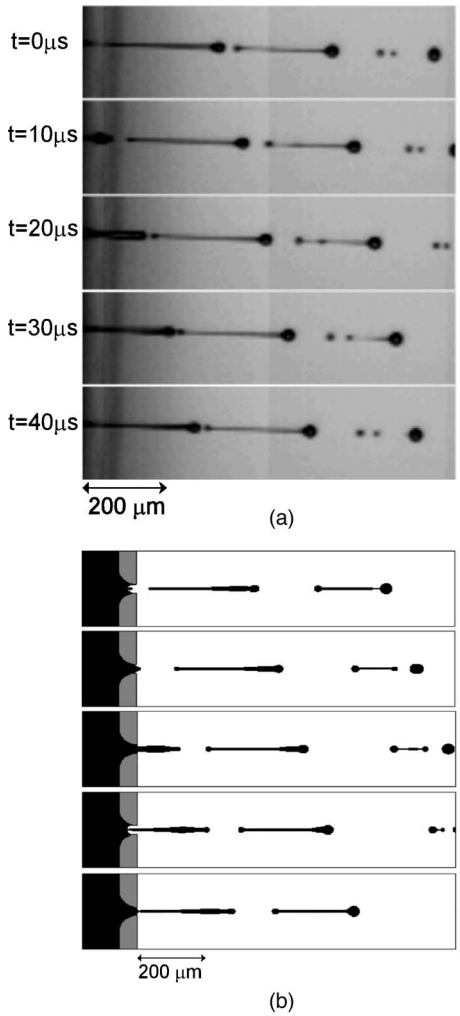


FIG. 3. A comparison of the droplet formation at 20 kHz. (a) Droplet formation, recorded by high speed imaging, (b) Droplet formation, simulated with Flow-3D. The same piezo-deformation as in experiment has been chosen, leading to the pressure profile in the ink channel. In both the experimental and the numerical case the droplet velocity is about 7 m/s.

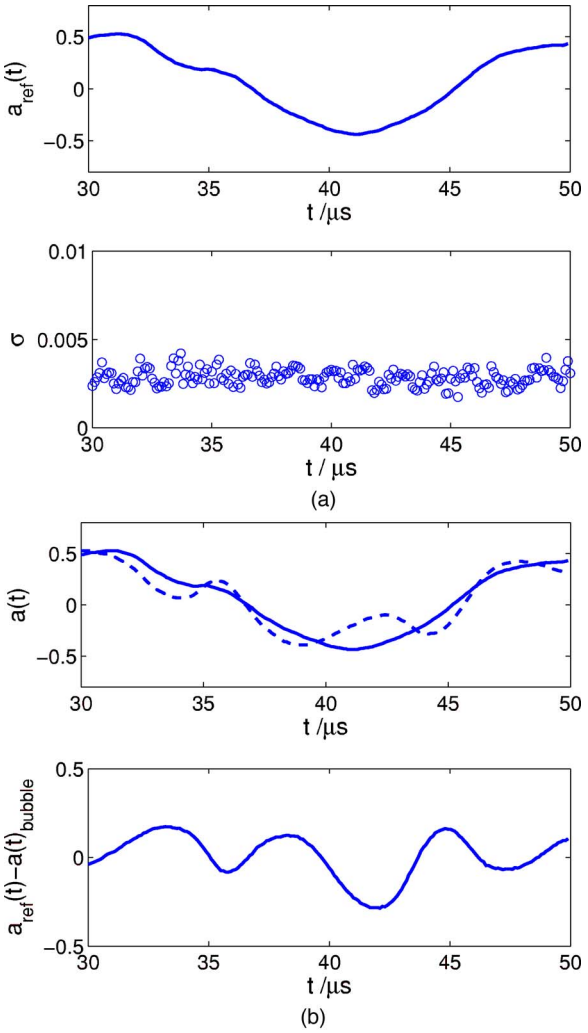


FIG. 4. (a) (Color online) The acoustical signal, $a_{ref}(t)$, averaged over 25 standard droplets (upper), and the corresponding standard deviation that is more than 100 times smaller (lower), showing the reproducibility of the signal. Note the 50 times enlarged scale on the y axis. (b) Upper: Acoustical signals with (dashed) and without (solid) an air bubble present in the ink channel. The lower panel shows the difference of these two signals.

over the whole length and cannot be directly translated back to the acoustic pressure. However, provided that the droplet is fired without distortion, the signal is perfectly reproducible. This is demonstrated in Fig. 4(a). In the top figure the acoustical signals of 25 successively jetted droplets are displayed. The standard deviation for the 25 signals is calculated in the bottom plot of Fig. 4(a). The standard deviation stays well under 0.005 (less than one percent of the maximal acoustical signal), showing the close to perfect reproducibility of the acoustical signals. Therefore this method is potentially very well suited to monitor flow deviations in the channel and at the nozzle.

As an entrained air bubble modifies the acoustical signal in a characteristic way, above acoustical monitoring method should also reveal when an air bubble is present inside the ink channel. In Fig. 4(b) we present such a case: The top figure displays both the standard acoustical signal (solid line) and the acoustical signal with an entrapped air bubble (dashed line). The modification of the signal caused by the entrapped air bubble is shown in the bottom figure. It can clearly be seen that the air bubble adds an additional frequency component to the acoustical signal. Further evidence that the signal modification is due to a bubble is obtained from the following experiment: The acoustic field is turned off for a specific time interval when the modification is present. If the time interval is long enough for the bubble to entirely dissolve, the signal modification disappears. When the bubble is not completely dissolved once the actuation is resumed; the bubble will again grow by rectified diffusion and cause nozzle failure.

Of course, the acoustic sensoring method allows only for an indirect detection of the bubble, not a direct visualization. Such a visualization is not possible in standard ink channels, as they are optically not accessible. However, recently we have succeeded in direct high-speed visualization of entrained air bubbles in channels equipped with nozzle plates partly made out of glass. A correlation between these images and the acoustical signal fully confirms our interpretation that the origin of the modified acoustical signal are entrained air bubbles.[23]

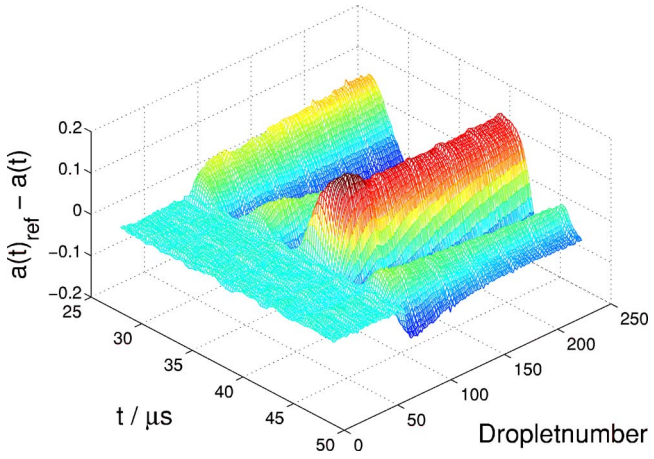


FIG. 5. (Color online) Development of the acoustical signal $a(t)$, as compared to the reference signal $a_{ref}(t)$ without a bubble, after air bubble entrapment.

The development of the acoustical signal after entrapment of an air bubble is shown in Fig. 5. The difference compared to the reference acoustical signal (i.e., the signal without air bubbles) is plotted. This difference occurs rather suddenly, within only a few cycles. This development of the acoustical signal corresponds to the moving and growing of the air bubble. The bubble growth is presumably due to rectified diffusion, which causes a net flow of dissolved air into the bubble during one acoustic cycle. The larger the bubble, the more it influences the acoustic waves traveling in the ink channel. Before a stationary state of the acoustical signals and therefore presumably also the bubble is reached, it can take several thousands of cycles.

To analyze the acoustical signal, the variance

$$\sigma^2 = \frac{1}{T}\int_0^T [a(t) - \bar{a}]^2\, dt,$$

over one period $T=50$ μs of the acoustical signal is calculated and plotted versus the droplet number; see Fig. 6. Remarkably, a large disturbance during two acoustic cycles is occurring just before air entrapment. This disturbance in the variance corresponds to a variation of only 2% in the amplitude of the acoustical signal. It is perfectly reproducibly, i.e., always occurs before air entrapment. On the other hand, not all these disturbances actually lead to air entrapment (as will be seen from Fig. 9, later). As shown by a trigger based on the variance, the disturbances occur randomly during actuations, and, in fact, only rarely result in air entrapment. What is the origin of this disturbance in the acoustical signal, which seems to be a necessary (but not sufficient) condition for air entrapment?

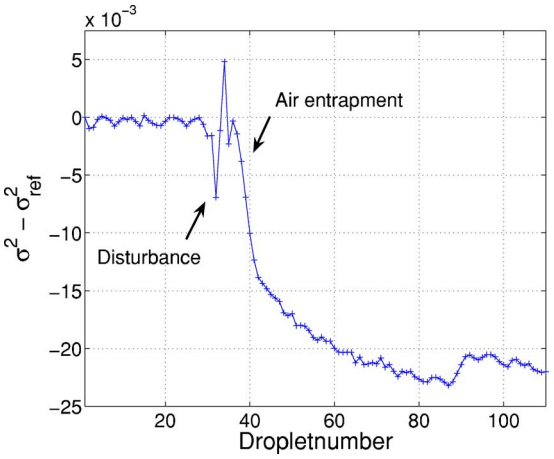


FIG. 6. (Color online) Variance of acoustical signals, when air is entrapped. First the acoustical signals show undisturbed droplets, then a disturbance occurs, and after the disturbance air is entrapped. The variance of the reference signal is subtracted.

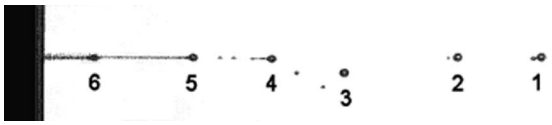


FIG. 7. Distorted droplet formation. Droplets 1 and 2 are standard droplets, droplet 3 corresponds to a disturbance of the acoustical signal (such as, e.g., seen in Fig. 6 for droplets number 34–36 of that figure). Here droplets 4–6 are standard again, but an air bubble is now present in the nozzle. After hundreds of cycles, this can lead to the breakdown of the jetting process.

IV. VISUALIZATION OF DROPLET FORMATION

To answer this question, high speed imaging is employed to visualize the droplets, once the deviation of the variance is detected in the acoustical signal. A Phantom V7 high speed camera from Vision Research is used to monitor the firing of the droplets. The camera is capable of making recordings up to 160 kfps and has pre- and post-triggering. The camera is triggered by the acoustical signal: After a disturbance in the variance has been detected, it records hundreds of droplets before and after the acoustical signal deviation. What can be seen is that the droplet formation is slightly altered before air is entrapped. The droplets differ in speed and sometimes the droplet has a deviant angle. A typical example is shown in Fig. 7.

The first two droplets (1) and (2) are standard droplets. Then a disturbance occurs in droplet (3), which corresponds to the disturbance in the variance shown in Fig. 6. The droplets fired after the disturbance seem to be standard again, but now an air bubble is present inside the nozzle, which can be concluded from the acoustical signal. (The actual air entrapment mechanism will be explained in the next section.) Just after the entrapment the air bubble is presumably very small. Therefore it has a relatively small influence on the droplet formation. After the air bubble is entrapped, it will grow by rectified diffusion and the speed of the droplets is gradually reducing. This is a relatively long-term effect due to the slow time scale of (rectified) diffusion. This can be seen in Fig. 8. The droplet speed reduces by 15% from 6.5 to 5.5 m/s within 20 ms, which corresponds to 400 actuations.

V. ORIGIN OF THE ENTRAINED BUBBLE

A. Air entrapment with regular jetting

As shown in the previous section, it is clear that distorted droplet formation can result in air entrapment. To find out what the actual disturbance is, recordings are made with the Phantom V7 high speed camera. Again, the acoustical signal is used to monitor the droplet formation and to trigger

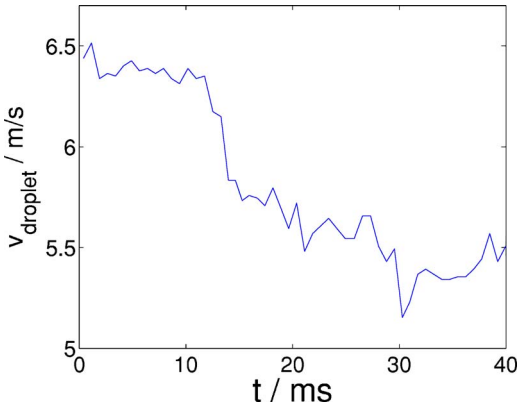


FIG. 8. (Color online) Droplet velocity as a function of time. When an air bubble is entrapped at $t=0$, the droplet speed decreases from 6.5 to 5.5 m/s within 20 ms, corresponding to 400 actuation cycles.

the camera. The shutter was set to 2 μs to prevent motion blur. A typical example recorded with 100 kfps is shown in Fig. 9(a). The corresponding variance of the acoustical signal is plotted in Fig. 9(b). Droplet 26 was created normally, but droplet 27 has a small twist in the tail of the drop. The reason becomes clear when looking at droplet 28, which shows a large disturbance being jetted out. This disturbance caused the droplet formation of droplet 27 to be modified before it was jetted out in the next actuation. After this the droplet formation returns to normal.

Figure 9 typifies a disturbance that potentially leads to air entrapment. To find the origin of the disturbance, some deviating droplets were caught and analyzed with a microscope. The capturing of the droplets was done by creating a piezoelectric device that can move several microns within several microseconds. A microscope glass cover was attached to the tip of the piezoelectric device. This device was placed beneath the nozzle plate, a few microns away from the stream of droplets. The acoustical signal was used as a trigger for the piezoelectric device. When triggered, the piezo moves into the droplet stream and captures a single droplet on the glass cover. This glass cover was then removed from the piezoelectric device and placed under a microscope. The captured transparent ink droplets contain relatively large particles (± 20 μm) that are not present in undisturbed droplets. It is therefore believed that these particles disturb the normal jetting process by creating an asymmetric meniscus, which then could lead to the entrapment of air. The chance of air entrapment depends on the actual position of the particle and the precise time within an acoustical cycle. We conclude that small particles in the ink can result in an altered droplet formation, which then can result into air entrapment.

B. Air entrapment with an ink layer

An ink layer on the nozzle plate can also induce air entrapment while jetting droplets. Therefore experiments are set up to clarify the exact role of an ink layer on air entrapment. During the experiments the time between consecutive actuations, and thus the created droplets, is kept constant at 100 μs. The ink layer is created by applying a higher ambient pressure to the ink reservoir inside the printhead for about one second (typically, $\Delta p = 100$ mbar enhancement, as compared to the standard slight reduction of 8 mbar, which under operational conditions is employed to avoid ink leakage from the nozzles; see Sec. II). This higher ambient pressure is negligible compared to the pressure created by the piezos of the actuated channels. It results in ink flowing out of the nonactuating neighboring nozzles onto the nozzle

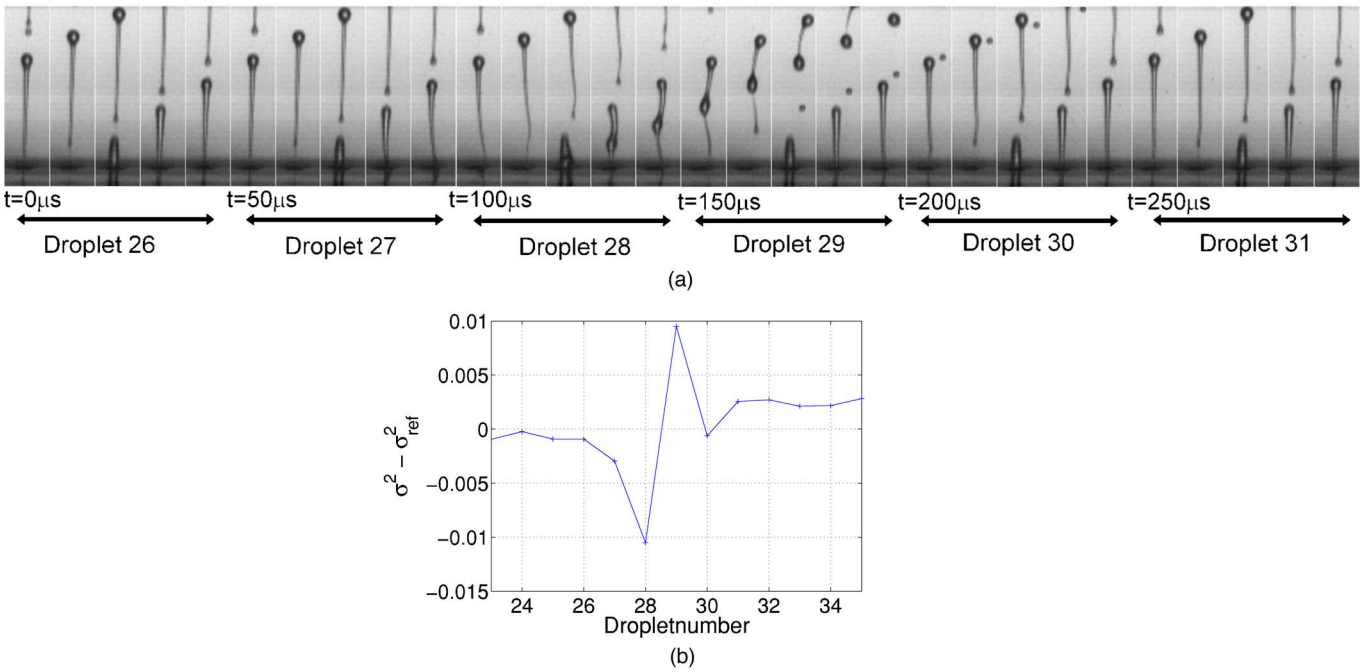


FIG. 9. (Color online) (a) Droplet formation recorded at 100 kfps showing the disturbance that can result in air entrapment. Droplet 26 shows regular droplet formation. Droplet 27 displays a slight deviation in the tail. Droplet 28 shows a large disturbance being jetted out. Droplets 29 and 30 display regular droplet formation again. (b) The variance of the acoustical signals is plotted for the corresponding droplets. The variance signals a large deviation at droplets 27 and 28, corresponding to the disturbance in the droplet formation process.

plate. The ink flows over the nozzle plate and reaches the jetting nozzle. Its effect on the droplet formation is documented in Fig. 10: The ink layer causes a reduction of the droplet velocity. The jet speed reduces with increasing ink layer thickness. At a critical ink layer thickness d_2 the nozzle ceases to fire droplets. When the pressure applied to the ink reservoir is back to its slightly reduced value, the surrounding nozzles suck the ink back into the ink channels. When the ink layer becomes thinner, at the critical thickness d_2 droplets can be jetted again. Finally, the ink layer on the nozzle plate is fully removed. However, it turns out that now air is entrapped in the actuating ink channel. This can be seen in the acoustical signal, such as shown in Fig. 4(b), but also by the lower droplet velocity, which is caused by the acoustical damping through the air bubble. Therefore it is concluded that an ink layer thickness window, $d_1 < d < d_2$, exists, in which air bubbles are entrapped. For smaller thicknesses, $d < d_1$, in general, there is no air entrapment, and for larger thicknesses, $d > d_2$, the meniscus motion is neither sufficient to jet droplets nor to entrap air.

Two typical examples of this process recorded at 10 kfps are shown in Fig. 10 (side view) and in Fig. 11 (top view). Notice the air bubbles on the nozzle plate in the last image of Fig. 11. When the air bubbles are on the nozzle plate, they slowly dissolve in the ink layer, as the pressure fluctuations are not sufficient to support growth by rectified diffusion. As several air bubbles turn up on the nozzle plate, apparently not only one air bubble has been entrapped in the ink channel. This is probably caused by the difference in time scale between the changing of the thickness of the ink layer and the actuation. As the change in the ink layer thickness occurs slowly compared to the actuation frequency, an ink layer with thickness $d_1 < d < d_2$ is present during multiple actuations. If an actuation occurs while the ink layer is within this thickness range, an air bubble is created. So multiple actuations lead to multiple air bubbles. The size of the air bubbles could possibly even reveal information on whether it has moved into the ink channel or has remained on the nozzle plate.

Though these measurements show the existence of an ink layer thickness window for air entrapment, they do not allow us to determine the exact values of the critical thicknesses d_1 and d_2, since the ink layer thickness varies over the nozzle plate. The determination of the critical thicknesses d_1 and d_2 is the subject of the next subsection.

C. Determination of the critical ink layer thickness

In order to find out and to determine the ink layer thicknesses where air entrapment occurs, an additional electroformed nickel nozzle plate is placed on top of the original nozzle plate, as shown in Fig. 12. Different thicknesses of the additional nozzle plate are used to vary the ink layer thickness.

One problem, however, arises: The volume of the additional nozzle is small compared to the volume of a complete ink layer. While jetting, the droplet formation could also change the amount of ink in the additional nickel nozzle of Fig. 12. To overcome this problem, a burst of actuations is

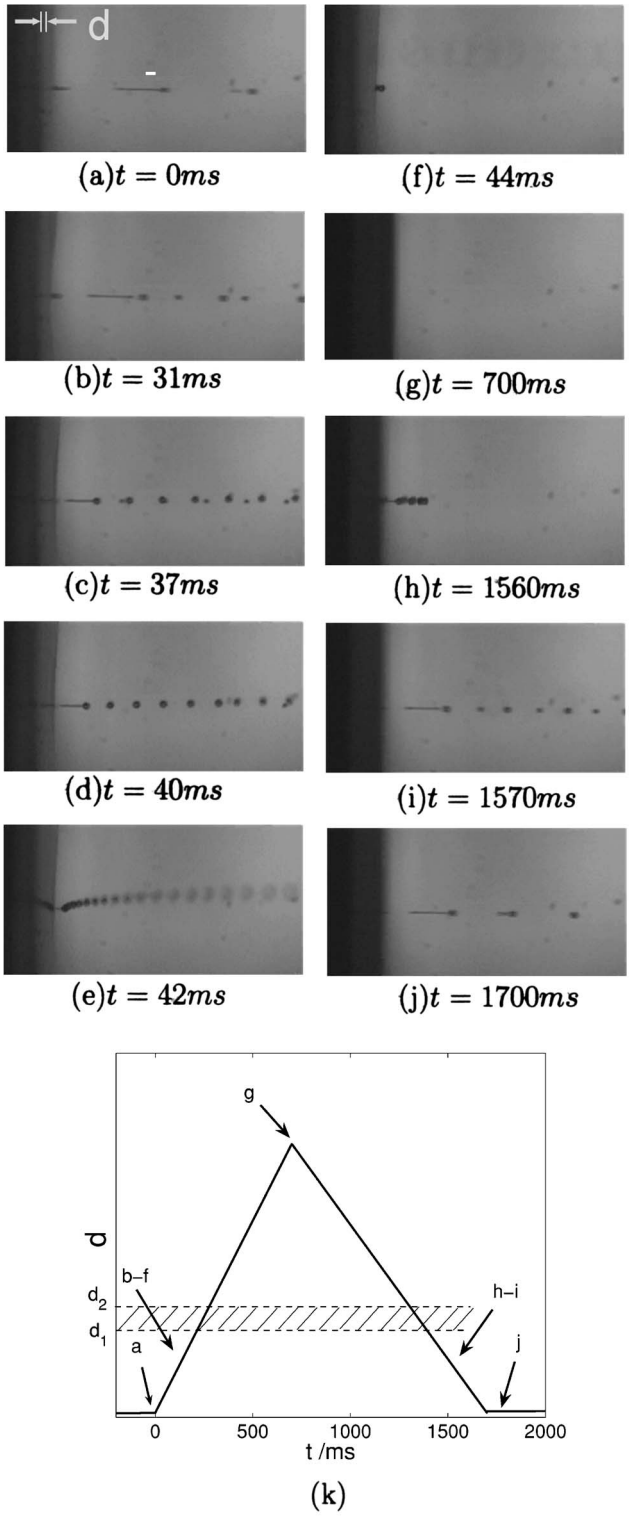


FIG. 10. High speed recordings at 10 kfps of droplets when a temporary ink layer is created on the nozzle plate by applying a higher ambient pressure to the ink reservoir for 700 ms. (a) Regular jetting behavior with no ink layer present. (b)–(e) The higher ambient pressure is applied, resulting in a growing ink layer. The velocity of the droplets is reduced through the ink layer. (f) No droplets are jetted anymore because all acoustic energy is dissipated in the ink layer. (g) The ink layer is very thick and no droplets are created at all. (h)–(j) The enhancement of the ambient pressure has ended, resulting in a reducing ink layer thickness. When the layer is thin enough, the jetting starts again, but now with an air bubble in the ink channel. (k) A schematic diagram of the increasing and decreasing ink layer.

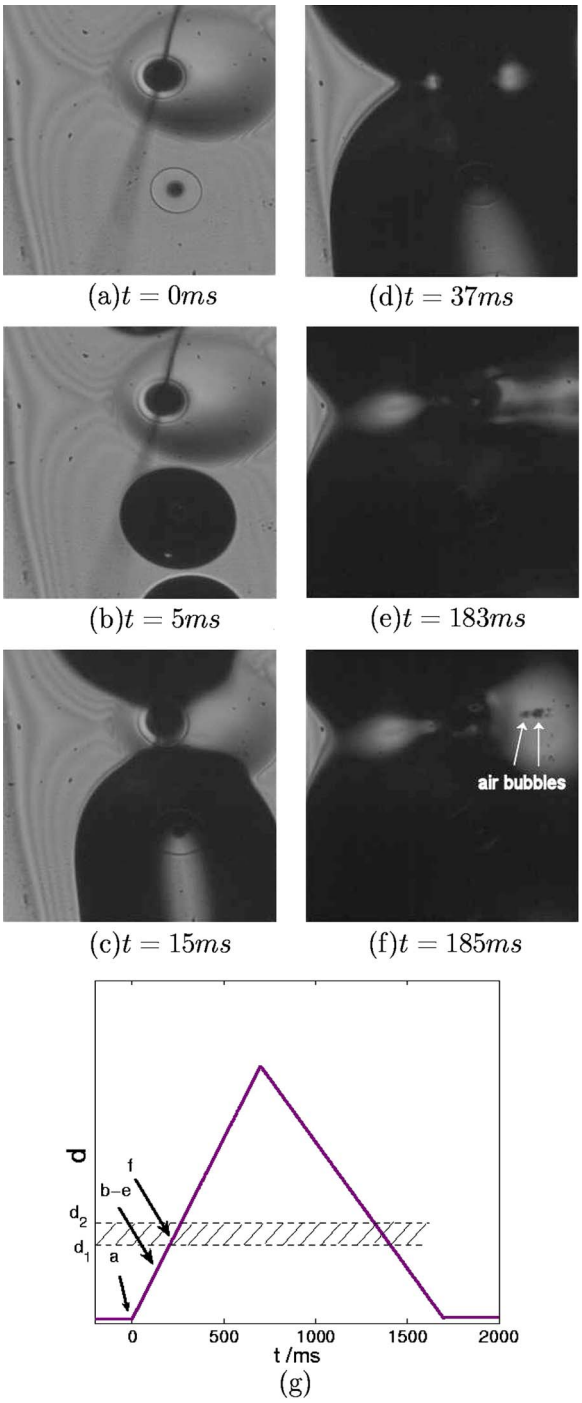


FIG. 11. High speed recordings of droplets at 10 kfps when a temporary ink layer is created on the nozzle plate by applying a higher ambient pressure to the ink reservoir for 700 ms. (a) Normal jetting with no ink layer present. (b)–(d) The higher ambient pressure is applied and the ink layer on the nozzle plate is growing. (e) The droplets now have a reduced speed and a deviant angle of jetting. (f) When the ink layer reaches the critical thickness d_1, air bubbles are created and are present on the nozzle plate. For $d > d_2$ jetting completely breaks down. (g) A schematic diagram of the increasing and decreasing ink layer.

applied. The burst consists of three actuations, after which a waiting period of 3 ms is set. The three actuations are chosen because not only air has to be entrapped, but the air bubble also needs to be detected through the acoustical signal. By controlling the ambient pressure of the ink reservoir, it is ensured that the additional nozzles (Fig. 12) are completely

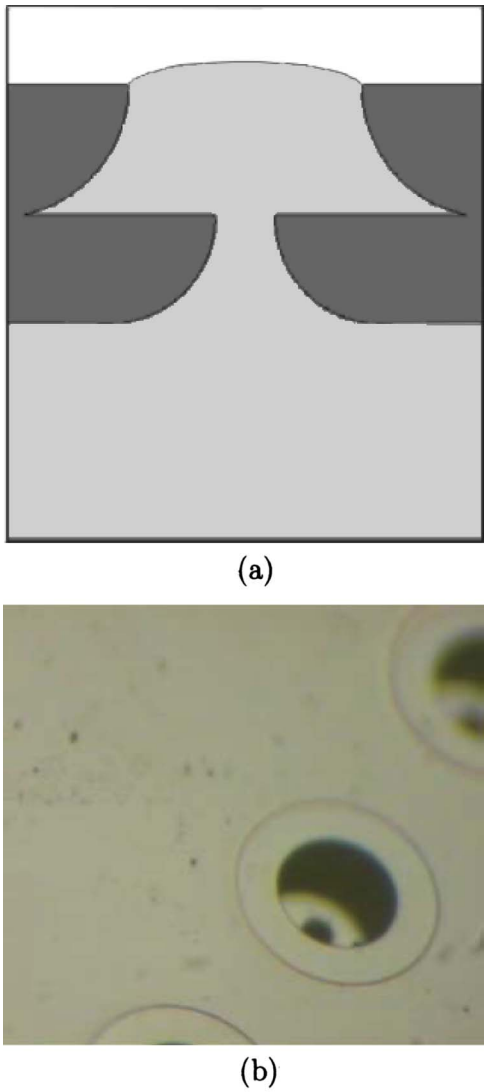


FIG. 12. (Color online) (a) Schematic side view of the additional nozzle plate on top of the original one. (b) Top view photograph of the setup.

filled by the setting of the ambient pressure of the ink reservoir. Measurements show that when the ink layer is relatively thin (<30 μm), or relatively thick (>40 μm), no air entrapment occurs. The region in between is the ink layer thickness window where air entrapment occurs.

D. Numerical simulations on ink layer triggered air entrainment

To qualitatively check whether air entrapment induced by ink layers on the nozzle is also reflected in numerical simulations, a Flow-3D simulation is set up, similar to the one described in Sec. II. For this simulation an ink layer is positioned on the nozzle plate with a thickness of 30 μm. It is assumed that the influence of the ink layer on the acoustical waves inside the printhead is negligible. Two actuations are modeled.

The result is presented in Fig. 13. It can be seen that indeed no air is entrapped during the first acoustic cycle, which lasts for 50 μs. During the second acoustic cycle, air is, however, entrapped. The reason for the delay is that the simulation is started with no fluid movement at all. The ac-

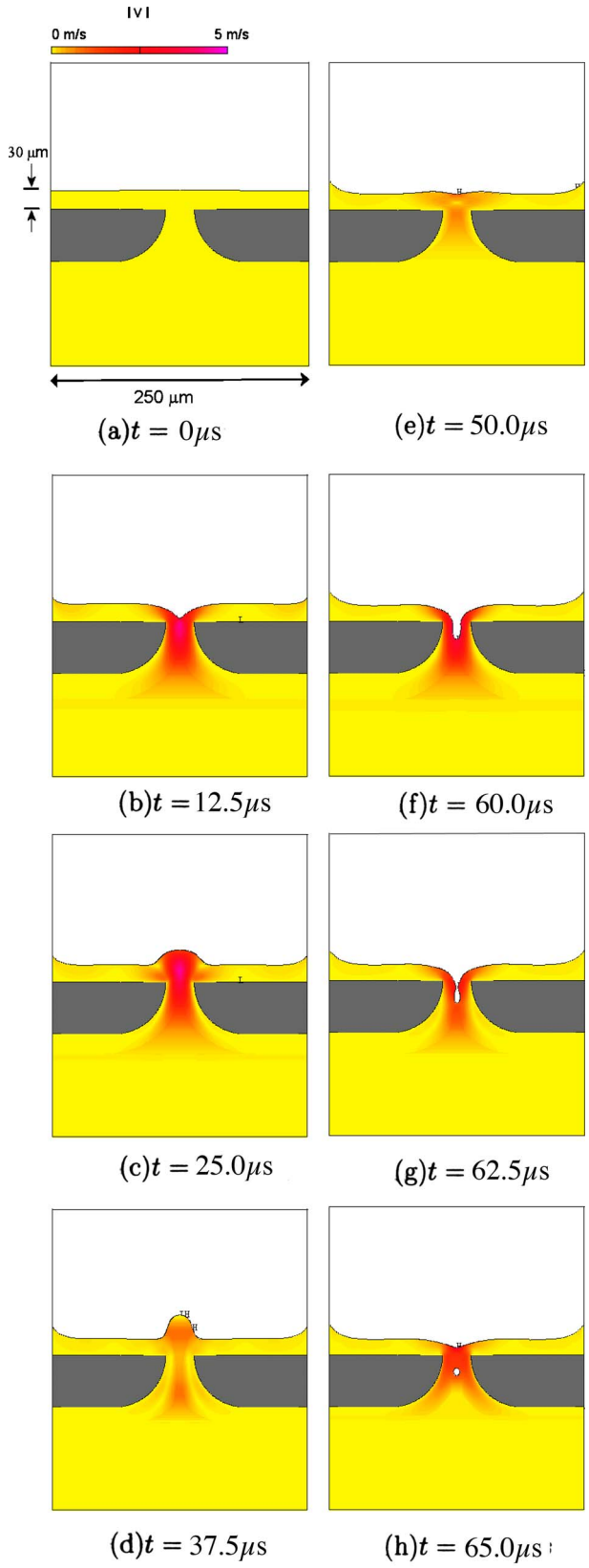


FIG. 13. (Color online) Air entrapment with an ink layer simulated with Flow-3D. (a) A 30 μm ink layer is placed on the nozzle plate. The actuation starts at $t=0$ μs. (b) The meniscus is pulling back due to the first negative pressure wave. (c)–(d) The positive pressure waves pushes the ink out of the nozzle. (e) Start of the second actuation. (f) The negative pressure wave pulls back the meniscus. (g) Due to the larger surrounding pressure, the void is closed. (h) An air bubble is entrapped.

tual air entrapment occurs when the meniscus is pulling back and ink is flowing into the nozzle. If no ink layer was present, the meniscus would retract into the nozzle. However, in the case of an ink layer, also the ink on top of the nozzle plate is pulled into the nozzle. The void is then closed at the top because of the ink flowing in from the sides due to the overpressure.

VI. SUMMARY, DISCUSSION, CONCLUSIONS, AND OUTLOOK

By using the piezoactuator as a sensor, the channel acoustics and the droplet formation can be accurately monitored. In particular, disturbances that can result in air entrapment and the entrapped bubbles themselves are detected by the acoustical signal. Since hitherto the only solution for recovering the regular droplet formation process is the dissolution of the bubble, it is crucial to detect the bubble as soon as possible after entrapment. In that case the bubble has not yet grown by rectified diffusion, and its dissolution time is therefore shorter. By employing the acoustical signal as a trigger, two air entrapment mechanisms are found:

- Air entrapment triggered by particles in the ink, disrupting the droplet formation
- Air entrapment caused by an ink layer on top of the nozzle plate

The first mechanism shows that particles in the ink, for example dust particles, can influence the droplet formation in such a way that air is entrapped. The size of the particles was found to be around 20 μm. Other sized particles will not result in air entrapment since small particles ($\ll$20 μm) will have no influence at all, and large particles ($\gg$20 μm) will block the nozzle and thus the droplet formation. The 20 μm particles probably cause a local surface tension distortion and thus an asymmetry of the droplet formation. This asymmetry of the completely retracted meniscus in combination with the next symmetric pressure wave then could cause air entrapment. The timing, size, and position of the particle is expected to play a crucial role in the air entrapment mechanism.

In the second mechanism an ink layer on top of the nozzle plate of approximately 30–40 μm results in air entrapment. When the meniscus is pulled back due to the negative pressure inside the nozzle, ink from the nozzle plate flows into the nozzle and air is entrapped. This process is modeled with Flow-3D. Measurements show that not all the air bubbles enter the ink channel, but some also move onto the nozzle plate. The exact criterium for entering the nozzle is not clear. Probably parameters like the size, position, and timing play an important role.

Note that small air bubbles cause no problems in jetting. Only once they grow, problems arise. The air bubble starts to oscillate due to the acoustical waves. Because of the oscillations, the bubble will experience a net force directed into the ink channel, the primary Bjerknes force. The oscillations also cause the bubble to grow by rectified diffusion. The further the bubbles moves into the ink channel, the larger the pressure amplitude and therefore also the oscillations and the

growth of the bubble. Note that the properties of bubbles in a constrained geometry such as the ink channel can differ considerably from those of free bubbles.[20–22]

In this paper a method to detect air bubbles in piezo-driven printheads is presented and mechanisms of how bubbles are entrained are identified. From a practical point of view the ultimate goal, of course, must be to prevent air entrapment (or to immediately get rid of the entrained bubbles), in order to improve the jetting stability. The two identified air-entrainment mechanisms suggest two requirements:

- Particles with a size of 20 μm or larger should be prevented to reach the nozzle. Inside the printhead, this can be achieved through filters. But dust particles from outside also form a threat. Printing in a clean environment should prevent particles reaching the nozzle. Fortunately, for printers in an industrial environment it is often possible to control the cleanness of the ambient air (clean-room facilities).
- The ink layer on the nozzle plate should stay below a thickness of 30 μm. This may be accomplished by a special design of the nozzle plates.

More research is needed to clarify the process from air entrapment of a small bubble to a full grown bubble, which eventually prevents the normal jetting of droplets. We have meanwhile optically monitored the trajectory, growth, and size of the bubble with high-speed imaging.[23] The next step is to understand its effect on the pressure field, in order to quantitatively describe the reduction of the droplet jetting velocity and to find out the exact conditions when the jetting will totally break down. Research in this direction is on its way.

ACKNOWLEDGMENTS

We thank the staff of Océ Technologies B.V. for support and Andrea Prosperetti for discussions. This study has been financed by the Fundamenteel Onderzoek der Materie (FOM) of The Netherlands under Grant. No. 02MFS39 and by Océ Technologies B.V.

[1]H. P. Le, "Progress and trends in ink-jet printing technology," J. Imaging Sci. Technol. **42**, 49 (1998).
[2]D. B. Bogy and F. E. Talke, "Experimental and theoretical study of wave propagation phenomena in drop-on-demand ink jet devices," IBM J. Res. Dev. **28**, 314–320 (1984).
[3]J. M. Meacham, M. J. Varady, F. L. Degertekin, and A. G. Fedorov, "Droplet formation and ejection from a micromachined ultrasonic droplet generator: Visualization and scaling," Phys. Fluids **17**, 100605 (2005).
[4]W. T. Berggren, M. S. Westphall, and L. M. Smith, "Single-pulse nano-electrospray ionization," Anal. Chem. **74**, 3443 (2002).
[5]C.-Y. Lung, M. D. Barnes, N. Lermer, W. B. Whitten, and J. M. Ramsey, "Single-molecule analysis of ultradilute solutions with guided streams of 1-μm water droplets," Appl. Opt. **38**, 1481 (1999).
[6]A. U. Chen and O. A. Basaran, "A new method for significantly reducing drop radius without reducing nozzle radius in drop-on-demand drop production," Phys. Fluids **14**, L1 (2002).
[7]S. A. Elrod, B. Hadimioglu, B. T. Khuri-Yakub, E. G. Rawson, E. Richley, C. F. Quate, N. N. Mansour, and T. S. Lundgren, "Nozzleless droplet formation with focused acoustic beams," J. Appl. Phys. **65**, 3441 (1989).
[8]D.-Y. Shin, P. Grassia, and B. Derby, "Oscillatory limited compressible fluid flow induced by the radial motion of a thick-walled piezoelectric tube," J. Acoust. Soc. Am. **114**, 1314–1321 (2003).
[9]J. F. Dijksman, "Hydro-acoustics of piezoelectrically driven ink-jet print heads," Flow, Turbul. Combust. **61**, 211–237 (1999).
[10]J. F. Dijksman, "Hydrodynamics of small tubular pumps," J. Fluid Mech. **139**, 173–191 (1984).
[11]J. D. Brock, I. M. Cohen, I. P. Ivanov, H. P. Le, and J. Roy, "Oscillations of an air bubble in an ink jet," J. Imaging Sci. Technol. **10**, 127–129 (1984).
[12]N. P. Hine, "Deaeration system for a high-performance drop-on-demand ink jet," J. Imaging Technol. **17**, 223–227 (1991).
[13]C. E. Brennen, *Cavitation and Bubble Dynamics* (Oxford University Press, Oxford, 1995).
[14]M. M. Fyrillas and A. J. Szeri, "Dissolution or growth of soluble spherical oscillating bubbles," J. Fluid Mech. **277**, 381–407 (1994).
[15]S. Hilgenfeldt, D. Lohse, and M. P. Brenner, "Phase diagrams for sonoluminiscescing bubbles," Phys. Fluids **8**, 2808–2826 (1996).
[16]Flow-3D is CDF software developed by Flow Science Inc., Santa Fe, New Mexico.
[17]C. W. Hirt and B. D. Nichols, "Volume of Fluid (VOF) method for the dynamics of Free Boundaries," J. Comput. Phys. **39**, 201–225 (1981).
[18]J. Eggers, "Nonlinear dynamics and breakup of free-surface flow," Rev. Mod. Phys. **69**, 865–929 (1997).
[19]H. Wijshoff, "Free surface flow and acousto-elastic interaction in piezo inkjet," *Proceedings of NSTI Nanotech 2004 Conference*, 2004, Vol. **2**, pp. 215–218 (http://www.flow3d.com/pdfs/bib2-04.pdf).
[20]X. M. Chen and A. Prosperetti, "Thermal processes in the oscillations of gas bubbles in tubes," J. Acoust. Soc. Am. **104**, 1389–1398 (1998).
[21]H. N. Oguz and A. Prosperetti, "The natural frequency of oscillation of gas bubbles in tubes," J. Acoust. Soc. Am. **103**, 3301–3308 (1998).
[22]X. Geng, H. Yuan, H. N. Oguz, and A. Prosperetti, "The oscillation of gas bubbles in tubes: Experimental results," J. Acoust. Soc. Am. **106**, 674–681 (1999).
[23]J. de Jong *et al.*, in preparation (2006).
[24]In principle, the units of $a(t)$ would be Nm^2/s; however, the signal is linearly converted into a voltage, which is measured in Volts. We decided to omit the units of $a(t)$ in this paper.

Experimental characterization of fatigue damage in a nickel-base superalloy using nonlinear ultrasonic waves

Jin-Yeon Kim, Laurence J. Jacobs,[a] and Jianmin Qu
G. W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

Jerrol W. Littles
Pratt & Whitney, Materials and Processes Engineering, 400 Main Street, M/S 114-40, East Hartford, Connecticut 06108

(Received 10 February 2006; revised 14 June 2006; accepted 16 June 2006)

This research develops a robust experimental procedure to track the evolution of fatigue damage in a nickel-base superalloy with the acoustic nonlinearity parameter, β, and demonstrates its effectiveness by making repeatable measurements of β in multiple specimens, subjected to both high- and low-cycle fatigue. The measurement procedure developed in this research is robust in that it is based on conventional piezoelectric contact transducers, which are readily available off the shelf, and it offers the potential for field applications. In addition, the measurement procedure enables the user to isolate sample nonlinearity from measurement system nonlinearity. The experimental results show that there is a significant increase in β linked to the high plasticity of low-cycle fatigue, and illustrate how these nonlinear ultrasonic measurements quantitatively characterize the damage state of a specimen in the early stages of fatigue. The high-cycle fatigue results are less definitive (the increase in β is not as substantial) due to increased uncertainties involved in the high-cycle fatigue tests, but still show a clear relationship between β and remaining fatigue life. One application of the measured β versus fatigue-life data is to potentially serve as a master curve for life prediction based on nonlinear ultrasonic measurements. [DOI: 10.1121/1.2221557]

PACS number(s): 43.25.Zx, 43.25.Dc, 43.25.Ba [MFH] Pages: 1266–1273

I. INTRODUCTION

Recent experimental studies and new physical models are demonstrating the potential of nonlinear ultrasonics (or the second-harmonic generation technique) to quantitatively detect and characterize fatigue damage in metals.[1–10] This fatigue damage first appears in the form of dislocation substructures, such as veins and persistent slip bands (PSBs), and these PSBs accumulate at grain boundaries to produce strain localization and, then finally, microcrack initiation with increasing fatigue cycles. These dislocations (and resulting microplastic deformation) do not cause a large change in the linear macroscopic properties (such as elastic moduli, sound speed, and attenuation) of a material; the changes in the linear ultrasonic values are not large enough to be accurately measured with conventional linear ultrasonic techniques. However, the accumulation of dislocations throughout the continuum (with increasing fatigue) will cause a nonlinear distortion in an ultrasonic wave propagating in the material, and thus generate higher harmonic components in an initially monochromatic ultrasonic wave signal. For this reason, nonlinear ultrasonic (acoustic) waves can be used to quantify the presence and the density of dislocations in a metallic material, and thus measure fatigue damage in a quantitative fashion. In addition, nonlinear ultrasonics has the potential to promote an understanding of the evolution and accumulation of the dislocation substructures in the very early stages of fatigue.

To date, a number of investigators[1–8] have applied nonlinear ultrasonic techniques to assess fatigue damage in different materials under relatively controlled laboratory conditions. Yost and Cantrell[1] and Cantrell and Yost[6] experimentally observed changes of the acoustic nonlinearity parameter, and attributed the changes to the effects of fatigue-induced dislocations. Frouin *et al.*[5,8] performed *in situ* nonlinear ultrasonic measurements during fatigue testing, and related the measured increase in the acoustic nonlinearity parameter—in the vicinity of the fracture surface—to an increase in the dislocation density. Among these studies, only Frouin *et al.*[8] reported using nonlinear ultrasonic results to track fatigue damage throughout the entire fatigue life of a specimen. One field application of nonlinear ultrasonics examined fatigue damage in stainless-steel turbine blades.[7] In spite of the recognized potential of nonlinear ultrasonics, there are very few examples of its successful application to monitor fatigue damage. This is probably due to instrumentation issues that make accurate and consistent nonlinear ultrasonic measurements difficult, plus a lack of flexibility in the measurement setup needed to interrogate real fatigue test specimens. A critical next step—for the incorporation of nonlinear ultrasonic techniques into life-prediction strategies of structural components—is a system-

[a]Also at: School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0355; electronic mail: laurence.jacobs@ce.gatech.edu

atic study that quantifies the robustness, accuracy, and validity of nonlinear ultrasonics to detect the early stages of fatigue damage (prior to crack initiation) in metallic materials. Of particular interest is the development of an experimental procedure with the capability of performing field inspections of an absolute and repeatable nature.

The objective of the current research is to develop a robust experimental procedure to track the evolution of fatigue damage in metallic materials with the acoustic nonlinearity parameter, β. The effectiveness of this proposed measurement procedure is demonstrated by making repeatable acoustic measurements of β, in nickel-base superalloy specimens, subjected to three types of damage: Quasi-static monotonic, low-, and high-cycle fatigue. These nonlinear ultrasonic measurements are used to track the evolution of damage in multiple specimens with a series of interrupted mechanical tests—first by making a baseline measurement of β in an undamaged specimen, then introducing some damage into the specimen, repeating the measurement of β in this same (unloaded) specimen, then introducing more damage into the specimen, and repeating the procedure.

It is important to note that the acoustic nonlinearity parameter, β, is an absolute material constant, which can be related to the higher-order elastic constants, of a material; the β parameter is a directly measurable acoustic parameter that is linked to the state of material damage. As a result, nonlinear ultrasonics is unparalleled in its potential to provide a robust and quantitative characterization of fatigue damage in in-service structural components. However, the acoustic nonlinearity associated with fatigue damage is very small, and can be easily overwhelmed by a number of other factors (especially instrumentation nonlinearity) inherent to the measurement procedure. Therefore, a critical contribution of this research is a systematic experimental procedure that can identify and remove spurious sources of nonlinearity, isolating only those contributions due to the material and associated damage.

II. GENERATION OF HIGHER HARMONICS AND THE ACOUSTIC NONLINEARITY PARAMETER, β

The equations of motion of a solid element, in the absence of body forces, are written in material coordinates, $\boldsymbol{X}$, as

$$\rho\frac{\partial^2 u_i}{\partial t^2}=\frac{\partial \sigma_{ij}}{\partial X_j}, \tag{1}$$

where t is time, ρ is the mass density, u_i is the displacement vector, and σ_{ij} is the stress tensor. The stress in a nonlinear (fatigued) solid can, in general, be written as

$$\sigma_{ij}=\sigma_{ij}^0+A_{ijkl}\frac{\partial u_k}{\partial X_l}+\frac{1}{2}A_{ijklmn}\frac{\partial u_k}{\partial X_l}\frac{\partial u_m}{\partial X_n}+\cdots, \tag{2}$$

where σ_{ij}^0 is the residual stress in the material, and A_{ijkl} and A_{ijklmn} are the Huang coefficients,[9] which are related to the second- and third-order elastic constants by $A_{ijkl}=\sigma_{jl}^0\delta_{ik}+\bar{C}_{ijkl}$ and $A_{ijklmn}=\bar{C}_{ijklmn}+\bar{C}_{jlmn}\delta_{ik}+\bar{C}_{ijnl}\delta_{km}+\bar{C}_{jnkl}\delta_{im}$. $\bar{C}_{ijkl}$ and $\bar{C}_{ijklmn}$ are modified by fatigue damage (dislocation substructures)[9] from their initial values, C_{ijkl} and C_{ijklmn}.[11] Expressions for the modified elastic constants during fatigue have been presented in terms of the residual stress and plastic strain.[12] The density of a material undergoing finite deformation is given by $\rho=\rho_0/\det\mathbf{F}$ where ρ_0 is the constant density in the unstressed configuration and $\mathbf{F}$ is the deformation gradient tensor, defined as $F_{ij}=\delta_{ij}+\partial u_j/\partial X_i$. Substituting Eq. (2) into Eq. (1), and considering one-dimensional wave propagation of a longitudinal wave in an isotropic solid, one gets

$$\frac{\partial^2 u_1}{\partial t^2}=c^2\frac{\partial^2 u_1}{\partial X_1^2}\left(1+\beta\frac{\partial u_1}{\partial X_1}\right), \tag{3}$$

where $c=\sqrt{(\bar{C}_{1111}+\sigma_{11}^0)/\rho}$ is the longitudinal wave speed and β is the acoustic nonlinearity parameter defined as

$$\beta=\frac{\bar{C}_{111111}+3\bar{C}_{1111}}{\bar{C}_{1111}+\sigma_{11}^0}. \tag{4}$$

It is well known that the second-order elastic constant [$\bar{C}_{1111}$ in Eq. (4)] changes very little, and that the residual stress (σ_{11}^0) is relatively small compared to the elastic constants. Therefore, it is the third-order elastic constant [$\bar{C}_{111111}$ in Eq. (4)] which causes the increase in the acoustic nonlinearity parameter, β, during fatigue.

Consider a time-harmonic plane (displacement) wave $A_1\cos(kX_1-\omega t)$, where A_1 is the amplitude, k is the wave number, and ω is the angular frequency. Assuming that the nonlinearity in the solid is small, the solution to Eq. (3) for this time-harmonic wave is obtained by a perturbation analysis as[13]

$$\begin{aligned}u_1&=-\tfrac{1}{8}\beta k^2A_1^2X_1+A_1\cos(kX_1-\omega t)\\&\quad+\tfrac{1}{8}\beta k^2A_1^2X_1\cos[2(kX_1-\omega t)]+\cdots\\&=A_0+A_1\cos(kX_1-\omega t)+A_2\cos[2(kX_1-\omega t)]+\cdots.\end{aligned} \tag{5}$$

It is noted that the amplitude of the second-harmonic displacement is proportional to the acoustic nonlinearity parameter and a subharmonic; that is, the static displacement is induced by the material nonlinearity.[14] The acoustic nonlinearity parameter is determined experimentally by measuring the absolute amplitudes of the fundamental (A_1) and the second-harmonic (A_2) displacement signals, or

$$\beta=\frac{8A_2}{k^2X_1A_1^2}. \tag{6}$$

Finally, note that Eq. (6) neglects the effect of attenuation losses that may be present in the fundamental and second-harmonic. If the difference in attenuation rates at the fundamental and the second-harmonic frequencies is large, then a correction factor must be included in the measurement of β. The specific superalloy examined in this research is IN100, which is produced by powder metallurgy and has a very fine grain structure. Attenuation measurements are made in IN100 through the range of 1–15 MHz before fatigue tests. These results show that the attenuations at the fundamental and the second-harmonic frequencies are about

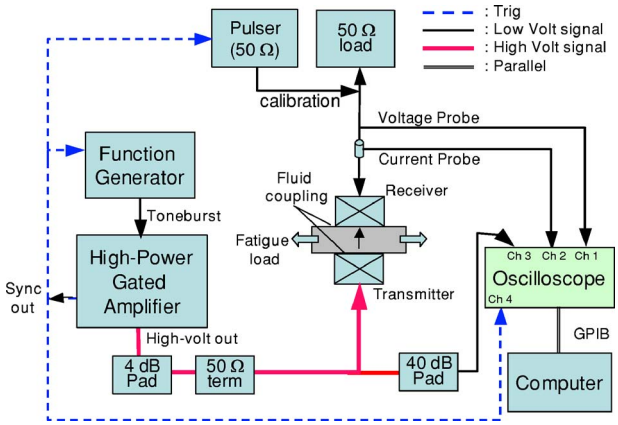


FIG. 1. (Color online) Experimental setup.

0.14 Neper/cm and 0.34 Neper/cm, respectively, which correspond to a maximum correction of less than 2% in β. Furthermore, the fatigue specimens show no noticeable change in attenuation in the frequency range considered here, so no attenuation corrections are made for the following β calculations.

III. EXPERIMENTAL PROCEDURE

A. Measurement system and procedure

Figure 1 shows a schematic of the proposed nonlinear ultrasonic measurement system. A tone burst signal of 7–9 cycles (depending on the specimen thickness) at 5.3 MHz is generated by a function generator (80 MHz Agilent 33250A) and is fed into a high-power gated amplifier (Ritec RAM-10000). In order to ensure one-dimensional wave propagation in a single direction (only right or left propagating), the exact number of cycles of the tone burst is selected as the maximum number of cycles that can fit within the thickness of the specimen—the spatial length of the tone burst is less than the specimen thickness. This eliminates any possible spurious (apparent) higher harmonics generated by the interference of the incident and reflected wavefronts, as well as the effects of boundary conditions. The amplified high-voltage signal passes through a 4 dB attenuator (pad) and a 50 Ω termination to suppress the transient behavior due to the mismatch in electrical impedances between the amplifier and the transducer. Commercial narrow-band PZT (Lead Zirconate Titanate)-base piezoelectric transducers, with center frequencies of 5 MHz and 10 MHz, are used as a transmitter and a receiver, respectively. The transducers are coupled to the specimen with light lubrication oil. A special fixture is designed to keep both the transmitting and receiving transducers aligned on the same centerline axis, and to also allow for the removal of either transducer (transmitter or receiver) without disturbing the coupling (and position) of the other; this capacity is critical for the calibration procedure described next. The receiver is terminated with a 50 Ω passive load to have the same terminal load in the calibration. Both voltage and current signals of the transmitted ultrasonic waves are recorded and averaged 256 times with an oscilloscope, and then transferred to a computer for further signal processing. Then, diffraction corrections are made to the measured fundamental and the second-harmonic signal amplitudes.

The calibration procedure for the (piezoelectric) receiving transducer is based on the principle of self-reciprocity,[15] and is employed in order to obtain a conversion transfer function (from the measured electrical signal to the absolute amplitude of the particle displacement), and to compensate for any (small) variations in the coupling of the receiving transducer. Note that this calibration is performed prior to every nonlinear measurement, with the transmitter transducer removed. A 50 MHz pulser/receiver (Panametrics, 5072PR) is used to transmit (through the receiver transducer) a wide-band ultrasonic pulse through the specimen. The current and voltage signals of the incident and the reflected pulse from the bottom surface of the specimen that is kept stress-free (when the transmitter is removed), are measured and used to calculate a transfer function that converts the measured current signal to the particle displacement of the incident wave at the receiver.[15]

Finally, the pulse-inversion technique[16,17] is applied to accentuate the contribution of the even (second) harmonic signal, while reducing the dominance of the fundamental contribution. The pulse-inversion technique is very efficient in extracting this second-harmonic amplitude by canceling out the odd harmonics (which are mainly due to the instrumentation); the even harmonic signal is extracted by adding two 180° out-of-phase input signals.[18] Figure 2 illustrates the pulse-inversion technique by showing both the 0° phase, and the 180° out-of-phase (inverted) signals, the respective Fourier spectra before and after addition (in the time domain), and the second-harmonic signal extracted. For the actual procedure, first, two transmitted time domain signals with 180° out-of-phase inputs are measured consecutively with all other conditions unchanged. A function generator performs phase inversion of the input pulse. Then, two separately measured output signals are combined in the time domain, extracting the second-harmonic signal. Note that this combination is performed with two raw signals without introducing any adjustments, such as time shifts or amplitude modification. Figure 2 clearly demonstrates how the fundamental frequency contribution is completely canceled out, leaving only the second-harmonic contribution. Note that the remaining subharmonic component (at zero frequency) corresponds to the first term in Eq. (5), and appears as a result of the static displacement induced by the acoustic radiation; this component should have an amplitude proportional to the amplitude of the second-harmonic,[19] but it is not systematically analyzed in this study. The frequency spectra of the signal originally transmitted, and the extracted signal shown in Fig. 2(b) are independently calculated with a rectangular window. Finally, to obtain a more accurate estimation of the amplitudes of the fundamental and second-harmonics, the signals are digitally filtered in the frequency domain and inverse Fourier transformed. An additional advantage of using the pulse-inversion technique is that one can readily monitor the shape of the second-harmonic signal. Since the amplitude of the second-harmonic signal produced by material nonlinearity is very small in comparison to the amplitude of the fundamen-

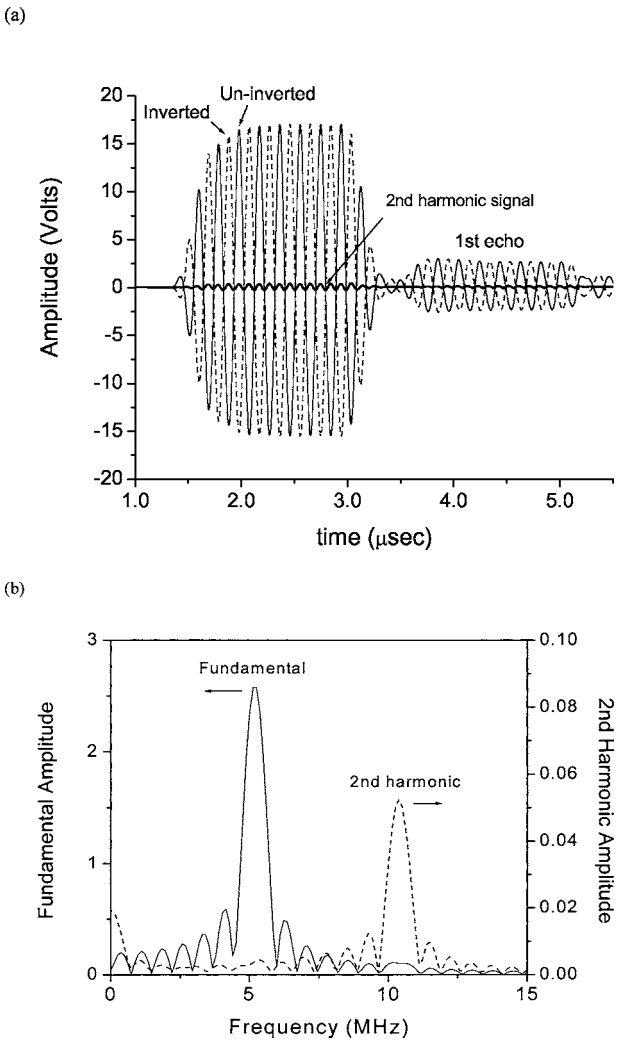


FIG. 2. (a) Typical time domain signals: Thin continuous and dotted lines are the transmitted signals with 0° phase (uninverted) and 180° out-of-phase (inverted) inputs, respectively. The thick line is the second-harmonic signal extracted by the pulse-inversion technique. (b) Fourier spectra of the original transmitted signal (fundamental) and the second-harmonic signal extracted by the pulse-inversion technique, demonstrating that the second-harmonic amplitude can be measured without being influenced by the large fundamental amplitude.

tal, small variations in coupling—that are usually accompanied by spurious interface nonlinearity—can have a significant influence on the repeatability of the proposed measurement procedure. Experience shows that the shape of this second-harmonic signal is an excellent indicator of the quality of the transducer to specimen coupling.

Figure 3 illustrates the linear relationship that exists between the measured absolute amplitudes of the second-harmonic and the squared fundamental (both displacements), as a function of increasing input voltage amplitude. These absolute displacement amplitudes are calculated using the transfer function described previously. Figure 3 shows the results of two independent measurements on the same specimen, where the transducers and the couplant are completely removed, and then reattached and recalibrated for the second measurement. It is seen that the slopes from these two independent measurements are nearly constant, which confirms that the measurements are repeatable, and that removal and

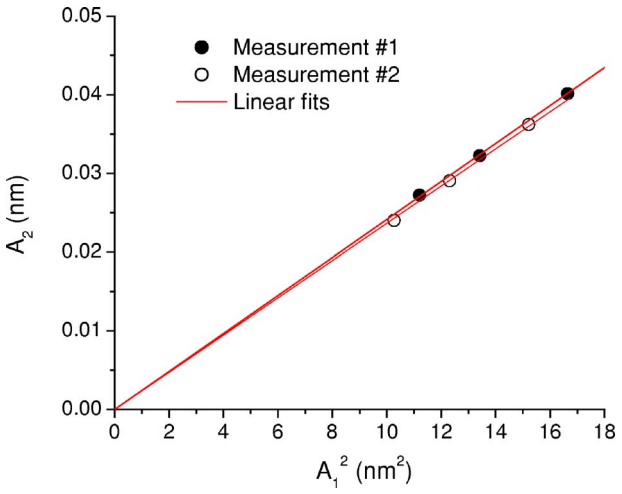


FIG. 3. (Color online) Second-harmonic amplitude (A_2) versus the amplitude of the fundamental squared $[(A_1)^2]$ for increasing input voltage obtained from two independent measurements.

replacement of the transducers (and couplant) will not have a dominant (negative) influence on the results. The variability (error bars) due to measurement error is determined by averaging five measurements on the same undamaged specimen, and results in a variability of ±0.45 on all β values reported henceforth. Finally, Fig. 3 can be used as a guide for the required input voltage needed to avoid inconsistencies caused by a low fundamental amplitude.[20]

The measurement system is calibrated by measuring β in borosilicate. It is known that borosilicate has a very low degree of nonlinearity, and researchers[21] have shown that the ratio of the second-harmonic amplitude to the fundamental amplitude is on the order of −120 dB (this unpublished reference value illustrates that the β of borosilicate is almost zero). A β of 9.0 is measured in borosilicate using the proposed measurement procedure, and this nonzero value of β is believed to be associated with the inherent nonlinearity of the transmitting piezoelectric transducers used in the measurement system. This is in agreement with previous researchers[22] who examined the nonlinear properties of PZT (polarized K1) and measured its β to be on the order of 8.0. Therefore, a β of 9.0 will be used to calibrate the measurement system by subtracting this value from all measured β values. Although such a calibration method neglects the interactions between different frequencies in a nonlinear system, it can be easily argued that the effects of such interactions on β are higher ordered. This is further verified by performing a β measurement on fused silica. The directly measured value of β for fused silica is 21.0. After calibration (subtracting 9.0), a β of 12.0 is obtained for fused silica, which is in agreement with published values.[23]

The good agreement between these results and published values, plus the consistency and repeatability of the results reported in Sec. IV, validate the accuracy of the proposed procedure as a working method to track changes in β as a function of fatigue life in multiple specimens.

B. Specimens

Three different types of specimens are used; each type machined from IN100 cylindrical rods—128 mm long and

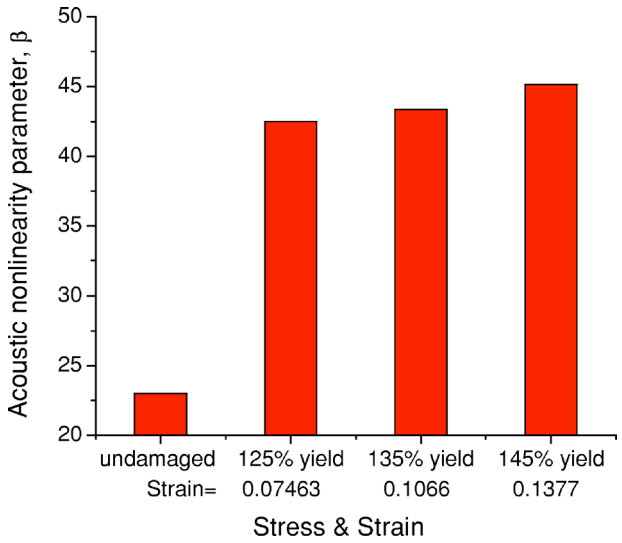


FIG. 4. (Color online) Monotonic load results—acoustic nonlinearity parameter β versus applied stress (or strain) level.

27 mm in diameter. Note that the surface finishes on all specimens in this study are "as-machined." The first type is a standard fatigue specimen with a constant rectangular cross section—a constant gauge width of 12.5 mm and a thickness of 6.4 mm. One of these specimens is used for the low-cycle fatigue tests, and one is used for the high-cycle fatigue tests. The second specimen has an hour-glass shape (starting from a width of 12.5 mm and a constant thickness of 6.4 mm) with a varying cross section that gradually reduces to create a region of higher stress at its center; this specimen is used exclusively for the high-cycle fatigue tests. The third specimen is a nonstandard rectangular bar specimen; it is simply a 120 mm long by 14.3 mm wide rectangular bar—having a constant thickness of 4.7 mm. This specimen is used for both the monotonic and low-cycle fatigue tests, and for a concurrent set of Rayleigh wave measurements.[24]

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Monotonic load results

First, consider a quasi-static monotonically loaded (nonstandard rectangular bar) specimen. This specimen is used to validate the repeatability of a set of nonlinear ultrasonic measurements made on a specimen subjected to an interrupted mechanical test—a test specimen that is mechanically loaded, removed to make a set of nonlinear ultrasonic measurements, and then the procedure is repeated at specified intervals, typically until the specimen fails. In this monotonic test, the specimen is loaded (at a rate of 890 N/s) to a first-load equivalent to 125% of the yield stress (absolute strain of 7.463%), and then is unloaded at the same rate. The nonlinear ultrasonic measurements are then performed on the unloaded specimen. The same procedure is repeated for increasing maximum loads equivalent to 135% and 145% of yield stress (strains of 10.66% and 13.77%, respectively). These calibrated results are presented in Fig. 4, and note that these measured acoustic nonlinearity parameters, β, are absolute values. It is important to note that the β value of 23.1—measured in the undamaged specimen (before any mechanical load is applied)—is a measure of the intrinsic nonlinearity of the undamaged IN100 material; and that the nonlinearity associated with the transmitting piezoelectric transducers (β=9.0) has already been subtracted from this and all other values. Figure 4 shows that there is a significant increase in β with increasing plastic stress; the increase is largest from the unloaded (undamaged) state to 125% yield stress, and then the increase is less substantial at the higher stresses. This observed behavior of a large increase in the acoustic nonlinear parameter, once the specimen is loaded above its yield stress, makes sense because dislocations (or microplasticity) create significant material nonlinearity; the literature reports that the second- (order) harmonic amplitudes associated with dislocations should be larger than the intrinsic material nonlinearity due to the elastic lattice anharmonicity.[25] Most importantly, the results in Fig. 4 show that the proposed measurement procedure is capable of making an absolute measurement of the evolution of the acoustic nonlinear parameter, β (as a function of stress in this case) in these interrupted mechanical tests on an IN100 specimen.

B. Low-cycle fatigue results

Low-cycle fatigue in this paper refers to a fatigue test where the maximum stress is above yield, so there is plastic deformation even at the beginning of the fatigue test. Of equal importance is that cyclic loading promotes the formation of dislocation dipoles, which is the strongest source of nonlinearity among a list of potential sources.[6,9,10] The frequency of cyclic loading is 0.5 Hz, $R(=\sigma_{min}/\sigma_{max})$ is zero (strain controlled), the maximum stress level is 105% of the yield stress (strain of 0.48%), and the fatigue tests are interrupted to perform the nonlinear ultrasonic measurements at different numbers of fatigue cycles. Three different specimens are tested, and there will be some level of variability associated with the initial microstructure of each specimen. As a result, the measured acoustic nonlinearity parameters will be normalized by the value measured in each undamaged specimen (β_0), before any mechanical load is applied. This normalization procedure (which will be repeated for the high-cycle fatigue results) removes some of the variability associated with the initial microstructures of each specimen, enables a direct comparison of the evolution of the acoustic nonlinearity of all the specimens tested, and normalizes the nonlinearity associated with the transmitting piezoelectric transducers. The evolution of the normalized acoustic nonlinearity parameter (β/β_0), as a function of normalized fatigue life (fatigue cycle normalized to the total number of cycles, where 100% means the total fatigue life), together with a best-fit curve, is shown in Fig. 5. Note that the specimens failed at 12,640, 13,012, and 50,221 cycles (the third specimen is the standard fatigue specimen; while the first two are the nonstandard rectangular bars) and the (calibrated) β_0 measured in each specimen is 21.4, 22.2, and 22.4, respectively.

Figure 5 shows a rapid increase in β/β_0 (up to 30%) during the first 40% of fatigue life, which demonstrates that these nonlinear ultrasonic measurements can be used to quantitatively characterize the damage state of this material

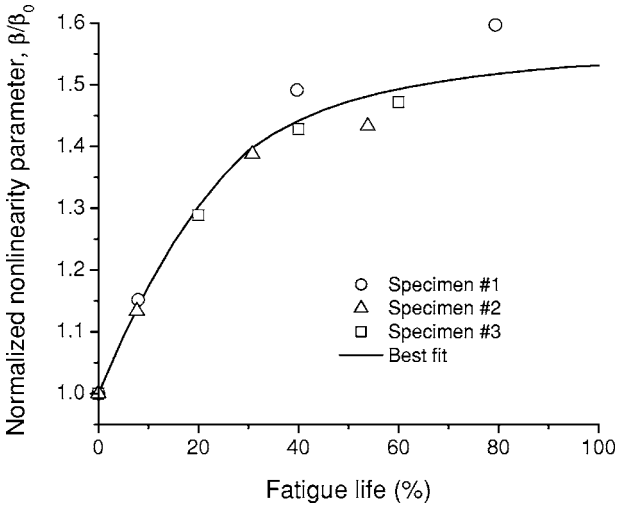


FIG. 5. Low-cycle fatigue results—normalized acoustic nonlinearity parameter β/β_0 as a function of the percentage of fatigue life for three different fatigue specimens. The continuous line is the best-fit curve obtained from the discrete experimental data.

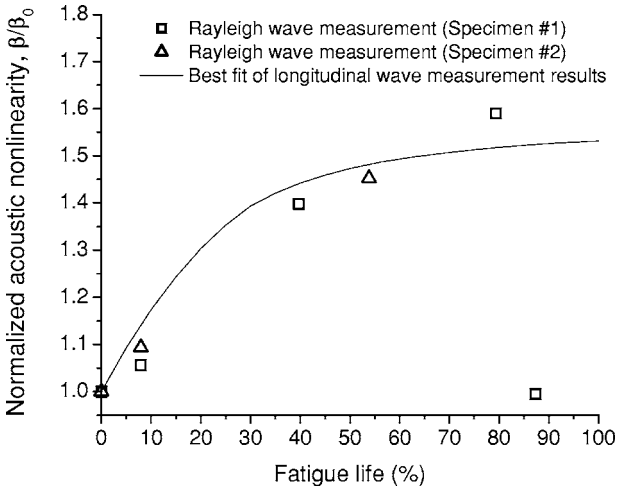


FIG. 6. Comparison of longitudinal and Rayleigh wave (Ref. 24) results showing the normalized acoustic nonlinearity parameter β/β_0 as a function of the percentage of fatigue life for the low-cycle fatigue results.

in the early stages of fatigue life. This is somewhat different from other experimental results,[5] which show a slower initial increase in β. The difference in behavior is most likely due to the high maximum stress (strain) level beyond the yield stress (strain), and a significant amount of plasticity—due to the dislocation motions—probably starts accumulating in the specimen from the first loading cycle, which reduces the time period for dislocation reassociation at the beginning of fatigue. The measurement data show increasing scatter with increasing number fatigue cycles, which is most likely due to a combination of two factors: The intrinsic material behavior, and issues with the measurement procedure. There is an inherent randomness in the progression of fatigue damage during fatigue testing (more so in high-cycle fatigue, as discussed in the next section), which should manifest itself as a corresponding randomness in the resulting acoustic nonlinearity. There is a somewhat unrelated issue with the measurement procedure in the later stages of fatigue—the surface deformation associated with the increased plasticity makes it difficult to consistently couple the transducers to the specimen surface. Finally, note that a best-fit curve, such as the one developed in Fig. 5 (but based on a larger number of specimen and data points), has the potential to serve as a master curve for life prediction based on nonlinear ultrasonic measurements.

A companion study makes nonlinear ultrasonic measurements with Rayleigh surface waves on the first two specimens; the procedure used to make these (relative) nonlinear Rayleigh wave measurements is reported elsewhere.[24] Figure 6 shows a comparison of the best-fit curve from Fig. 5 (longitudinal waves) with those from the nonlinear Rayleigh wave measurements. There is excellent agreement with these two sets of results, demonstrating that both longitudinal and Rayleigh waves can be used to track nonlinear material behavior. Note that the sharp drop in the acoustic nonlinearity (the single data point in Fig. 6), at 87% of fatigue life for Specimen No. 1, is most likely due to the emergence of surface-breaking microcracks whose depths are larger than the wavelength of the second-harmonic.[24]

C. High-cycle fatigue results

High-cycle fatigue in this study refers to fatigue tests where the maximum stress level is below yield; in this case, the maximum stress is 95% of yield stress, the frequency of cyclic loading is 1 Hz, and R $(=\sigma_{\min}/\sigma_{\max})$ is zero (load controlled). Five different fatigue specimens are tested, with failure occurring at 55,432, 102,392, 203,220, 328,341, and 350,985 cycles. As with the monotonic and low-cycle fatigue tests, these fatigue tests are interrupted to perform the nonlinear ultrasonic tests. Figure 7 shows the change in the normalized acoustic nonlinearity parameter (β/β_0) over the normalized fatigue life of each specimen. Note that the (calibrated) β_0 measured in each specimen is 22.1, 21.3, 19.8, 22.4, and 21.2, respectively.

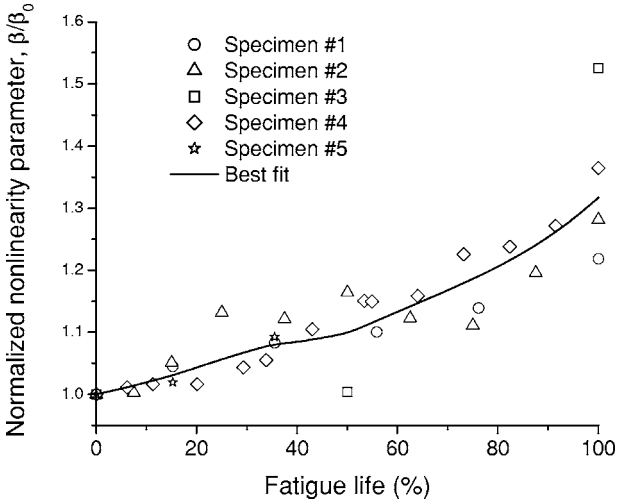


FIG. 7. High-cycle fatigue results—normalized acoustic nonlinearity parameter β/β_0 as a function of the percentage of fatigue life for five different fatigue specimens. The continuous line is the best-fit curve obtained from the discrete experimental data.

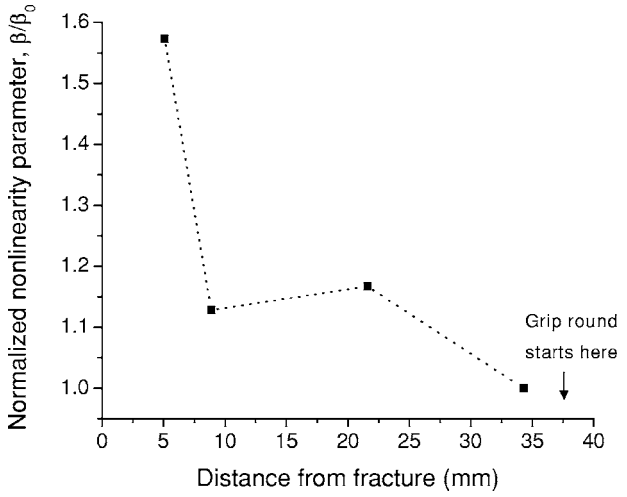


FIG. 8. Spatial variation of the normalized acoustic nonlinearity parameter on one failed high-cycle fatigue specimen, shown as a function of distance from the fracture site.

The critical observations from Fig. 7 are: The absolute increase in acoustic nonlinearity is much lower than that of the low-cycle fatigue tests; there is much more scatter in the high-cycle fatigue data (a larger variability at a given percentage of fatigue life than the data from the low-cycle fatigue tests); and the overall increase in acoustic nonlinearity is more gradual in the high-cycle specimens. Some of these trends may be inherent to the high-cycle fatigue behavior of this material since fatigue damage during high-cycle fatigue is more localized than that during low-cycle fatigue. So it is not surprising that there is a corresponding variability in the acoustic nonlinearity. Similar results were obtained in a titanium alloy (Ti-6-4) by Frouin *et al.*[8] A full explanation of this behavior may only be possible by coupling these nonlinear ultrasonic results with a material model that predicts the formation of dislocation substructures in this material under different loading conditions. For Specimen No. 5, additional nonlinear measurements beyond 35% of fatigue life were not possible due to severe surface distortion, but the fatigue test was continued to obtain its fatigue life (55,432).

Nonlinear ultrasonic evidence of this localized damage behavior in high-cycle fatigue is given in Fig. 8 by plotting the spatial variation (four discrete locations) of the acoustic nonlinearity in the gauge section of a failed high-cycle fatigue specimen (the specimen with a uniform cross section). Note that each acoustic nonlinearity parameter measured with this experimental procedure is representative of the cylindrical volume of material under the transducer (12.7 mm diameter for these tests). Figure 8 shows that the acoustic nonlinearity, at a location about 5 mm away from the fracture surface, is approximately 58% higher than those in the region near the grip. It is also shown that the fatigue damage in this material is quite localized (on the order of 10 mm), which could be the source of the large variation in the later stages of the high-cycle fatigue tests. It is important to note that this level of spatial variation in the acoustic nonlinearity parameter (and thus, localized damage) is primarily present in the high-cycle fatigue specimens; and for these specimens, mainly in the later stages of fatigue life. The monotonic and low-cycle fatigue specimens do not show this level of spatial variation in the measured acoustic nonlinearity parameter. These values are very consistent, regardless of where they are measured in the gauge section, so the damage in the monotonic and low-cycle fatigue specimens seems to be more evenly distributed.

V. CONCLUSION

This research develops a robust experimental procedure to track the evolution of fatigue damage in a nickel-base superalloy with the acoustic nonlinearity parameter, β, and demonstrates its effectiveness by making repeatable measurements of β on multiple specimens subjected to both high- and low-cycle fatigue. The measurement procedure developed in this research is robust in that it is based on conventional piezoelectric contact transducers—which are readily available off the shelf—and offers the potential to be used in field applications. In addition, it allows the user to isolate sample nonlinearity from measurement system nonlinearity. The experimental results show that there is a significant increase in β associated with the high plasticity of low-cycle fatigue, and illustrate how these nonlinear ultrasonic measurements quantitatively characterize the damage state of a specimen in the early stages of fatigue. A companion study with Rayleigh surface waves demonstrates that both longitudinal and Rayleigh waves can be used to track nonlinear material behavior. The high-cycle fatigue results are less definitive (the increase in β is not as substantial), but still show a clear relationship between β and remaining fatigue life. Overall, these nonlinear ultrasonic results indicate that the fatigue damage associated with high-cycle fatigue is relatively localized, while the damage associated with low-cycle fatigue is more evenly distributed. One application of the measured β versus fatigue life data is to potentially serve as a master curve for life prediction based on nonlinear ultrasonic measurements.

ACKNOWLEDGMENTS

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA), Defense Sciences Office (DSO), Engine Systems Prognosis—under a subcontract to Prime Contract No. HR0011-04-C-0001, and is approved for public release with unlimited distribution. The authors thank Jeff Davies, Jan Herrmann, and Marc Erwin, for their help with the measurements, and Professor Peter Nagy for his advice and the use of his Ritec RAM-10000 system.

[1]W. T. Yost and J. H. Cantrell, "The effects of fatigue on acoustic nonlinearity in aluminum alloys," Proc. IEEE **2**, 947–954 (1992).
[2]J. K. Na, J. H. Cantrell, and W. T. Yost, "Linear and nonlinear properties of fatigued 410Cb stainless steel," Rev. Prog. Quant. Nondestr. Eval. **15**, 1479–1488 (1996).
[3]D. J. Barnard, G. E. Dace, and O. Buck, "Acoustic harmonic generation due to thermal embrittlement of Inconel 718," J. Nondestruct. Eval. **16**, 67–75 (1997).
[4]P. B. Nagy, "Fatigue damage assessment by nonlinear ultrasonic material characterization," Ultrasonics **36**, 375–381 (1998).
[5]J. Frouin, S. Sathish, T. E. Matikas, and J. K. Na, "Ultrasonic linear and nonlinear behavior of fatigued Ti-6Al-4V," J. Mater. Res. **14**, 1295–1298 (1999).

[6]J. H. Cantrell and W. T. Yost, "Nonlinear ultrasonic characterization of fatigue microstructures," Int. J. Fatigue **23**, S487–S490 (2001).
[7]Y. L. Hinton, J. K. Na, W. T. Yost, and G. L. Kissel, "Field measurement of the acoustic nonlinearity parameter in turbine blades," NASA Technical Memorandum, TM-2000-210303 ⟨http://library-dspace.larc.nasa.gov/dspace/jsp/bitstream/ 2002/13099/1⟩ (2000).
[8]J. Frouin, T. E. Matikas, J. K. Na, and S. Sathish, "*In situ* monitoring of acoustic linear and nonlinear behavior of titanium alloys during cyclic loading," Proc. SPIE **3585**, 107–116 (1999).
[9]J. H. Cantrell, "Substructural organization, dislocation plasticity, and harmonic generation in cyclically stressed wavy slip metals," Proc. R. Soc. London, Ser. A **460**, 757–780 (2004).
[10]J. H. Cantrell, "Quantitative assessment of fatigue damage accumulation in wavy slip metals from acoustic harmonic generation," Philos. Mag. **86**, 1539–1554 (2006).
[11]M. A. Breazeale and J. Philip, "Determination of third-order elastic constants from ultrasonic harmonic generation measurements," Phys. Acoust. **17**, 1–60 (1984).
[12]J.-Y. Kim, J. Qu, L. J. Jacobs, J. Littles, and M. F. Savage, "Acoustic nonlinearity parameters due to microplasticity," J. Nondestruct. Eval. **25** (2006).
[13]R. E. Green, Jr., "Ultrasonic investigation of mechanical properties," *Treatise on Materials Science and Technology* (Academic, New York, 1973), Vol. **3**.
[14]J. H. Cantrell, Jr., "Acoustic-radiation stress in solids. I. Theory," Phys. Rev. B **30**, 3214–3220 (1984).
[15]G. E. Dace, R. B. Thompson, and O. Buck, "Measurement of the acoustic harmonic generation for materials characterization using contact transducers," Rev. Prog. Quant. Nondestr. Eval. **11**, 2069–2076 (1992).
[16]Y. Ohara, K. Kawashima, R. Yamada, and H. Horio, "Evaluation of amorphous diffusion bonding by nonlinear ultrasonic method," Rev. Prog. Quant. Nondestr. Eval. **23**, 944–951 (2004).
[17]S. Krishnan, J. D. Hamilton, and M. O'Donnell, "Suppression of propagating second harmonic in ultrasound contrast imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 704–711 (1998).
[18]T. Mueller, "Characterization of microstructure evolution of nickel-base superalloys using nonlinear ultrasonic techniques." M.S. thesis, School of Civil and Environmental Engineering, Georgia Institute of Technology, 2005.
[19]W. T. Yost and J. H. Cantrell, Jr., "Acoustic-radiation stress in solids. II: Experiment," Phys. Rev. B **30**, 3221–3227 (1984).
[20]D. J. Barnard, "Variation of nonlinearity parameter at low fundamental amplitudes," Appl. Phys. Lett. **74**, 2447–2449 (1999).
[21]B. B. Chick and M. McKenna (RITEC Corporation, private communication, 2005).
[22]K. K. Na and M. A. Breazeale, "Ultrasonic nonlinear properties of lead zicronate-titanate ceramics," J. Acoust. Soc. Am. **95**, 3213–3221 (1994).
[23]D. C. Hurley and C. M. Fortunko, "Determination of the nonlinear ultrasonic parameter β using a Michelson interferometer," Meas. Sci. Technol. **8**, 634–642 (1997).
[24]J. Herrmann, J.-Y. Kim, L. J. Jacobs, J. Qu, J. W. Littles, and M. F. Savage, "Assessment of material damage in a nickel-base superalloy using nonlinear Rayleigh surface waves," J. Appl. Phys. **99**, 124913 (2006).
[25]A. Hikata and C. Elbaum, "Generation of ultrasonic second and third harmonics due to dislocations," Phys. Rev. **144**, 469–477 (1966).

Diffraction of sound due to moving sources by barriers and ground discontinuities

M. Buret and K. M. Li[a)]
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

K. Attenborough
Department of Engineering, The University of Hull, Cottingham Road, Hull HU6 7RX, United Kingdom

(Received 2 December 2005; revised 14 June 2006; accepted 14 June 2006)

The solution for diffraction of sound by a wedge is extended to sources in uniform motion parallel to the obstacle. By means of an auxiliary transformation the problem is reduced to that of a set of stationary sources so that the resulting solution is in accordance with well-known models. Applications to prediction and abatement of transportation noise are considered. The formulation for the diffracted wave is combined with the solution for ground reflection of sound due to moving sources. New models for the sound field due to a source in motion along a barrier above the ground and for a source moving parallel to an impedance discontinuity are derived. In both situations, greater sensitivity to motion of the resulting sound pressure levels is found on source approach. However, attenuation of sound from moving sources by barriers is found to be so little affected by motion that design schemes for stationary source are relevant. On the other hand, it is noted that noise predictions must account for source motion. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2221535]

PACS number(s): 43.28.Js, 43.28.Fp, 43.20.Dk, 43.20.El [VEO] Pages: 1274–1283

I. INTRODUCTION

Diffraction by obstacles is a common problem in wave physics and fundamental solutions in acoustics have been derived by borrowing from other fields such as electromagnetics.[1] Since they are a traditional means for noise abatement, an obvious acoustical application is to noise barriers.

Although simplified schemes[2,3] are generally considered sufficient for predicting noise barrier efficiency, their scope is usually limited to basic situations such as stationary omnidirectional point sources in the presence of nonabsorptive surfaces. In more complex cases, although semiempirical approaches can lead to analytical models,[4] it is necessary to consider the theoretical foundation of the problem, particularly when higher accuracy is sought.

Pierce[5] has treated the classic problem of sound wave diffraction by wedges. In particular he relates the solution to the general context of the geometrical theory of diffraction. This solution has been extended to situations relevant to environmental noise predictions and abatement such as wide barriers,[5] barriers above impedance ground and impedance discontinuities,[6] and the diffraction of sound by absorbing barriers.[7] However, except for a recent formulation for diffraction of sound due to dipoles[8] and a unifying model for point sources, point source arrays and line sources,[9] few works have focused on extension of the models at the source end.

In particular, to the authors' knowledge, little can be found on the diffraction of sound due to moving sources. Noise prediction schemes typically neglect the Doppler and convection effects related to source motion.[10] For barriers in urban areas, involving high traffic density and speed restrictions, it possible to neglect the Doppler and convection effects when formulating the expressions for the sound field. However, for high speed trains, which can reach speeds beyond 400 km/h, these assumptions are less acceptable.

In this context, an expression—hereinafter referred to as the *Doppler-Weyl-Van der Pol formula*—was recently introduced by Buret *et al.*[11,12] for the sound field due to a source at high speed motion above the ground. Following Morse and Ingard's method[13,14] for deriving the sound field due to a source in uniform motion, use of an auxiliary coordinate system in well-known expressions for stationary sources, enabled the problem to be tackled from the governing equations. It was shown that, although moderate, the effects of source motion were significant even for relatively low Mach numbers.

In this paper, a similar approach is used to derive the expression for the sound wave due to a source in motion in the presence of a barrier—or a diffracting element, extending Pierce's expression.[5] The derivation of this fundamental solution is presented in Sec. II. Section III is dedicated to applications of the new model to situations relevant to environmental acoustics. First, the effects of motion on barrier attenuation are discussed; the sound field due to a moving source in the presence of a hard screen above arbitrary ground is then derived. Eventually, De Jong *et al.*'s model for propagation of sound in the presence of a ground discontinuity[6] is extended to account for source motion.

[a)]Corresponding author. Present address: Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, 140 S. Intramural Drive, West Lafayette, IN 47907-2031. Electronic mail: mmkmli@purdue.edu

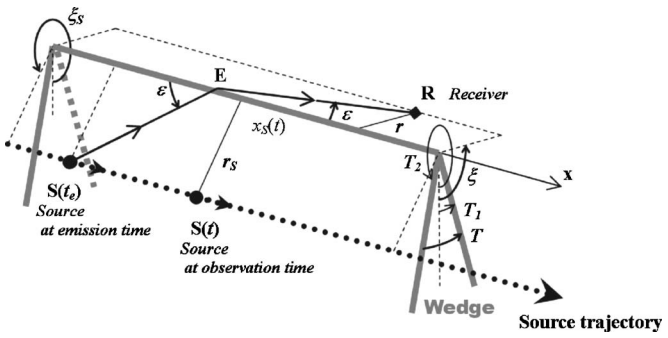


FIG. 1. Geometry for a source S in uniform motion along a semi-infinite wedge of top angle T. The x axis coincides with the edge of the wedge and the cylindrical coordinates of the source $[\xi_S, r_S, x_S(t)]$ and the motionless receiver R located at (ξ, r, x) are explicated. The diffracted ray is represented by the broken segment $S(t_e)$-E-R.

II. DIFFRACTION OF SOUND DUE TO A POINT SOURCE IN UNIFORM MOTION ALONG A RIGID WEDGE

A. Kinematics

Consider a semi-infinite rigid wedge with constant cross section in the vertical plane y-z. Its faces are oriented at arbitrary angles T_1 and T_2 ($T_1 < T_2$) with respect to the latter (see Fig. 1). The top angle of the wedge is $T = 2\pi + T_1 - T_2$ and takes the value $2T_1$ in the particular case of a symmetrical wedge. Let S be a sound source moving along the x axis, hence, in parallel to the edge of the wedge. A cylindrical coordinate system (ξ, r, x) attached to a motionless receiver can be introduced, in which $\mathbf{r_s} \equiv (\xi_S, r_S, x_S)$ represents the source coordinates. For uniform source motion at Mach number M, the only varying parameter is the source-receiver offset x-x_S. Let the initial position of the source be located at $x=0$ and c_0 be the speed of sound; then

$$x_S(t) = c_0 M t. \tag{1}$$

The projection of the geometry in the cross-sectional plane y-z remains unchanged with source motion and the analogy with a stationary source and/or a two-dimensional situation is then obvious. According to the geometrical theory of diffraction, the diffracted sound ray is a segment linking the source S to the receiver R via the edge of the wedge E. The diffracted path length is then $L = \overline{SE} + \overline{ER}$ The diffraction angle ε is the same on both source and receiver side.[5] Let the receiver be motionless and located at cylindrical coordinates $(\xi, r, 0)$; then

$$\cos\varepsilon = -\frac{c_0 M t_e}{L}, \quad \sin\varepsilon = \frac{r + r_s}{L}, \quad x_E = -\frac{r}{\tan\varepsilon}, \tag{2}$$

where t_e is the time of emission of the diffracted wave reaching the receiver at time t, L is the total length of the diffracted ray, and x_E defines the position E at which the diffracted ray intersects the edge of the wedge. L and t_e are the solutions of

$$c_0(t - t_e) = L = \sqrt{(r + r_S)^2 + (c_0 M t_e)^2}, \tag{3}$$

which resolves as

$$t_e = \frac{c_0 - \Lambda}{c_0(1 - M^2)}, \quad L = \frac{\Lambda - M(c_0 M t)}{1 - M^2}, \tag{4a}$$

$$\Lambda = \sqrt{(c_0 M t)^2 + (1 - M^2)(r + r_S)^2} = L(1 - M_E), \tag{4b}$$

where $M_E = M\cos\varepsilon$ is the component of the Mach number along the diffracted ray.

B. Diffracted pressure for a source in uniform motion

In the case of a stationary point source, the pressure field in the vicinity of a wedge can be expressed as the sum of up to four components:[5] a direct pressure field p_{dir}, two reflections on the faces of the wedge $p_{\mathrm{refl},1}$ and $p_{\mathrm{refl},2}$, and a diffracted wave p_{diffr}. The three first contributions may or may not occur, depending on the relative source and receiver positions with respect to the wedge that may screen the sound source or its images from the receiver[5,8]

$$p = U[\pi - |\xi - \xi_S|]p_{\mathrm{dir}} + U[\pi + 2T_1 - (\xi + \xi_S)]p_{\mathrm{refl},1} + U[(\xi + \xi_S) + \pi - 2T_2]p_{\mathrm{refl},2} + p_{\mathrm{diffr}}. \tag{5}$$

In Eq. (5), the Heaviside step function U ($U[x]=1$ if $x>0$; $U[x]=0$ if $x\leqslant 0$) determines the presence or absence of the direct and reflected waves. Remarkably, the arguments of U—and thence the composition of the sound field—depend only on the angular coordinates of the source, the receiver, and the faces of the wedge in the cross-sectional plane.

In the case of a source moving parallel to the edge of a wedge, the geometry in this plane remains constant and therefore the relationships determining the presence or absence of the direct and reflected pressure waves remain the same as in Eq. (5). When present, the direct wave is expressed as the pressure due to the source in motion in the free field.[13,14] The potential contributions due to the reflected waves can be formulated considering either a formulation for reflection of the sound field due to a source in motion above a plane,[11] or the acoustic pressure field due to the relevant image sources with respect to the faces of the wedge radiating in the absence of obstacle. Derivation of these contributions is straightforward and will not be detailed here for brevity, and because in practical applications (noise barriers), the sound field is composed of the diffracted wave only.

Morse and Ingard have derived the sound field due to a monopole in uniform motion in the free field by means of an auxiliary transformation, analogous to a Lorentz transformation.[13] Buret *et al.*[11] have used the same transformation to formulate the Doppler-Weyl-Van der Pol formula for a source moving above an impedance ground. The coordinates and time (x_L, y_L, z_L, t_L) in the auxiliary coordinate system are

$$\begin{aligned} x_L &= \gamma^2(x - c_0 M t), \\ y_L &= \gamma y, \\ z_L &= \gamma z, \\ t_L &= \gamma^2(t - Mx/c_0), \end{aligned} \quad \text{with } \gamma = 1/\sqrt{1 - M^2}. \tag{6}$$

Hereinafter, subscript L characterizes corresponding variables in the auxiliary coordinate system. In the case of uniform source motion, the wave equation for the velocity po-

tential field φ takes the same form as that for a stationary monopole when expressed in the auxiliary coordinate system.[13] The acoustic pressure is then calculated as the time derivative of the velocity potential[13] $p=\rho\partial\varphi/\partial t$ where ρ is the air density. Noting that

$$\frac{\partial}{\partial t}=\gamma^2\frac{\partial}{\partial t_L}-\gamma^2 c_0 M\frac{\partial}{\partial x_L}, \tag{7}$$

it is shown that the acoustic pressure field can be expressed as the sum of contributions from two stationary sources: a monopole and a dipole oriented in the direction of source motion.[12] Denoting these relative contributions by $p^{(0)}$ and $p^{(1)}$, respectively, the total pressure p due to a harmonic source of strength P_0 can be expressed as[11]

$$p=P_0\left(\gamma^4 p^{(0)}-\gamma^4\frac{M}{ik_0}p^{(1)}\right), \tag{8}$$

where k_0 is the wave number, $k_0=\omega/c_0$, ω being the angular frequency of the source.

With constant Mach number, the auxiliary transformation involves uniform expansion along the three dimensions, as well as a time dependent translation along the direction x of source motion. As a result, the transformed geometry is analogous to that for a stationary source in the original coordinate system. A cylindrical coordinate system (ξ_L, r_L, x_L) is introduced, relating to that attached to the motionless receiver by

$$\xi_L=\xi,\quad r_L=\gamma r,\quad x_L=\gamma^2(x-c_0 Mt). \tag{9}$$

Remarkably, the polar angle is unaffected by the transformation. As a result, the top angle of the wedge has same value in both the auxiliary and original coordinate system. Using Pierce's formulation,[5,15] the expression for the monopole component of the diffracted wave is

$$p^{(0)}=\frac{\exp[-i\omega(t_L-L_L/c_0)]}{4\pi L_L}\frac{e^{i\pi/4}}{\sqrt{2}}[A_D(X_{L+})+A_D(X_{L-})], \tag{10a}$$

where L_L is the diffracted path in the transformed coordinates and A_D is the diffraction integral

$$A_D(X_{L\pm})=\frac{1}{\pi\sqrt{2}}\int_{-\infty}^{\infty}\frac{e^{-u^2}}{(\pi/2)^{1/2}X_{L\pm}-e^{-i\pi/4}u}du, \tag{10b}$$

$$X_{L\pm}=\Gamma_L N_\nu(\alpha_\pm),\quad \alpha_-=\xi+\xi_S-2T_1,\quad \alpha_+=\xi-\xi_S, \tag{10c}$$

$$\Gamma_L=\sqrt{k_0 r_L\rho_S/\pi L_L},\quad N_\nu(\alpha_\pm)=\frac{\cos\nu\pi-\cos\nu\alpha_\pm}{\nu\sin\nu\pi}, \tag{10d}$$

where $\nu=\pi/(2\pi-T)$ is the wedge index and $\rho_S=\gamma r_S$.

The expression for the three-dimensional (3D) dipole field diffracted by a wedge has been derived and has been validated by laboratory measurements elsewhere.[8] For a dipole lying along the x_L axis, the sound field can be expressed using[8,16]

$$p^{(1)}=\frac{\exp[-i\omega(t_L-L_L/c_0)]}{4\pi L_L}\frac{\cos\varepsilon_L}{L_L}\frac{e^{i\pi/4}}{\sqrt{2}}\Bigg\{(1-ik_0L_L)\times[A_D(X_{L+})+A_D(X_{L-})]+\frac{i}{2}[X_{L+}(1-\pi X_{L+}A_D(X_{L+}))+X_{L-}(1-\pi X_{L-}A_D(X_{L-}))]\Bigg\}. \tag{11}$$

Both X_{L^+} and X_{L^-}take small values for long diffraction paths and high frequencies. As a result, the second term inside the curly brackets in Eq. (11) can be neglected.[8] This approximation is particularly applicable for source motion in parallel to the edge of the wedge, as the diffracted path length increases with the time-dependent source-receiver offset x-x_S. The dipole component of the pressure then takes a form similar to the monopole pressure given in Eq. (10), except for the directivity coefficient $\cos\varepsilon_L$ and the strength factor $(1-ik_0L)/L_L$,

$$p^{(1)}=\frac{\exp[-i\omega(t_L-L_L/c_0)]}{4\pi L_L}\frac{\cos\varepsilon_L}{L_L}\frac{e^{i\pi/4}}{\sqrt{2}}(1-ik_0L_L)\times[A_D(X_{L+})+A_D(X_{L-})]. \tag{12}$$

The total diffracted pressure is then expressed by substituting Eqs. (10) and (12) into Eq. (8),

$$p=\gamma^4 P_0\frac{\exp[-i\omega t_L]}{4\pi}\frac{\exp(ik_0L_L)}{L_L}\frac{e^{i\pi/4}}{\sqrt{2}}\times\left[1+M\cos\varepsilon_L\left(1-\frac{1}{ik_0L_L}\right)\right][A_D(X_{L+})+A_D(X_{L-})]. \tag{13}$$

Transforming back into the coordinate and time system (x,y,z,t) attached to the receiver is straightforward, noting that

$$L_L=\sqrt{(\gamma r+\gamma r_S)^2+x_L^2}=\gamma^2 L(1-M\cos\varepsilon), \tag{14a}$$

$$t_L-\frac{L_L}{c_0}=t_e=t-\frac{L}{c_0}, \tag{14b}$$

and

$$\cos\varepsilon_L=\frac{x_L}{L_L}=\frac{\cos\varepsilon-M}{1-M\cos\varepsilon}. \tag{14c}$$

The diffracted acoustic pressure wave due to a monopole in uniform motion in parallel to a rigid wedge is represented by

$$p=P_0\frac{\exp(-i\omega t)}{4\pi}\frac{\exp(ik_0L)}{L(1-M\cos\varepsilon)^2}\times\left\{\frac{e^{i\pi/4}}{\sqrt{2}}\left[1-\frac{1}{ik_0L}\frac{M\cos\varepsilon-M^2}{1-M\cos\varepsilon}\right]\left[A_D\left(\frac{X_+}{\sqrt{1-M\cos\varepsilon}}\right)+A_D\left(\frac{X_-}{\sqrt{1-M\cos\varepsilon}}\right)\right]\right\}, \tag{15}$$

where A_D is defined as in Eq. (10b) with

$$X_\pm=\Gamma N_\nu(\alpha_\pm),\quad \Gamma=\sqrt{k_0 r r_S/\pi L}. \tag{16}$$

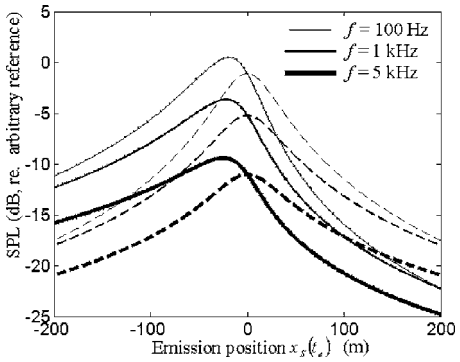


FIG. 2. Diffracted pressure for a source in uniform motion at Mach 0.3 along a half plane at discrete frequencies 100 Hz (thin lines), 1 kHz (thick lines), and 5 kHz (thicker lines). Coinciding solid and dotted lines of the same thickness correspond to the approximate solution in Eq. (15) and the full expression with dipole component expressed using Eq. (11), respectively. Dashed lines correspond to predictions for the corresponding stationary source at position of emission. The source trajectory is 1 m below the edge, with horizontal separation 10 m; the receiver is 0.8 m below the edge at distance 20 m.

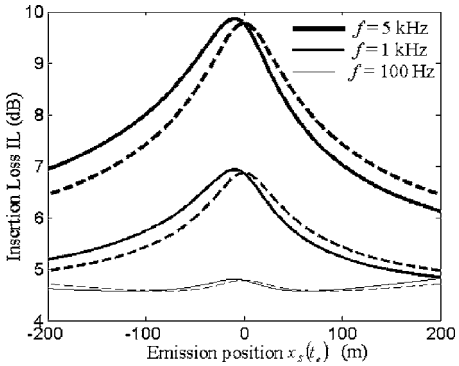


FIG. 3. IL for the same source and obstacle as for Fig. 2, for discrete frequencies 100 Hz (thin lines), 1 kHz (thick lines), and 5 kHz (thicker lines). Solid lines show the IL for a source moving at Mach number 0.3, dashed lines show the IL for the corresponding stationary source.

In Fig. 2, the sound pressure level (SPL) for the diffracted wave is plotted for discrete frequencies 100 Hz, 1 and 5 kHz, for a point source moving at Mach number 0.3 ($\sim$360 km/h which is a plausible speed for a high train, for example) in the vicinity of a semi-infinite rigid half plane ($T=0$; $\nu=1/2$) with source, receiver, and obstacle edge at heights 1, 1.2, and 2 m, respectively. The shortest horizontal separation from the obstacle to the source is 10 m and that to the receiver on the other side is 20 m. Coincident solid and dotted lines show, respectively, the results of calculations by means of the approximation in Eq. (15) and using the full dipole component given in Eq. (11). Such coincidence validates the approximation used to derive Eq. (15) along the whole source trajectory. Dashed lines show the SPL calculated for a *corresponding stationary source*—i.e., a stationary monopole located at the position $x_S(t_e)$ of emission. Comparison with the predictions for the moving source enables an assessment of the effects of source motion other than the geometrical change in location of the source position. The SPL is higher than for a stationary source when the source approaches the receiver $[x_S(t_e) \leqslant 0]$, and lower when it is receding $[x_S(t_e) \geqslant 0]$.

It should be noted that the maximum in SPL is predicted to occur when the source is approaching the receiver, hence, before the point of closest source-receiver approach. These effects are directly related to the Doppler factor $(1-M\cos\varepsilon)^{-1}$ and the convection coefficient $(1-M\cos\varepsilon)^{-2}$ in Eq. (15). Both are larger than 1 on approach—leading to higher pressure amplitude [cf. Eq. (15)]—and smaller than 1 on recession resulting in lower SPL. The shift of the peak towards the approaching source position is due to the nonlinearity (and, hence, nonsymmetry) of the logarithm function used to express the SPL in decibels (dB) while $(1-M\cos\varepsilon)$ shows symmetry around the value 1. These observations are consistent with previous results for a source moving in the absence of obstacles[13] for which similar observations can be made in accordance to variations in the Doppler factor around 1.

III. APPLICATIONS TO OUTDOOR ACOUSTICS

A. Effect of source motion on barrier attenuation

Barrier attenuation is typically characterized by the insertion loss (IL): the ratio in dB of the total pressure wave perceived at the receiver [Eq. (15)] to the direct sound field p_{dir} that would occur in the absence of any obstacle

$$\mathrm{IL} = -10\log[|p(t)/p_{\mathrm{dir}}(t)|]. \tag{17a}$$

In the case of a source in motion[11,13,14]

$$p_{\mathrm{dir}}(t) = P_0\frac{\exp(-i\omega t)}{4\pi}\frac{\exp(ik_0R)}{R(1-M\cos\psi\sin\phi)^2} \times\left[1-\frac{M}{ik_0R}\left(\frac{\cos\psi\sin\phi-M}{1-M\cos\psi\sin\phi}\right)\right], \tag{17b}$$

where R is the length of the direct source-receiver path, ψ and ϕ are its azimuthal and elevation angles, such that $\cos\psi\sin\phi=-x_S(t_{\mathrm{dir}})/R$, with $t_{\mathrm{dir}}=t-R/c_0$ the emission time of the direct wave. Due to the path difference between the direct and the diffracted wave the latter has been emitted earlier: $t_e<t_{\mathrm{dir}}$.

Figure 3 shows the IL predicted for the same source and receiver locations as in Fig. 2 and calculated for the same discrete frequencies. The solid lines show the IL predicted for a source moving at Mach number 0.3, whereas dotted lines show the IL predicted for the corresponding stationary source. The IL predicted for the moving source is within only 1 dB of that predicted for a stationary source. Recalling the expression for the diffracted wave given in Eq. (15), we point out that the magnitude of the IL depends strongly on the ratio of the convection factors $(1-M\cos\varepsilon)^2/(1-M\cos\psi\sin\phi)^2$. The latter is close to 1 when the path difference between direct and diffracted pressures is not large $(L/R\approx 1)$, which is the case for large source offset, for example. The expression for the magnitude of the IL due to a source in motion is then close to that for the corresponding stationary source, except for the Doppler factor in the argument of the diffraction integrals A_D. The general trend, showed clearly in Fig. 3 for 1 and 5 kHz is explained by the fact that the IL is predicted to increase with the argument of decreasing function A_D. When the source is approaching, the Doppler factor $(1-M\cos\varepsilon)^{-1}$ is greater than 1 and the pre-

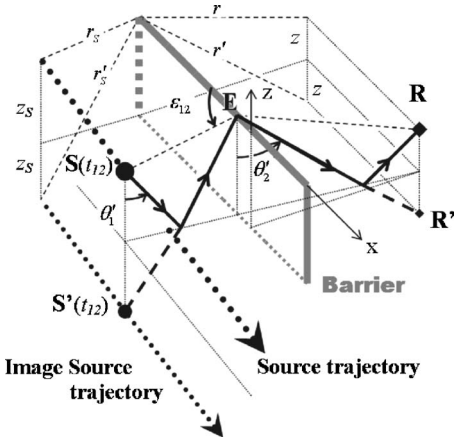


FIG. 4. Geometry for a source in uniform motion in parallel to a thin barrier above the ground. The diffraction path "12" with reflection on both sides of the barrier is represented by thick lines.

dicted insertion loss for a source in motion will be larger than for a stationary source located at the point of emission $x_S(t_e)$. This trend is reversed when the source recedes from the receiver, with smaller predicted IL than in the motionless case.

The predicted influences of the Doppler and convection factors on the barrier attenuation are small and the effects of motion on the IL are mainly related to the change in source position over the trajectory. It can be concluded therefore that usual schemes to estimate barrier attenuation for stationary sources—such as Maekawa's[3]—can apply to screen design in situations involving sources in motion, particularly at low speed.

On the other hand, for the sake of accuracy, noise predictions must account for source displacement and convection in order to properly describe the absolute pressure field. As shown in Fig. 2, the maximum in predicted SPL occurs before the source reaches the shortest separation to the receiver, hence, the resulting sound field is somewhat affected by source motion and it is necessary to account for the convection and Doppler effects in its formulation. For typical outdoor situations it is important to allow for the presence of ground around the barrier. This is the subject of the next section. A further related analysis of the effects of ground discontinuity on the sound field of a moving source follows this.

B. Barrier in the presence of ground

Sound propagation around a rigid barrier on the ground is a classical problem. In the case of hard ground on both sides of the barrier, the solution may be obtained in a straightforward manner by means of mirror images. When the source is screened from the receiver, the total sound field is the sum of the four diffracted waves, corresponding to the paths linking the source S and its image S' with respect to the ground to the receiver R and its image R' via the top edge of the barrier (Fig. 4). Denoting with subscript 1 the paths including a reflection on the source side and by subscript 2 those for which there is a reflection on the receiver side

$$p = p_0 + p_1 + p_2 + p_{12}, \tag{18}$$

where p_0 stands for the diffracted wave obtained for path 0, linking source to edge to receiver with no ground reflection. Subscript $K \equiv 0,1,2,12$ is used in the following for general formulas valid for all four diffraction paths.

Using the same reference systems as in Sec. II, the coordinate vectors of the image source and the image receiver are $\mathbf{r}'_S$ and $\mathbf{r}'$, respectively. In the Cartesian coordinate system (x,y,z):

$$\mathbf{r}'_S(t) \equiv (-c_0 M t,\, y_S, -z_S), \quad \mathbf{r}'(t) \equiv (0, y, -z), \tag{19a}$$

whereas in the cylindrical coordinate system (r,ξ,x):

$$\mathbf{r}'_S(t) \equiv (r'_S, 2\pi - \xi_S, -c_0 M t), \quad \mathbf{r}'(t) \equiv (r', 2\pi - \xi, 0). \tag{19b}$$

Let z_E be the barrier height; then

$$r'_S = r_S \sqrt{1 - 4\frac{z_E}{r_S}\left(\cos \xi_S - \frac{z_E}{r_S}\right)}, \tag{20a}$$

$$r' = r \sqrt{1 - 4\frac{z_E}{r}\left(\cos \xi - \frac{z_E}{r}\right)}. \tag{20b}$$

Obviously the four diffracted paths have different lengths and as a result, each contribution in Eq. (18) has been emitted at a different time t_K, solution of

$$L_K = c_0(t - t_K), \tag{21a}$$

where the L_K's are the diffracted path lengths

$$L_0 = \sqrt{(r_S + r)^2 + (c_0 M t_0)^2},$$

$$L_1 = \sqrt{(r'_S + r)^2 + (c_0 M t_1)^2},$$

$$L_2 = \sqrt{(r_S + r')^2 + (c_0 M t_2)^2},$$

$$L_{12} = \sqrt{(r'_S + r')^2 + (c_0 M t_{12})^2}. \tag{21b}$$

The components of the Mach number in the direction of the diffracted rays M_K's are determined from the diffraction angles ε_K by

$$M_K = M \cos \varepsilon_K = -M \frac{c_0 M t_K}{L_K}. \tag{22}$$

Solving Eq. (21), each pressure wave component is calculated from Eq. (15), substituting the appropriate (image) source and (image) receiver coordinates, as well as the corresponding diffracted path and angle, L_K and ε_K. The total pressure is then calculated by substitution into Eq. (18).

In the presence of soft ground on either or both source and receiver side of the barrier, the corresponding reflection coefficients differ from 1 (hard ground). Equation (18) must be amended to account for ground reflections.[4] It is then necessary to consider the monopole and dipole components of the pressure field as expressed in Eq. (8) separately. Let β_1 and β_2 be the admittance of the ground on the source and receiver side, respectively. As in the previous section, subscript 1 denotes a reflection on the source side, whereas subscript 2 corresponds to a reflection on the receiver side.

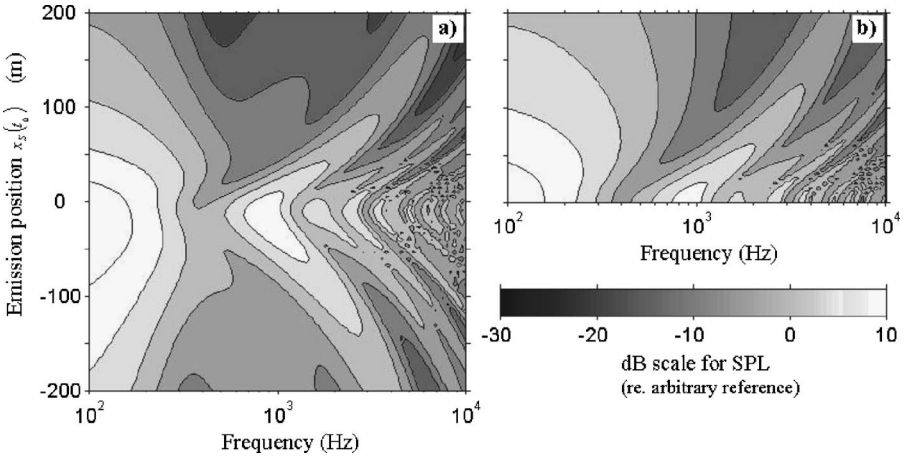


FIG. 5. (a) Sound field due to a source moving at Mach number 0.3 along a thin barrier of height 2 m. The source path is 10 m away from the barrier, at height 1 m. The receiver is in the shadow zone, 20 m away from the barrier and at height 1.2 m. The ground is hard ground on the source side and soft ground ($\sigma_e = 140$ kPa s m^{-2} and $\alpha_e = 35$ m^{-1}) on the receiver side. (b) Sound field due to a corresponding stationary source in the same geometry as (a), with offset varying from 0 to 200 m.

The monopole component expressing the sound field due to a stationary monopole in the presence of a barrier over the ground is given by Jonasson.[4] In the auxiliary coordinates

$$p^{(0)} = p_0^{(0)} + Q_1 p_1^{(0)} + Q_2 p_2^{(0)} + Q_1' Q_2' p_{12}^{(0)}. \tag{23}$$

The diffracted monopole pressures $p_K^{(0)}$ are calculated after Eq. (10). Dashes on the reflection coefficients applied to the last term of Eq. (23) denote the fact that the emission time for p_{12} is different from those for p_1 and p_2. In accordance with Huygens's principle that considers the diffracted wave as emitted by a secondary source located at the edge of the obstacle,[5] the spherical wave reflection coefficients are calculated for the paths joining the image source and the image receiver to the edge of the barrier,[7] respectively

$$\rho_1 = \frac{r_S'}{r + r_S'} L_1,$$

$$\rho_2 = \frac{r'}{r' + r_S} L_2, \quad \rho_1' = \frac{r_S'}{r' + r_S'} L_{12}, \quad \rho_2' = \frac{r'}{r' + r_S'} L_{12}. \tag{24}$$

Use of the Doppler-Weyl Van der Pol formula,[11] with $i \equiv 1,2$, gives

$$Q_i = R_{Pi} + (1 - R_{Pi}) F\left(\frac{w_i}{\sqrt{1 - M_i}}\right), \quad M_i = M \cos \varepsilon_i, \tag{25a}$$

where R_{Pi} is the plane wave reflection coefficient

$$R_{Pi} = \frac{\cos \theta_i - \beta_i}{\cos \theta_i + \beta_i}, \tag{25b}$$

and F is the boundary loss function

$$F\left(\frac{w_i}{\sqrt{1 - M_i}}\right) = 1 + i\sqrt{\pi} \frac{w_i}{\sqrt{1 - M_i}} \exp\left(-\frac{w_i^2}{1 - M_i}\right) \times \operatorname{erfc}\left(-i \frac{w_i}{\sqrt{1 - M_i}}\right), \tag{25c}$$

where erfc is the complementary error function and w_i is the numerical distance,

$$w_i = \left(-\tfrac{1}{2} i k_0 \rho_i\right)^{1/2} (\cos \theta_i + \beta_i), \tag{25d}$$

$$\cos \theta_1 = (z_S + z_E)/\rho_1, \quad \cos \theta_2 = (z + z_E)/\rho_2. \tag{25e}$$

Equations (25a)–(25e) hold for the calculation of Q_1' and Q_2' after substituting dashed parameters.

The dipole component of the pressure field is expressed in the auxiliary coordinate system, recalling the solution for the stationary dipole by Buret *et al.*[8] For a dipole along the x_L axis

$$p^{(1)} = p_0^{(1)} + Q_1 p_1^{(1)} + Q_2 p_2^{(1)} + Q_1' Q_2' p_{12}^{(1)} + \frac{\partial Q_1}{\partial x_S} p_1^{(0)} + \frac{\partial Q_1'}{\partial x_S} Q_2' p_{12}^{(0)}, \tag{26}$$

where the same notations as in Eq. (23) are used, with superscript (1) denoting pressure due to dipole sources. The diffracted pressures $p_K^{(1)}$ are calculated after Eq. (12). The last two terms in Eq. (26) involving the diffracted monopole pressures $p_1^{(0)}$ and $p_{12}^{(0)}$ can be neglected due to small variations of the spherical wave reflection coefficient in the horizontal plane. The acoustic pressure field due to a point source moving along a barrier over arbitrary ground is then straightforward, summing the monopole and the dipole contributions and transforming back into the coordinate system attached to the receiver

$$p = p_0 + Q_1 p_1 + Q_2 p_2 + Q_1' Q_2' p_{12}, \tag{27}$$

$$p_K = P_0 \frac{\exp(-i\omega t)}{4\pi} \frac{\exp(i k_0 L_K)}{L_K (1 - M \cos \varepsilon_K)^2} \times \left\{ \frac{1 + i}{2} \left[1 - \frac{1}{i k_0 L_K} \frac{M \cos \varepsilon_K - M^2}{1 - M \cos \varepsilon_K} \right] \left[A_D\left(\frac{X_{K+}}{\sqrt{1 - M \cos \varepsilon_K}}\right) + A_D\left(\frac{X_{K-}}{\sqrt{1 - M \cos \varepsilon_K}}\right) \right] \right\}, \tag{28a}$$

$$K \equiv 0,1,2,12. \tag{28b}$$

Calculation of the pressure field is given in Fig. 5(a) for hard ground on the source side and soft ground on the receiver side, and in Fig. 6(a) for soft ground on both sides. Figures 5(b) and 6(b) show the calculation of the SPL for a stationary source in the corresponding geometry, with offset

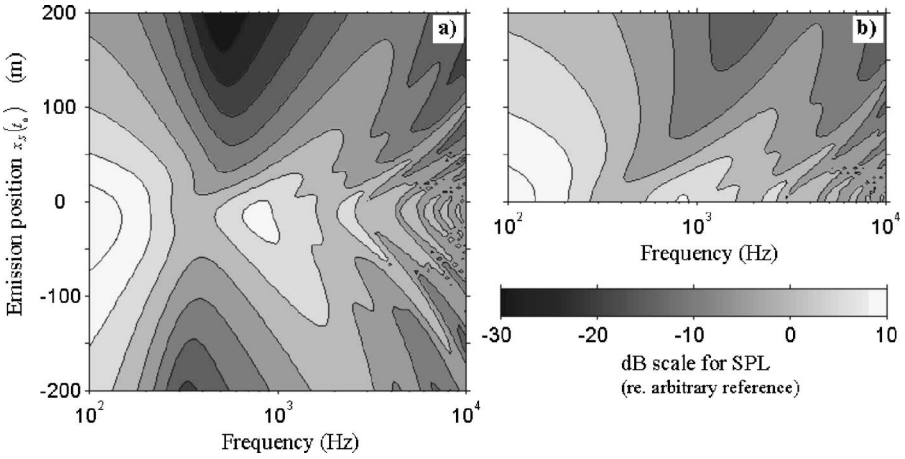


FIG. 6. Same as Fig. 5 but with soft ground on both sides of the barrier.

varying from 0 to 200 m. (The symmetrical pattern for offset −200 m to 0 is not shown). With common y axis, comparing with Figs. 5(a) to 5(b), to 6(a) and 6(b) allows assessing the effect of source convection which are not related to the change of source location along the trajectory—that is the influence of the Doppler and convection factors in the formulation of the sound field. For both Figs. 5 and 6, soft ground is characterized by a two parameter model admittance[17] with flow resistivity $\sigma_e = 140$ kPa s m^{-2} and porosity rate $\alpha_e = 35$ m^{-1}. The pressure field is found to be asymmetric along the source trajectory. Lower levels are predicted when the source recedes from the receiver $[x_S(t_0) \geqslant 0, (1 - M\cos\varepsilon_K)^{-2} \leqslant 1]$, whereas sensitivity of the sound field to source motion is predicted to be more important on source approach $[x_S(t_0) \leqslant 0, (1 - M\cos\varepsilon_K)^{-2} \geqslant 1]$, for which the deformation of the SPL pattern is stronger. Expectedly, strong interference between the four components of the acoustic pressure is predicted at high frequencies. The two dips related to the reflections on each side of the barrier are identifiable, particularly at large offsets and for source recession. Comparing Figs. 5 and 6, it is also predicted that sensitivity to motion is somewhat more important for soft ground than it is for hard ground.

C. Source in motion parallel to a ground discontinuity

A semiempirical formula for the sound field due to a stationary source in the presence of a ground discontinuity has been derived by De Jong *et al.*,[6] considering the superposition of two half-planes with different admittance. This model was shown to be valid in 3D situations for which the propagation path is not necessarily perpendicular to the discontinuity,[18] and was later extended to dipole sources.[8] Derivation of the acoustic pressure field for source motion in parallel to the discontinuity is then straightforward, after Eq. (8).

Consider an admittance step coinciding with the x axis, as shown in Fig. 7. Let a source S be in uniform motion at height z_S, in parallel to the discontinuity. The ground on the source side is characterized by its specific admittance β_1 and that on the receiver side by β_2. The direct wave reaching receiver R at time t has been emitted at instant t_e, following path $R = c_0(t - t_e)$ with azimuthal angle ψ and elevation angle ϕ. The reflected wave reaching R at time t (denoted by dashed parameters in the coordinate system attached to the receiver) has been emitted at retarded time t'_e and has followed path $R' = c_0(t - t'_e)$ with azimuthal and elevation angles $\psi' = \psi(t'_e)$ and $\theta' = \theta(t'_e)$, respectively. The expressions for R, ψ, ϕ and R', ψ', θ' are not detailed here for brevity, their derivation from a kinematics analysis for the source and its image with respect to the ground is straightforward.[12,13]

The diffracted wave follows path L as per defined in Eq. (3). Its time of emission t''_e is calculated after Eqs. (3) and (4). In the following, a double dash denotes corresponding parameters in the coordinate system attached to the receiver.

In the auxiliary coordinate system, the monopole component of the acoustic pressure for a single discontinuity is, after De Jong *et al.*[6]

$$p^{(0)} = \frac{\exp(-i\omega t_L)}{4\pi}\left[\frac{\exp(ik_0 R_L)}{R_L} + Q'_G\frac{\exp(ik_0 R'_L)}{R'_L} + (Q'_1 - Q'_2)\frac{1+i}{2}\frac{\exp(ik_0 L_L)}{L_L}D_L\right], \tag{29a}$$

where L_L is the diffraction path in the auxiliary coordinates for a horizontal half-plane whose edge coincides with the ground discontinuity and $D_L = A_D(X_{L+}) + A_D(X_{L-})$ the sum of the two diffraction integrals given in Eq. (10). The quantity Q'_G is the spherical wave reflection coefficient at the point of specular reflection, with $G \equiv 1$ or 2 depending on the geometry. The quantities, Q'_1 and Q'_2, are the spherical wave reflection coefficients for each type of ground. They are calculated after the Doppler-Weyl-Van der Pol formula[11]

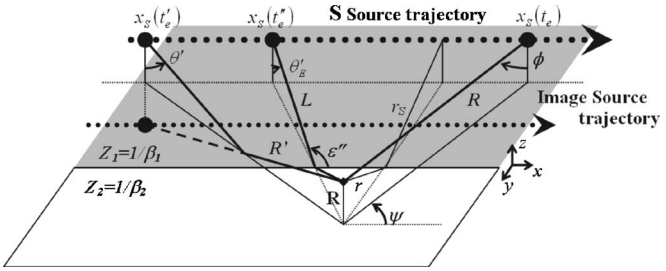


FIG. 7. Geometry for a source in uniform motion in parallel to a ground discontinuity.

$$Q'_i = R'_{Pi} + (1 - R'_{Pi})F\left(\frac{w'_i}{\sqrt{1 - M'_R}}\right), \tag{29b}$$

with M'_R as the component of the Mach number in the direction of propagation of the reflected wave

$$M'_R = M \cos \psi' \sin \theta', \tag{29c}$$

$$R'_{Pi} = \frac{\cos \theta' - \beta_i}{\cos \theta' + \beta_i}. \tag{29d}$$

The expression for the boundary loss function F is analogous to that given in (25c), with

$$w'_i = \left(-\tfrac{1}{2} i k_0 R'\right)^{1/2} (\cos \theta' + \beta_i). \tag{29e}$$

When the point of specular reflection coincides with the admittance step, $R'_L = L_L$ and $X_{L+} = 0$. With $A_D(0) = -(1-i)/2$, the continuity of the monopole component of the acoustic pressure is ensured.

The dipole component is also expressed as the sum of a direct, a reflected, and a diffracted pressure wave to which are applied the appropriate reflection coefficients. The formulation for the direct and the reflected fields are straightforward after Li *et al.*[19] The diffracted wave is calculated after Buret *et al.*[8] and the total dipole contribution is expressed as

$$\begin{aligned} p^{(1)} = \frac{\exp(-i\omega t_L)}{4\pi} \Bigg\{ & \Theta_R \frac{1 - i k_0 R_L}{R_L} \frac{e^{i k_0 R_L}}{R_L} \\ & + \Theta'_R \frac{1 - i k_0 R'_L}{R'_L} \frac{e^{i k_0 R'_L}}{R'_L} Q'_G \\ & + \frac{1+i}{2} \frac{e^{i k_0 L_L}}{L_L} \Bigg[\Theta_L \frac{1 - i k_0 L_L}{L_L} (Q'_1 - Q'_2) \\ & + \frac{\partial (Q'_1 - Q'_2)}{\partial x_L} \Bigg] D_L \Bigg\}, \end{aligned} \tag{30a}$$

with the directivity factors given by

$$\begin{aligned} \Theta_R &= \cos \psi_L \sin \phi_L, \\ \Theta'_R &= \cos \psi_L \sin \theta'_L, \\ \Theta_E &= \cos \psi_L \sin \theta_{E,L}. \end{aligned} \tag{30b}$$

In the earlier equation, ψ_L, ϕ_L, and θ_L are, respectively, the azimuthal angle, the elevation angles for the direct ray and the elevation angle for the reflected ray in the auxiliary coordinates. The quantity, $\theta_{E,L}$ is the elevation angle of the ray linking the source to the admittance step. Corresponding angles in the coordinates system attached to the receiver are shown in Fig. 7. Remarkably, when the specular reflection occurs at the admittance step: $\Theta'_R = \Theta_E = \cos \varepsilon_L$ and continuity of the acoustic pressure is ensured. The very last term in Eq. (30a) can be neglected, as variations of the spherical wave coefficient in the horizontal directions are known to be small.[8] Equation (30) then simplifies to an expression analogous to that of the monopole component given in Eq. (29).

The total pressure field is derived by substituting Eqs. (29) and (30) into Eq. (8). Transforming back into the coordinate system attached to the receiver is straightforward by means of the relations in Eq. (14) and noting that[11]

$$R_L = \gamma^2 R(1 - M \cos \psi \sin \phi), \quad t_L - \frac{R_L}{c_0} = t - \frac{R}{c_0},$$

$$\Theta_R = \frac{\cos \psi \sin \phi - M}{1 - M \cos \psi \sin \phi}, \tag{31a}$$

$$R'_L = \gamma^2 R'(1 - M \cos \psi' \sin \phi'), \quad t_L - \frac{R'_L}{c_0} = t - \frac{R'}{c_0},$$

$$\Theta'_R = \frac{\cos \psi' \sin \theta' - M}{1 - M \cos \psi' \sin \theta'}, \tag{31b}$$

and

$$\Theta_E = \frac{\cos \varepsilon'' - M}{1 - M \cos \varepsilon''} = \frac{\cos \psi'' \sin \theta''_E - M}{1 - M \cos \psi'' \sin \theta''}. \tag{31c}$$

The pressure field for a source in uniform motion along an admittance discontinuity is then expressed as

$$p = p_{\text{dir}} + Q_G p_{\text{refl}} + (Q_1 - Q_2) p_{\text{diffr}}, \tag{32a}$$

where p_{dir}, p_{refl}, and p_{diffr} represent the direct, reflected, and diffracted waves, respectively,

$$\begin{aligned} p_{\text{dir}} = P_0 \frac{\exp(-i\omega t)}{4\pi} & \frac{\exp(i k_0 R)}{R(1 - M \cos \psi \sin \phi)^2} \\ & \times \left[1 - \frac{M}{i k_0 R} \left(\frac{\cos \psi \sin \phi - M}{1 - M \cos \psi \sin \phi} \right) \right], \end{aligned} \tag{32b}$$

$$\begin{aligned} p_{\text{refl}} = P_0 \frac{\exp(-i\omega t)}{4\pi} & \frac{\exp(i k_0 R')}{R'(1 - M \cos \psi' \sin \phi')^2} \\ & \times \left[1 - \frac{M}{i k_0 R'} \left(\frac{\cos \psi' \sin \phi' - M}{1 - M \cos \psi' \sin \phi'} \right) \right], \end{aligned} \tag{32c}$$

$$\begin{aligned} p_{\text{diffr}} = P_0 \frac{\exp(-i\omega t)}{4\pi} & \frac{1+i}{2} \frac{\exp(i k_0 L)}{L(1 - M \cos \psi'' \sin \theta''_E)^2} \\ & \times \left[1 - \frac{M}{i k_0 L} \left(\frac{\cos \psi'' \sin \theta''_E - M}{1 - M \cos \psi'' \sin \theta''_E} \right) \right] \\ & \times \left[A_D \left(\frac{X_+}{\sqrt{1 - M \cos \psi'' \sin \theta''_E}} \right) \right. \\ & \left. + A_D \left(\frac{X_-}{\sqrt{1 - M \cos \psi'' \sin \theta''_E}} \right) \right]. \end{aligned} \tag{32d}$$

Figure 8(a) shows the calculation of the SPL calculated according to Eq. (32) as a function of the source position at emission of the reflected wave $x_S(t'_e)$ with a 33%-hard ground mix (when the proportion of hard ground below the sound propagation path is 33%). The source is moving at Mach number 0.3 with its track located 10 m away from the admittance step at height 1 m above hard ground. The receiver is located 20 m away from the discontinuity, 1.2 m above soft ground characterized by a two parameter model

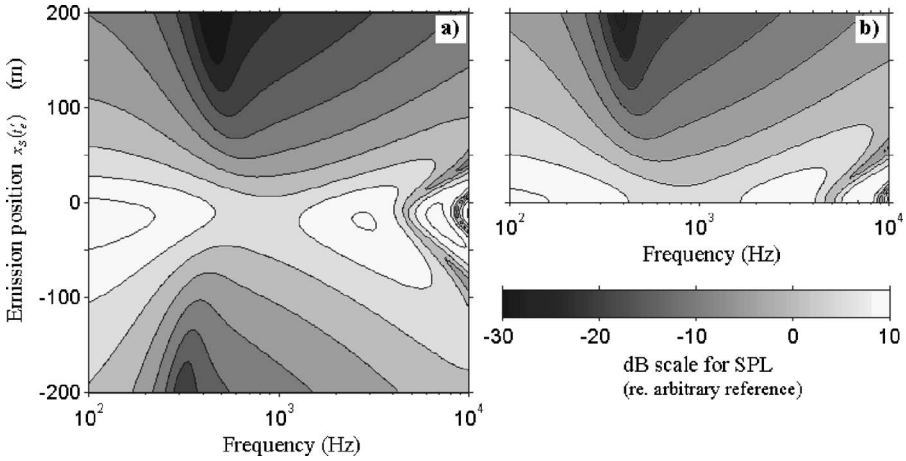


FIG. 8. (a) SPL (re. Arbitrary reference) for a source moving at Mach number 0.3 parallel to an admittance discontinuity, 10 m away from it, at height 1 m. The receiver is 20 m away from the step line at height 1.2 m. The ground on the source side is hard, the ground on the receiver side is characterized by $\sigma_e=140$ kPa s m^{-2} and $\alpha_e=35$ m^{-1}. The proportion of the sound path occurring above hard ground is 33%. (b) SPL for a stationary source in the same geometry as in (a), with offset varying from 0 to 200 m.

admittance[17] with flow resistivity $\sigma_e=140$ kPa s m^{-2} and porosity rate $\alpha_e=35$ m^{-1}. Figure 8(b) shows the SPL for the corresponding stationary source at offset 0–200 m and allows assessing the effects of source motion. The corresponding instantaneous excess attenuation EA $=10\log(|p(t)/p_{\mathrm{dir}}(t)|)$ is shown in Fig. 9(a), for the source at offset 100 m on approach [$x_S(t'_e)=-100$ m, thick solid line] and recession [$x_S(t'_e)=+100$ m; thin solid line] and for the corresponding stationary source (dashed line). In Fig. 9(a), dotted lines of different thickness showing the calculation at closet position [$x_S(t'_e)=0$] for moving and stationary source, respectively, are coinciding: effects of motion on ground attenuation when the source is close to the receiver can be neglected. Figure 9(b) shows the excess attenuation but for the source track over hard ground 20 m away from the admittance step and receiver over soft ground 10 m away (the proportion of the sound path occurring above hard ground is 66%).

As was noted previously in the case of homogenous ground,[11] the effects of motion are predicted to be small. Again, it is predicted that the SPL shows a dissymmetry along the source motion due to the Doppler shift and the convection factor that affect the location and magnitude of the ground effect dip. As clearly shown in Fig. 9, this is predicted to result in lower SPL than for a stationary source when the source approaches the receiver [$x_S(t'_e)\leqslant 0$], up to the frequency of the ground effect dip and higher levels from the dip up to the interference in the high frequencies. This trend is reversed when the source recedes from the receiver. On source approach, the ground effect dip is slightly sharper and deeper than that for a stationary source, whereas on recession it is slightly broader and shallower. This is due to the Doppler factor within the argument of the boundary loss factor as given in the Doppler-Weyl-Van der Pol formula [cf. Eq. (29b)]: on source approach the absorbing ground is seen as softer, whereas it is seen as harder on recession. On the other hand, comparing Figs. 9(a) and 9(b), the sound field is predicted to be little affected by the change of the proportion in the ground mix covering the source—receiver path.

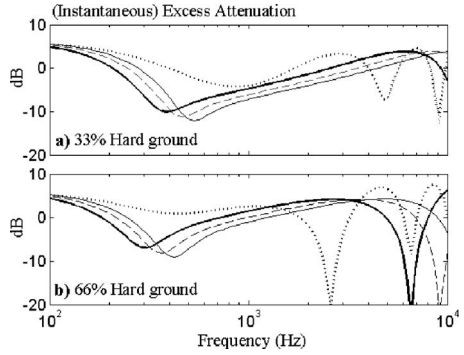


FIG. 9. Instantaneous excess attenuation (a) for the same configurations as in Fig. 8, (b) for a proportion of hard ground over the propagation path of 66% (source path over hard ground 20 m away from the discontinuity, receiver 10 m away, over soft ground). (—) Approaching source $x_S(t'_e)=-100$ m, (—) Receding source $x_S(t'_e)=100$ m, (---) stationary source at offset $x_S=100$ m, (·····) moving source at closest separation [$x_S(t'_e)=0$] and coinciding with the calculation for a stationary source at closest separation (·····).

IV. CONCLUDING REMARKS

It has been shown that the problem of diffraction of the sound field due to a source in motion parallel to a semi-infinite rigid wedge is strongly analogous to the case of a stationary source, as a result of the conservation of the geometry in the cross-sectional plane of the obstacle. By means of an auxiliary coordinate system and the formulations for monopole and dipole sources, a closed-form formula for the diffracted sound field due to a monopole in uniform motion parallel to a diffracting edge has been derived that is similar to Pierce's solution for a stationary monopole and its extension to dipole sources.

Although motion results in dissymmetry of the diffracted pressure along the source trajectory, its effects on barrier attenuation are predicted to be limited. In particular the effects on IL are predicted to be minimal so that design schemes used for stationary sources are valid also for sources in motion.

On the other hand, predictions for the absolute sound pressure levels must account for the influence of the convection and Doppler factors on the resulting sound field.

New models have been derived for propagation of sound due to sources moving parallel to barriers in the presence of the ground and greater sensitivity to motion is predicted on source approach and for softer grounds. De Jong's semi-empirical model for propagation in the presence of a ground discontinuity has been extended also to the case of source

motion along an admittance step. Sensitivity to source motion is also predicted to be stronger on source approach, whereas it is predicted to be little affected by the proportion of hard and soft ground covering the sound propagation path.

Further extensions of this work could include effects of absorptive barriers and impedance strips parallel to the axis of source motion.

ACKNOWLEDGMENTS

This work was prepared when one of the authors (M.B.) was on study leave in the Department of Mechanical Engineering, the Hong Kong Polytechnic University. The authors would like to thank the Research Committees of the Open University and the Hong Kong Polytechnic University for their financial support in this project. M.B. was supported by an Open University Competitive Studentship. One of the authors (K.M.L.) gratefully acknowledges a generous donation from MTR Corporation for partial support of this research.

[1]J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi, *Electromagnetic and Acoustic Scattering by Simple Shapes* (North-Holland, Amsterdam, 1969).
[2]U. Kurze, "Noise reduction by barriers," J. Acoust. Soc. Am. **55**(3), 504–518 (1974).
[3]Z. Maekawa, "Noise reduction by screens," Appl. Acoust. **1**, 157–173 (1968).
[4]H. G. Jonasson, "Sound reduction by barriers on the ground," J. Sound Vib. **22**(1), 113–126 (1972).
[5]A. D. Pierce, "Diffraction of sound around corners and wide barriers," J. Acoust. Soc. Am. **55**(5), 941–955 (1974).
[6]B. A. De Jong, A. Moerkerken, and J. D. Van der Toorn, "Propagation of sound over grassland and over an earth barrier," J. Sound Vib. **86**(1), 23–46 (1983).
[7]P. Koers, "Diffraction by an absorbing barrier or by an impedance transition," *Proc. InterNoise 83, 1983*, Vol. **1**, pp. 311–314.
[8]M. Buret, K. M. Li, K. Attenborough, "Diffraction of sound from a dipole source near to a barrier or an impedance discontinuity," J. Acoust. Soc. Am. **113**(5), 2480–2494 (2003).
[9]P. Menounou, J. L. Bush-Vishniac, and D. T. Blackstock, "Directive line source model: A new model for sound diffraction by half planes and wedges," J. Acoust. Soc. Am. **107**(6), 2973–2986 (2000).
[10]R. Makarewicz, J. Jarzecki, K. Berezowska-Apolinarska, and A. Preis, "Rail transportation noise with and without a barrier," Appl. Acoust. **26**, 135–147 (1989).
[11]M. Buret, K. M. Li, and K. Attenborough, "Optimisation of ground attenuation for moving sources," Appl. Acoust. **67**(2), 135–157 (2006).
[12]K. M. Li, M. Buret, and K. Attenborough, "The propagation of sound due to a source moving at high speed in a refracting medium," *Proc. Euronoise 98, 1998*, Vol. **2**, pp. 955–960.
[13]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw Hill, New York, 1968).
[14]A typographical error in Eq. (11.2.15) of Ref. 13 should be corrected. The denominator for the second term should read $R^2(1-\mathrm{M}\cos\theta)^3$ [in place of $R^2(1-\mathrm{M}\cos\theta)^2$].
[15]A. D. Pierce, *Acoustics, An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, New York, 1989).
[16]The orientation of dipole missing in Eq. 14a of Ref. 8 has been added in Eq. 11 where its direction cosines (ℓ_x,ℓ_y,ℓ_z) are given by $(\cos\varepsilon_L,0,0)$ in the present situation.
[17]R. Raspet and J. M. Sabatier, "The surface impedance of grounds with exponential porosity profiles," J. Acoust. Soc. Am. **99**(1), 147–152 (1996).
[18]D. C. Hothershall and J. N. B. Harriot, "Approximate models for sound propagation above multi-impedance plane boundaries," J. Acoust. Soc. Am. **97**(2), 918–926 (1995).
[19]K. M. Li, S. Taherzadeh, and K. Attenborough, "Sound propagation from a dipole source near an impedance plane," J. Acoust. Soc. Am. **101**(6), 3343–3352 (1997).

Internal wave effects on the ambient noise notch in the East China Sea: Model/data comparison

Daniel Rouseff and Dajun Tang
Applied Physics Laboratory, College of Ocean and Fishery Sciences, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105

(Received 29 November 2005; revised 1 June 2006; accepted 19 June 2006)

The vertical directivity pattern of the ambient noise field observed in shallow water is typically anisotropic with a trough in the horizontal. This trough, often called the ambient noise notch, develops because downward refraction steepens all rays emanating from near the sea surface. Variability in the environment has the potential to redistribute the noise into shallower angles and thereby fill the notch. In the present work, a model for the width and depth of the ambient noise notch is developed. Transport theory for acoustic propagation is combined with a shallow water internal wave model to predict the average output of a beamformer. Ambient noise data from the East China Sea are analyzed in the 1-to-5-kHz band. Good agreement between the model and the data for both the width and depth of the ambient noise notch is obtained at multiple frequencies, suggesting that internal wave effects are significant. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2225458]

PACS number(s): 43.30.Bp, 43.30.Ft, 43.30.Nb [DRD] Pages: 1284–1294

I. INTRODUCTION

The vertical directivity pattern of ambient noise often displays a pronounced trough in the horizontal direction. In shallow water, appropriate conditions for the existence of this trough include a mixed layer of warm water near the sea surface. Downward refraction then causes noise generated in this mixed layer by the wind or distant shipping to propagate at relatively steep angles. The resulting noise trough at relatively shallow angles is often called the ambient noise notch. When present, the ambient noise notch may have significant impact on passive detection schemes; in effect, it creates a window to look through the noise for targets of interest. Oceanographic phenomena like internal waves, however, have the potential to redistribute noise into shallow angles and thereby fill the noise notch. Consequently, being able to predict the extent to which a noise notch will exist under different oceanographic conditions is highly desirable.

Based on Belousov and Furduev,[1] Fig. 1 shows the vertical directivity pattern of ambient noise. While highly idealized, the figure is useful for describing the pattern's basic characteristics and for defining the quantities of interest. In shallow water, the noise is typically stronger looking above horizontal, B_{up}, than looking below horizontal, B_{down}. The ratio $B_{\mathrm{down}}/B_{\mathrm{up}}$ can be used to estimate the bottom reflectivity.[1–6] The average of B_{down} and B_{up} is B_{avg}. The ambient noise notch has width $2\theta_0$ and relative depth $B_{\mathrm{notch}}/B_{\mathrm{avg}}$. The notch width can be estimated from simple application of Snell's law,[1,7,8] but estimating the expected notch depth is more challenging. In the present work, a model is derived for the width and depth of the noise notch in the presence of shallow water internal waves. The emphasis is on frequencies between 1 and 5 kHz.

There have been numerous approaches to modeling the ambient noise field in general and the ambient noise notch in particular. The early emphasis was on deep-water, low-frequency applications, but in recent years shallow water and higher frequencies have attracted attention. Ray-based analytical models include those derived by Dashen and Munk[9] and by Harrison.[10] A recent ray-based numerical model has been developed by Ferat and Arvelo.[11] Kuperman and Ingenito's work[12] has been the basis for models using normal modes. The parabolic equation method has been used alone,[13] in conjunction with normal modes,[14] and in conjunction with Monte Carlo techniques.[15] Dating apparently to a paper by Kuryanov and Klyachin,[16] Russian noise notch models use simplified versions of transport theory to derive very simple formulas for the noise notch depth.[1,2,17,18] Evans[19] recently used a fuller version of transport theory to predict the noise notch characteristics in deep water at 50 Hz.

In the present work, a fuller implementation of transport theory is combined with a modern shallow water internal wave model to calculate the noise notch characteristics. Of concern is not the specific mechanisms for generating noise in the 1-to-5-kHz band, but rather for the way in which it is coupled into the sound speed duct and thereby affects the ambient noise notch. In Sec. II, the model is derived. The Dozier-Tappert transport theory for acoustic propagation is combined with a shallow water version of the Garrett-Munk internal wave model. Computationally efficient techniques for implementing the combined model are discussed. In Sec. III, the model is tested using data collected in the East China Sea as part of the 2001 ASIAEX experiment.[20] Assuming only a typical internal wave strength and with no further adjustable parameters, excellent agreement is obtained between the model and the data at multiple frequencies. Implications of the results and suggestions for future research are in Sec. IV.

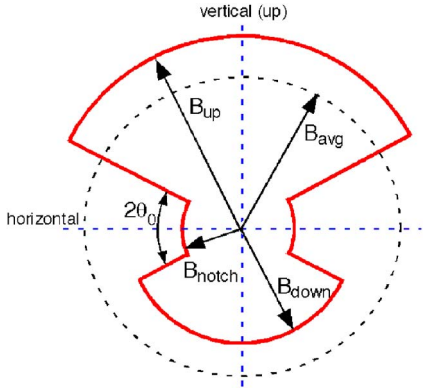


FIG. 1. (Color online) Idealized vertical directivity pattern of ambient noise. The ambient noise notch has relative depth $B_{\text{notch}}/B_{\text{avg}}$ and width $2\theta_0$. Figure modeled after Belousov and Furduev.[1]

II. MODEL

Since internal waves are ubiquitous in the ocean, their effect on the noise notch must be considered. From the standpoint of acoustic modes, internal waves drive energy from the high-order modes that are excited by the noise sources into low-order modes that then propagate. A beamformer steered from the vertical into the horizontal direction is sensitive to these now populated low-order modes. The result is a noise notch that is diminished from what would exist in the absence of internal waves.

Developing a complete model for the ambient noise notch requires constituent models for the beamformer, for acoustic propagation, and for the internal waves. These constituent models are developed in Secs. II A–II C. The reader interested only in the final combined model is referred to Sec. II D. Comments on implementation are deferred to Sec. II E.

A. Beam-formed noise model

Let P_n represent the total time-harmonic $\exp(-i2\pi ft)$ noise field measured on the nth element of a vertical array. Assuming the array elements are uniformly spaced Δ apart, the output $B(\theta)$ of a Bartlett beamformer is

$$B(\theta) = \left| \sum_{n=1}^{N} w_n^* P_n \right|^2, \tag{1}$$

where the weights

$$w_n = \exp(ik\Delta(n-1)\sin\theta). \tag{2}$$

Beamformer steering angle $\theta=0$ corresponds to the horizontal and the presumed direction of the noise notch. The wave number k is proportional to the frequency f and is calculated using the sound speed in the vicinity of the array. The total noise field is modeled as being the sum of contributions from L noise sources

$$P_n = \sum_{l=1}^{L} p^{(l)}(z_n), \tag{3}$$

where the superscript is the noise source index and the dependence on the depth of the receiving element z_n is made explicit. The field from the $(l)^{\text{th}}$ source is written in a normal mode expansion

$$p^{(l)}(z_n) = \sum_{m=1}^{M} A_m^{(l)} \Psi_m(z_n)/\sqrt{\xi_m r^{(l)}}. \tag{4}$$

The real mode functions Ψ_m and wave numbers ξ_m are determined using the average background sound speed profile $c_0(z)$. For a range-dependent environment featuring shallow water internal waves, the complex mode amplitudes A_m include the effects of mode coupling. Implicitly, each A_m is a function of the range $r^{(l)}$ between that noise source and the array.

Combining Eqs. (1)–(4), the beamformer output can be written in matrix form

$$B(\theta) = \mathbf{w}^\dagger \mathbf{\Psi} \mathbf{C} \mathbf{\Psi}^\dagger \mathbf{w}, \tag{5}$$

where † denotes the conjugate transpose. The N elements in the column vector $\mathbf{w}$ are given by Eq. (2). In the usual matrix notation, the elements in $\mathbf{\Psi}$ and $\mathbf{C}$ are, respectively,

$$\Psi_{nm} = \Psi_m(z_n)/\sqrt{\xi_m}, \tag{6}$$

$$C_{mm'} = \sum_{l,l'} A_m^{(l)} A_{m'}^{(l')*} / \sqrt{r^{(l)} r^{(l')}}. \tag{7}$$

When the number of noise sources L is large, the dominant terms in $\mathbf{C}$ lie along the diagonal. Retaining only these terms and using the Kronecker delta function,

$$C_{mm'} \approx \delta_{mm'} \sum_{l=1}^{L} |A_m^{(l)}|^2 / r^{(l)}. \tag{8}$$

The validity of the approximation Eq. (8) will be tested in subsequent numerical calculations.

In a practical problem, the beamformed noise will be random and it is desirable to calculate the average of Eq. (5). This averaging is approximated in two steps, the first considering randomness in the environment and the second considering randomness in the position of the noise sources. The randomness in the environment is embedded in the complex mode amplitudes. For noise source l, define

$$\Gamma_m^{(l)} \equiv \langle |A_m^{(l)}|^2 \rangle. \tag{9}$$

The average mode intensity, Eq. (9), can be calculated using the transport theory outlined in the following section. Implicit in the calculation is that the location of noise source l is known. If the noise source locations are not known, they too are treated as being random. Statistical information about the range, depth, and bearing for the noise sources relative to the receiving array might be known for a particular location or scenario. As a simple but relevant example, assume that both the range and the depth for each of the L noise sources in Eq. (3) is an independent, identically distributed (i.i.d.) random variable. Using Eqs. (5), (8), and (9):

$$\langle B(\theta) \rangle = \mathbf{w}^\dagger \mathbf{\Psi} \langle \mathbf{C} \rangle \mathbf{\Psi}^\dagger \mathbf{w}, \tag{10}$$

where $\langle \mathbf{C} \rangle$ has elements

$$\langle C_{mm'} \rangle \approx \delta_{mm'} L \langle r^{-1} \Gamma_m \rangle. \tag{11}$$

Note that the i.i.d. assumption permits the superscript l to be dropped as each noise source has the same statistical distribution.

In the remainder of this section, models for acoustic propagation and shallow water internal waves are developed that permit Eqs. (10) and (11) to be evaluated.

B. Acoustic propagation model

To implement Eq. (10) for the average beamformed noise, it is necessary to calculate the average intensity Γ_m for each acoustic mode m as a function of propagation range. In this section, the pertinent results from the Dozier-Tappert modal transport theory[21–23] are summarized.

Equation (4) gives the coupled mode representation for the field due to an individual noise source. In the narrow angle parabolic approximation, the mode amplitude A_m satisfies

$$\frac{\partial A_m}{\partial r} - (\xi_m + i\alpha_m) A_m = -i \sum_n \rho_{mn} A_n, \tag{12}$$

where the mode coupling coefficient,

$$\rho_{mn}(r) = \frac{k^2}{\sqrt{\xi_m \xi_n}} \int (\delta c(r,z)/c_0(z)) \Psi_m(z) \Psi_n(z) dz, \tag{13}$$

quantifies the dependence on the range- and depth-varying sound speed perturbation δc. A perturbation expansion gives the modal attenuation α_m as

$$\alpha_m = \frac{2\pi f}{\xi_m} \int (\alpha(z)/c_0(z)) \Psi_m^2(z) dz, \tag{14}$$

where $\alpha(z)$ is the volume absorption profile. The original derivation by Dozier and Tappert[21] considered a deep-water problem and so neglected modal attenuation. Dozier[22] subsequently generalized the derivation to include modal attenuation, an important factor for shallow water applications. At low frequencies, the integration in Eq. (14) need only consider attenuation in the sediment. At the kilohertz frequencies of current interest, the volume absorption profile includes both the sediment and the water column.

The solution to Eq. (12) has a particularly simple form for the special case of a range-independent ($\delta c = 0$) environment:

$$A_m = \Psi_m(z_s) e^{i(\xi_m + i\alpha_m) r}, \tag{15}$$

where the initial condition is the mode function as evaluated at the source depth z_s. This special case will be used in subsequent sections to test some of the approximations made in developing the noise notch model.

For the general range-dependent case, a quantity of interest is the correlation function of the mode coupling coefficient ρ_{mn}. If the sound speed perturbations form a Gaussian stochastic process, it follows that the correlation function has the form

$$R_{mn}(r_2 - r_1) = \langle \rho_{mn}(r_2) \rho_{mn}^*(r_1) \rangle. \tag{16}$$

Assuming that the modal attenuation is small over the correlation length of R_{mn}, Creamer[23] shows how an important quantity is the cosine transform of Eq. (16),

$$a_{mn} = 2 \int_0^\infty R_{mn}(r) \cos(\xi_{mn} r) dr, \tag{17}$$

as evaluated at the wave number difference

$$\xi_{mn} = \xi_m - \xi_n. \tag{18}$$

Let $\mathbf{\Gamma}$ represent a column vector containing the average mode intensities, Eq. (9). Equations (12)–(18) provide the necessary ingredients to derive the transport equation

$$\frac{\partial \mathbf{\Gamma}}{\partial r} = -\mathbf{S}\mathbf{\Gamma}. \tag{19}$$

The transport matrix $\mathbf{S}$ has elements

$$S_{mn} = \begin{cases} -a_{mn}, & m \neq n, \\ 2\alpha_m + \sum_{n' \neq m} a_{mn'}, & m = n. \end{cases} \tag{20}$$

Note that only the diagonal $m=n$ elements contain the modal attenuation α_m.

The formal solution to Eq. (19) can be written using the matrix exponential

$$\mathbf{\Gamma}(r) = \exp(-\mathbf{S}r) \mathbf{\Gamma}(0). \tag{21}$$

The initial conditions $\mathbf{\Gamma}(0)$ are the mode intensities at the range $r=0$ and depth z_s of the source.

Dozier[22] and Creamer[23] were interested in the asymptotic behavior of the mode intensities. In the asymptotic limit, the smallest singular value of $\mathbf{S}$ determines the decay rates for all of the modes. In the present application, some noise sources will be too close to the receiving array for the asymptotic result to apply, so the more general Eq. (21) will be retained.

C. Internal wave model

In deep water, internal waves propagate in many directions. Since the waves are randomly phased, it is appropriate to represent their sum as a random process described by a spectrum. The Garrett-Munk spectral model[24] has found wide application in deep water for these background internal waves. In shallow water, the situation is more complicated. In addition to the background field, tidally driven packets of strong internal waves may propagate from distinct bathymetric features. These internal waves, sometimes called solibores,[25] are strong agents of acoustic mode coupling. In the present work, only background internal waves are considered. The choice is justified partly by the fact that the background internal waves are always present while the solibores are transient. A shallow-water version of Garrett-Munk model is sketched in a form compatible with the noise notch model.

To first order, the sound speed perturbation of Eq. (13) is linearly related to the internal wave displacement $\zeta(\mathbf{r})$ via[26]

$$\delta c(\mathbf{r}) = -\left(\frac{\partial c_0}{\partial z} - \frac{\partial c_a}{\partial z}\right)\zeta(\mathbf{r}) \equiv -\partial_z c_P \zeta(\mathbf{r}). \tag{22}$$

In Eq. (22), $\partial c_0/\partial z$ is the gradient of the background sound speed profile used previously to calculate the acoustic modes and $\partial c_a/\partial z = 0.016\ \mathrm{s}^{-1}$ is the gradient of the adiabatic part of the sound speed.

As with the acoustic field, it is convenient to use a normal mode expansion for the internal wave field. The internal wave modes W_j satisfy the eigenvalue equation

$$\frac{\partial^2 W_j}{\partial z^2} + \frac{N^2(z) - \omega^2}{\omega^2 - \omega_i^2} K^2 W_j = 0. \tag{23}$$

The boundary conditions are that the internal wave displacements are zero at the sea surface and at the bottom. The depth-dependent buoyancy frequency $N(z)$ is determined by the vertical gradient of the water density. The frequency ω and horizontal wave number K for the internal waves are related by a dispersion relationship. The inertial frequency ω_i is known from the latitude of a particular site.

Historically, it has often been assumed that the buoyancy is a simple function of depth. In deep water, a decaying exponential model is often used and the internal wave modes are then calculated using the WKB approximation. In shallow water, setting $N(z)$ to a constant lets one then express W_j using simple sinusoids. For the present application, both of these approximations are inadequate. The width and depth of the ambient noise notch will depend on both the background sound speed profile and the buoyancy profile; it is crucial that these two profiles are realistic and consistent with one another. For buoyancy profiles based on experimental data, Eq. (23) can be solved numerically.[27]

The spatial correlation for the internal waves is[28]

$$\langle \zeta(x,y,z)\zeta(x',y',z')\rangle = \sum_j \int F(\omega,j) W_j(K;z) W_j(K;z') J_0(K r_d) d\omega, \tag{24}$$

where the summation is over the mode index and the integration is over frequency. The Bessel function arises because the internal waves are assumed to be statistically homogeneous and isotropic in the horizontal. The magnitude-squared of the difference in horizontal position is $r_d^2 = (x-x')^2 + (y-y')^2$. Note that the correlation is not statistically homogeneous in the vertical as z and z' appear separately. The Garrett-Munk spectrum,

$$F(\omega,j) = bE_{GM}(bN_0)^2 \frac{2}{\pi}\frac{\omega_i}{\omega^3}\sqrt{\omega^2 - \omega_i^2} H(j), \tag{25}$$

is separable with dependence on mode index j contained in

$$H(j) = \frac{n_s}{j^2 + j_*^2}. \tag{26}$$

The canonical deep-water value for the characteristic mode number is $j_* = 3$. In shallow water, low-order internal wave modes dominate and a smaller value is appropriate. The normalization n_s is fixed so that the summation of $H(j)$ over all modes equals unity. Using $j_* = 0$, it follows[29] that $n_s = 6/\pi^2$.

As noted earlier, in deep water the buoyancy frequency is often modeled as an exponential function of depth where N_0 is the reference value and b, having units of length, is the e-folding depth. In undoing the effects of the exponential model, the product bN_0 occurring in Eq. (25) is defined by an explicit integration over the depth of the water column D:[28]

$$bN_0 \equiv \int_0^D N(z) dz. \tag{27}$$

For a buoyancy profile observed in an experiment, the integration can be approximated by numerical quadrature. For shallow water, bN_0 is typically on the order of 1 m/s.

The final parameter in Eq. (25) is the dimensionless internal wave energy density E_{GM}. In the present formulation, E_{GM} enters only as a product with b. Using the canonical deep-water values[24] for b and E_{GM} yields $bE_{GM} = 0.08$ m. In shallow water, bE_{GM} is typically on the order of 1 m or less. For data taken near Nantucket,[30] bE_{GM} was estimated to be 0.5 m.[28] For data taken on the New England shelf during the 1996 Coastal Mixing and Optics Experiment,[31] $bE_{GM} = 0.35$ m.[32,33]. For data taken in a less energetic internal wave field in the Mediterranean, $bE_{GM} = 0.135$ m.[34] In the model/data comparisons of Sec. III, bE_{GM} will be treated as the lone adjustable parameter.

D. Combined model

The primary quantity of interest, the average beamformed ambient noise, is given by Eq. (10). The averaging in Eq. (10) is accomplished in two steps: the matrix $\mathbf{\Gamma}$ of average acoustic mode intensities is first calculated, and the result is then averaged again over the location of the noise sources as in Eq. (11). To calculate these averages, it is necessary to combine the models of the previous sections for the beamformer, the acoustic propagation, and the ocean internal waves. In this section, these models are combined to yield an estimate for the average beamformed noise. The matrix elements a_{mn} needed for $\mathbf{\Gamma}$ are calculated for the shallow water internal wave model. The formal averages over the noise source locations are then evaluated.

Equation (24) gives a statistical model for the internal wave displacements. Using Eq. (22), this can be translated into a model for the sound speed perturbations that can, in turn, be used to calculate the correlation of the mode coupling term ρ_{mn} as defined in Eqs. (13) and (16). Combining these equations, taking the cosine transform via Eq. (17), then rearranging terms yields for the matrix elements

$$a_{mn} = \frac{2k^4}{\xi_m \xi_n} \sum_j \int F(\omega,j) |Y_{jmn}(K)|^2 I_K \, d\omega, \tag{28}$$

where

$$Y_{jmn}(K) = \int_0^D (\partial_z c_P / c_0) \Psi_m(z) \Psi_n(z) W_j(K;z) dz. \tag{29}$$

The integral over depth Y_{jmn} involves the mode functions for acoustic modes m and n and internal wave mode j. The integral over range in Eq. (28) can be evaluated[35]

$$I_K = \int_0^\infty J_0(Kr)\cos(\xi_{mn} r)\,dr$$
$$= \begin{cases} (K^2 - \xi_{mn}^2)^{-1/2}, & K > |\xi_{mn}|, \\ 0, & K < |\xi_{mn}|. \end{cases} \tag{30}$$

The integral over internal wave frequency ω in Eq. (28) can be approximated. Note that Y_{jmn} is an implicit function of internal wave frequency through the dispersion relationship linking K and ω. Furthermore, because I_K introduces an integrable singularity, the dominant contribution to the integral occurs near $K=|\xi_{mn}|$. Evaulating Y_{jmn} at the singularity and removing it from the integral yields

$$a_{mn} \approx \frac{2k^4}{\xi_m \xi_n} \sum_j |Y_{jmn}(\xi_{mn})|^2 \int F(\omega, j) I_K \, d\omega. \tag{31}$$

The approximation in Eq. (31) is effectively the same as the resonance approximation made by Dozier and Tappert: the internal wave spectrum is evaluated at a wave number equal to the difference of two acoustic mode wave numbers.[21–23] This fixes the physical length of the internal waves that contribute to coupling between a given pair of acoustic modes.

To complete the calculation of a_{mn}, it is necessary to use a specific form for the internal wave dispersion relationship. An approximate form[26] adequate for present purposes is

$$K = \frac{j\pi}{bN_0}(\omega^2 - \omega_i^2)^{1/2}, \tag{32}$$

where bN_0 is given by Eq. (27). Substituting Eq. (32) into Eq. (30),

$$I_K = \begin{cases} (\omega^2 - \omega_c^2)^{-1/2}, & \omega > \omega_c, \\ 0, & \omega < \omega_c, \end{cases} \tag{33}$$

with cutoff frequency

$$\omega_c = [(bN_0 \xi_{mn})^2/(j\pi)^2 + \omega_i^2]^{1/2}. \tag{34}$$

After substituting the expression for the internal wave spectrum [Eq. (25)] and rearranging terms, the mode coupling term a_{mn} reduces to

$$a_{mn} = \frac{4}{\pi}\frac{k^4}{\xi_m \xi_n} bE_{GM}(bN_0)^2 \sum_j |Y_{jmn}(\xi_{mn})|^2 H(j) I_\omega(j), \tag{35}$$

where the simplified integral over frequency is

$$I_\omega(j) = \int_{\omega_c}^\infty \frac{\omega_i}{\omega^3}\sqrt{\frac{\omega^2 - \omega_i^2}{\omega^2 - \omega_c^2}}\, d\omega. \tag{36}$$

This remaining integral can be evaluated in its limiting cases. When the cutoff frequency ω_c is large compared to the inertial frequency ω_i,

$$I_\omega(j) \approx \int_{\omega_c}^\infty \frac{\omega_i}{\omega^2}\sqrt{\frac{1}{\omega^2 - \omega_c^2}}\, d\omega = \frac{\omega_i}{\omega_c^2}. \tag{37}$$

In the other limit, as ω_c approaches ω_i,

$$I_\omega(j) \approx \int_{\omega_c}^\infty \frac{\omega_i}{\omega^3}\, d\omega = \frac{\omega_i}{2\omega_c^2}. \tag{38}$$

The two limiting cases differ only by a factor of 2. Consistent with the fact that the low-order internal wave modes are of primary importance in shallow water, numerical calculations show that Eq. (37) applies to most cases of practical interest. Using this approximate form for $I_\omega(j)$, together with the mode weighting Eq. (26) appropriate for $j_*=0$, yields the final result for a_{mn}:

$$a_{mn} = \frac{24}{\pi}\frac{k^4}{\xi_m \xi_n} bE_{GM} \sum_j \frac{|Y_{jmn}(\xi_{mn})|^2}{\xi_{mn}^2 + (j\pi\omega_i)^2/(bN_0)^2}. \tag{39}$$

Together with the modal attenuation Eq. (14), a_{mn} defines the elements in the transport matrix $\mathbf{S}$ used to calculate the average acoustic mode intensities via Eq. (21).

The remaining task is to average over the noise source locations. It will be convenient to use a singular value decomposition of the transport matrix

$$\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\mathbf{T}, \tag{40}$$

where $\mathbf{\Lambda}$ is a diagonal matrix with elements λ_m. Equation (21) for the mode intensities $\mathbf{\Gamma}$ is rewritten as

$$\mathbf{\Gamma}(r) = \exp(-\mathbf{S}r)\mathbf{\Gamma}(0) = \mathbf{U} \cdot [\mathrm{diag}(e^{-\lambda_m r})] \cdot \mathbf{V}^\mathbf{T}\mathbf{\Gamma}(0). \tag{41}$$

From Eq. (11), the quantity necessary for calculating the average beamformer response is $\mathbf{\Gamma}$ divided by range and then averaged over the noise source position. In general, both the range r and the depth z_s of the noise source may be random quantities. Note that $\mathbf{\Gamma}$ depends on z_s only through the initial conditions, $\mathbf{\Gamma}(0)$, and depends on r only through the diagonal matrix. Consistent with the earlier assumption that the noise source range and depth are i.i.d. random variables, it follows that upon averaging:

$$\langle r^{-1}\mathbf{\Gamma}(r)\rangle = \mathbf{U} \cdot \langle[\mathrm{diag}(r^{-1}e^{-\lambda_m r})]\rangle \cdot \mathbf{V}^\mathbf{T}\langle\mathbf{\Gamma}(0)\rangle. \tag{42}$$

Consider first the average over noise source depth. Define $f_z(z_s)$ as the probability density function in depth. Then the averaged elements in the $\mathbf{\Gamma}(0)$ matrix are

$$\langle \Gamma_m(0)\rangle \equiv \int_0^D \Gamma_m(0) f_z(z_s) dz_s = \int_0^D \Psi_m^2(z_s) f_z(z_s) dz_s, \tag{43}$$

where we recall that Ψ_m is the acoustic mode function used in Eq. (4). In the literature, it is often assumed that the noise sources are fixed at depth $\frac{1}{4}$ wavelength,[12]in which case $f_z(z_s)$ would be a delta function. The formulation Eq. (43) is more general and allows one to study the sensitivity of the final solution to the assumed depths of the noise sources. Consider next the average over noise source range

$$\langle r^{-1}e^{-\lambda_m r}\rangle = \int_0^\infty r^{-1}e^{-\lambda_m r} f_r(r) dr. \tag{44}$$

In lieu of other information, a reasonable assumption is that the noise sources are uniformly distributed in the horizontal (x,y) plane. Then the corresponding probability density function in range $f_r(r)$ has the form

$$f_r(r) = \begin{cases} n_r r, & r_{\min} < r < r_{\max} \\ 0, & \text{else.} \end{cases}, \tag{45}$$

Again to preserve some generality, Eq. (45) admits that the noise sources might be constrained to lie between some minimum and maximum ranges, $r_{\min}$ and $r_{\max}$, respectively. Combining the last two equations and evaluating the integral yields finally

$$\langle r^{-1} e^{-\lambda_m r} \rangle = \frac{n_r}{\lambda_m} [e^{-\lambda_m r_{\min}} - e^{-\lambda_m r_{\max}}]. \tag{46}$$

In the limiting case as $r_{\min} \to 0$ and $r_{\max} \to \infty$, the result reduces to n_r/λ_m.

To summarize, the objective has been to calculate the average output from a vertical array beamformer measuring ambient noise in the presence of internal waves. Transport theory has been used to model how the acoustic field from a noise source propagates through internal waves to the receiving array. The transport matrix **S** [Eq. (20)] contains elements that combine the acoustic and internal wave modes via Eq. (39). Averaging over the positions of the noise sources is accomplished in Eqs. (42)–(46). The result is substituted into Eq. (10), yielding the average beamformer output as a function of the steering direction θ. The specific interest is in the width and depth of the ambient noise notch.

E. Model implementation and testing

The primary computational burden associated with implementing the model is in calculating the transport matrix **S**. Recall that **S** is a symmetric matrix of size M where M is the number of acoustic modes retained in the calculation. To calculate **S**, it is necessary to evaluate numerous integrals of the form given in Eq. (29). The integrand involves two acoustic modes and one internal wave mode. Of specific interest is the internal wave mode evaluated at the resonant wave number for the two acoustic modes. Computationally, this means the internal wave modes must be recalculated for each unique pair of acoustic modes; if there are M acoustic modes, then the internal wave modes must be recalculated $M(M-1)/2$ times. Moreover, if J internal wave modes are retained, then there are $JM(M-1)/2$ integrals to calculate. At frequencies of a few kilohertz, M will be on the order of 100. If $J=20$, the number of integrals is on the order of 10^5.

Given the large number of integrals that must be calculated, it is not surprising that approximations have often been made in the literature to reduce the computational load. If both the background sound speed and buoyancy profiles are depth independent, all the modes can be expressed using sinusoids and the integrals evaluated analytically [23]. If more complicated profiles are used and the integrals are to be evaluated numerically, one approximation is to ignore mode coupling except for between adjacent pairs of modes;[36,18] this reduces the number of integrals to $J(\mathbf{M}-1)$. Another simplification is to retain only the first internal wave mode.[37]

In the present paper, the full calculation is made with all M acoustic modes and all J retained internal wave modes. This full calculation can be made computationally tractable. First, recognize that the acoustic modes must be calculated only once. The modes are all calculated on the same vertical grid with uniform fine spacing and stored. The internal waves must be calculated for the $M(M-1)/2$ pairs of acoustic modes, but a key point is that they can be calculated on a much coarser grid because invariably $J \ll M$. The internal wave modes are then splined onto the same fine grid as the acoustic modes. Finally, the integrals are evaluated using a simple closed-rule quadrature scheme. Order 10^5 integrals can be calculated on a laptop computer in a few minutes. Note also that the integrals are independent of both the internal wave strength bE_{GM} and the bottom loss. Consequently, the effect on the beamformer output of varying these two important parameters can be studied without having to recalculate the integrals.

Another important aspect to implementing the model is the size of the transport matrix **S**. As developed by Dozier and Tappert and used here, transport theory describes how the average intensity of each acoustic mode $\langle |A_m|^2 \rangle$ develops in range. Specifically neglected are cross terms $\langle A_m A_n^* \rangle$ where $m \neq n$. Were it necessary to include the cross terms, the size of **S** would increase from M to a potentially unmanageable $M(M+1)/2$. In the present application, the cross terms are neglected in going from Eq. (7) to Eq. (8). The validity of neglecting these terms can be tested, albeit for the special case of a range-independent medium, by comparing the average beamformer output as predicted by the present model to that predicted using a Monte Carlo approach that implicitly includes the cross terms. The calculation is also useful for defining the notch width and depth as the terms will be used when examining experimental data in Sec. III.

As a test case, 200 noise sources were randomly distributed in range between 0.5 and 25 km and in depth between 0.1 and 5.0 m. The simulated sound speed was 1520 m/s to depth 10 m and 1480 m/s below 30 m with a linear gradient connecting the two regions. While the sound speed profile in the water column is synthetic, the sea bed is modeled using parameter values similar to those observed in the East China Sea [38]. In the Monte Carlo calculation, the complex field at 1.48 kHz due to each noise source was calculated using a discrete mode sum and Eq. (15). The complex fields from the different sources were added at each element in a simulated 32-element receiving array with the combined field that was then beamformed using Eq. (1). Finally, the beamformed field was calculated for 100 different random realizations, and the result was then averaged. This Monte Carlo result was compared to the model result [Eqs. (10) and (11)] where modal cross terms are neglected. The two calculations for the simulated environment are compared in Fig. 2 using a common normalization.

The two calculations for the average beamformer output in Fig. 2 are quite similar for almost all beamformer steering directions θ. Both calculations show energy largely confined to a band of angles defined by the critical angle of the sediment. Within this band, both calculations show a notch at $\theta = 0$. Defining B_{up} as the local maximum looking up from horizontal (negative θ and B_{down} as the local maximum looking down from horizontal, both calculations have notch width $2\theta_0 \approx 28$ deg and notch depth $B_{\text{avg}}/B_{\text{notch}} \approx 20$ db. The

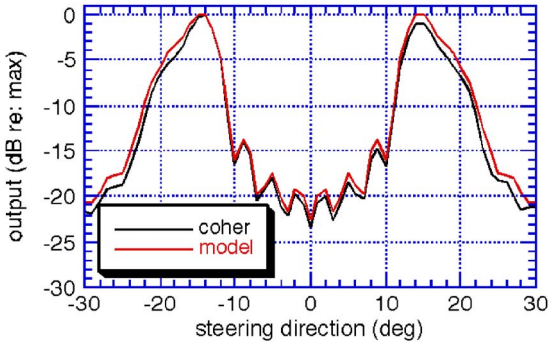


FIG. 2. (Color online) Evaluation of model for range-independent environment. Ambient noise beam pattern predicted by model compared to Monte Carlo calculation.

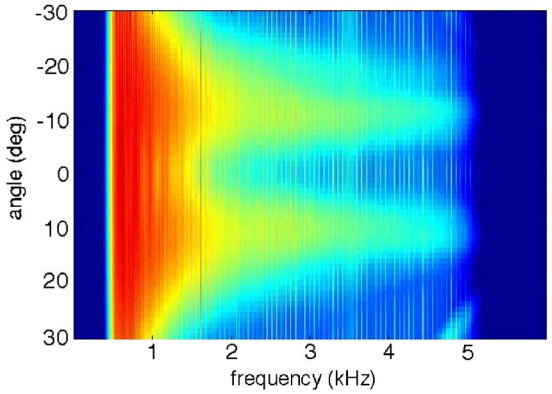


FIG. 3. (Color online) Example of ambient noise beam pattern observed in East China Sea. Result is average over 140 consecutive time windows each of 0.5-s duration. 50-dB dynamic range.

model calculation is symmetric, an even function of θ as would be expected from Eqs. (10) and (11). The Monte Carlo calculation, however, is slightly asymmetric with $B_{\text{up}} > B_{\text{down}}$. The ratio between B_{up} and B_{down} can be used to infer the bottom reflectivity.

On the basis of Fig. 2, the model will be used to estimate the width and depth of the noise notch. The calculations are made with the understanding that the model is unable to distinguish subtle differences between the up- and down-looking directions.

The noise notch in Fig. 2 is approximately 20 dB deep. For this range-independent environment, the notch depth is determined by the sidelobes of the beamformer. When internal waves are included in the calculation, the notch is substantially shallower as demonstrated in the following section.

III. RESULTS

In the previous section, a model was developed for the average beam formed ambient noise. The model can be used to predict both the depth and the width of the ambient noise notch in the presence of background internal waves. As a specific test, the model is now applied to data taken in the East China Sea.

Acoustic and environmental data were collected in the East China Sea (ECS) between 28 May and 9 June 2001 as part of the Asian Seas International Acoustics Experiment (ASIAEX).[20] The ECS is a particularly apt test for the model as the bathymetry is simple and the internal waves are relatively benign.[20,39] ASIAEX was designed as an acoustic propagation and reverberation experiment, but some ambient noise data were collected. In this section, ECS ambient noise data are analyzed in the 0.5-to-5-kHz band. The observed width and depth of the ambient noise notch is then compared to that predicted by the model. Finally, the internal wave strength is varied within the model with the resultant effect on the expected ambient noise notch quantified.

A. East China Sea noise data

East China Sea operations in ASIAEX were centered about 29° 40.67′ N and 126° 49.39′ E. The nominal upper mixed layer extended to depth 30 m in a nominally 105 m water column.

Ambient noise data were collected on a 31-element vertical array deployed from the R/V *Melville*. The deepest array element was 7.5 m above the seabed with 21.43 cm spacing between the elements. All the array elements were below the nominal thermocline. Noise data were sampled at 12 kHz, bandpass filtered between 0.5 and 5 kHz, and stored in 0.5-s-long sections. Nearby research vessels, numerous fishing boats, distant shipping, and the wind all contributed to the noise field. Other sections of data collected on the array were previously used to study the reverberation beam pattern.[5]

Figure 3 is a typical result showing the ambient noise beam pattern as a function of frequency and steering angle. Attention is restricted to ±30° of horizontal. The image is the result of averaging over 140 consecutive time windows, each 0.5 s in duration. The beam pattern calculated from individual 0.5-s time windows may fluctuate, but by averaging the image is smoothed. The noise notch at 0° is clearly visible over a broad frequency band as are the local maxima near ±10°. The local maxima above and below horizontal are designated as B_{up} and B_{down}, respectively. The angle spanned by the two maxima is the width of the noise notch.

Figure 4 shows how the beam pattern at 4 kHz evolves in time over 27 min. The data were taken on 6 June beginning at 12:00:45 local time. To show the effects of local shipping and beamformer grating lobes, the complete beam pattern between ±90° is shown. The figure also shows noise level in the notch and at the local maxima looking both immediately above and below the notch.

Several observations can be made from Fig. 4. The noise notch is stable over the duration of the measurement. The noise level in the notch (B_{notch}) is typically 3 to 5 dB down from the local maxima that are observed looking either above or below the horizontal. The noise peak looking above horizontal (B_{up}) is typically greater than noise peak looking below horizontal (B_{down}), consistent with the reflection loss from the seabed. The width of the noise notch is ~20° and changes little over the duration of the measurement. The strong noise level looking up at angles approaching −90° is presumably from the dynamic positioning of the R/V *Melville* stationed above the array. Because the array elements are spaced at greater than one-half wavelength for 4 kHz, the beamformer has significant grating lobes. As a

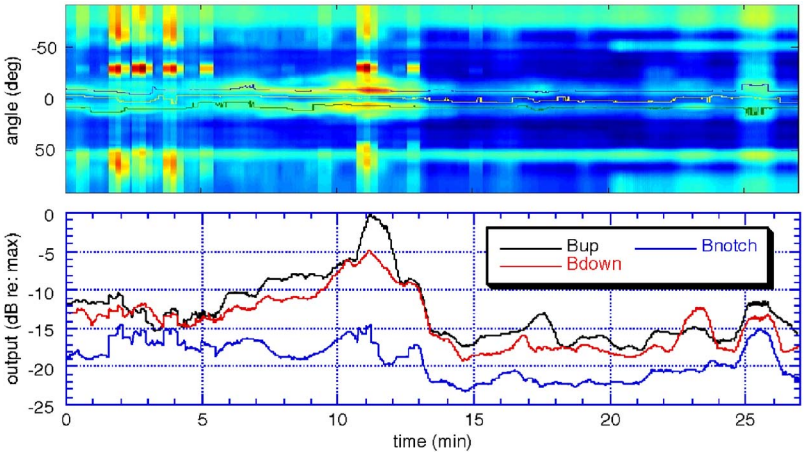


FIG. 4. (Color online) Top: Evolution of ambient noise beam pattern at 4 kHz. 30-dB dynamic range. Bottom: Maximum ambient noise level looking both above and below horizontal and minimum level in the ambient noise notch.

result, noise from the R/V *Melville* is also manifest in the anomalous noise peak near 50°.

The period between 9 and 13 min in Fig. 4 warrants comment. The high noise level apparent in many steering directions is believed due to a Chinese research vessel that approached within a few kilometers. While both B_{up} and B_{down} increase by several decibels during this period, B_{notch} is relatively unchanged. Presumably, the noise source is too close to the array for internal waves or other forms of variability to drive energy from the excited high-order acoustic modes into the low-order modes sampled by B_{notch}.

Figure 5 shows the ambient noise beam pattern at 2 and 3 kHz for the same 27-min period. The results in the vicinity of the notch are similar at all three frequencies.

The period between 14 and 24 min in Figs. 4 and 5 is relatively quiescent with no obvious local shipping activity. In Fig. 6, the width and depth of the notch at 2, 3, and 4 kHz is plotted for this quiescent period. (Note that beams were formed in 1-deg increments so the width of the notch takes integer values.) The results are stable. In general, the noise notch gets deeper and narrower as the frequency is increased from 2 to 4 kHz. At 3 kHz, the average notch depth is 4.7 dB and the average notch width is 19.7°.

Ambient noise data were also collected on June 7. A quiescent 7-min section of data beginning at 20:10:00 local time was studied in detail. Average values for the notch width and depth for the two dates are tabulated along with predictions made using the model in the following section.

B. Model/data comparison

As input to the model, certain environmental information must be either known or estimated. Extensive sediment data were collected as part of ASIAEX.[20,38] From these measurements, it is adequate for current purposes to represent the bottom as a single layer with sound speed 1600 m/s, density ratio 2.0, and attenuation 0.25 dB/m/kHz. The sound speed profile shown in Fig. 7 was measured in the water column and used to calculate the acoustic modes. To calculate the

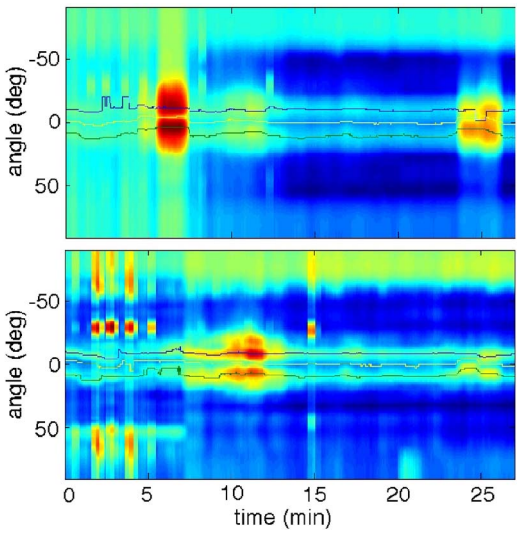


FIG. 5. (Color online) Evolution of ambient noise beam pattern. Top: 2 -kHz result with 48-dB dynamic range. Bottom: 3-kHz result with 30-dB dynamic range.

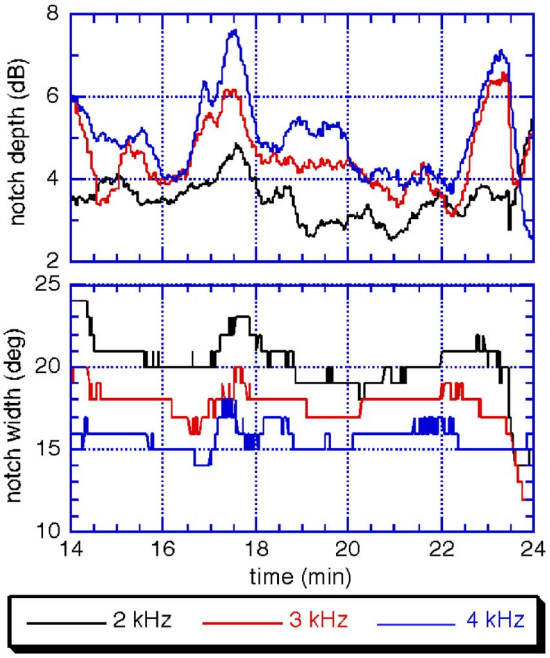


FIG. 6. (Color online) Depth (top) and width (bottom) of ambient noise notch. Result for the quiescent period between 14 and 24 min in Figs. 4 and 5.

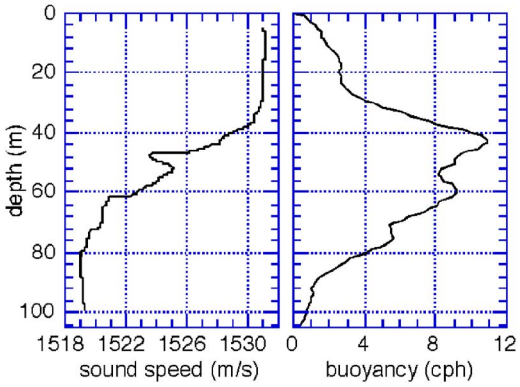


FIG. 7. (Color online) Left: Measured sound speed profile from East China Sea. Right: Corresponding estimated buoyancy frequency profile. Both profiles used as input to the model.

internal wave modes, the buoyancy frequency is needed. Ideally, the buoyancy profile would be known from direct synoptic measurement of density. As this information is not available, the buoyancy $N(z)$ is estimated from the measured sound speed profile c_0 using[26]

$$N^2(z) = -G^{-1}c_0^{-1}\left(\frac{\partial c_0}{\partial z} - \frac{\partial c_a}{\partial z}\right), \tag{47}$$

where recall $\partial c_a/\partial z$ is the gradient of the adiabatic sound speed [Eq. (22)]. To avoid large fluctuations in N, the measured c_0 is smoothed before estimating its derivative.[32] Equation (47) is an approximate relationship but should give an estimate for N accurate to within a factor of 2. The value $G=0.8$ s^2/m provided good agreement for the shallow water Coastal Mixing and Optics Experiment data and is used here. The resulting buoyancy profile used to calculate the internal wave modes is shown in Fig. 7. Integrating in depth [see Eq. (27)] gives $bN_0=0.86$ m/s, a reasonable value for shallow water.

The sole remaining parameter to describe the internal wave field is the energy level bE_{GM}. A value for bE_{GM} directly from ASIAEX measurements has not yet been determined. A reasonable estimate, however, may be gleaned from other experiments. The measured values for bE_{GM} summarized in Sec. II C vary from 0.08 to 0.5 m. Because the internal wave field observed in the East China Sea was relatively benign, the value $bE_{GM}=0.135$ m measured elsewhere by Kuperman and Jackson[34] is used. With this choice, the internal wave model is completely specified; there are no adjustable parameters in subsequent comparisons between the model and East China Sea data.

Figure 8 compares model and data for the beamformer output at 2, 3, and 4 kHz. The data curves represent vertical slices through Fig. 3 (7 June data) at the three frequencies. Each plot is individually normalized to the maximum output at that frequency. The model underpredicts the data for large steering angles because it does not include contributions from continuous modes.[40] Within ±10° of horizontal, however, the agreement between model and data is good. For each frequency, the model with no adjustable internal wave parameters accurately predicts the width and depth of the ambient noise notch.

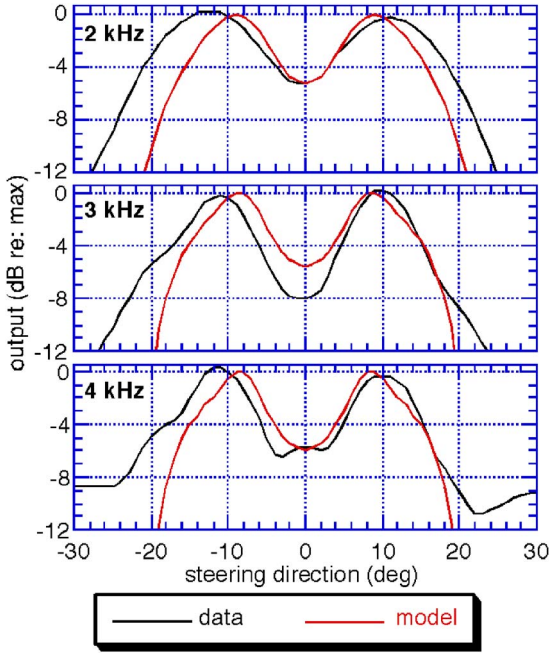


FIG. 8. (Color online) East China Sea model/data comparison of ambient noise notch at 2, 3, and 4 kHz. Data averaged over 140 consecutive time windows each of 0.5-s duration. See text for details of model calculation.

Table I compares the model to data averaged over longer time periods. Tabulated are the average width and depth of the noise notch at 2, 3, and 4 kHz as measured using the 6 June and 7 June data sets and as calculated using the model. Qualitatively, the model accurately predicts the tendency of the notch to grow deeper and narrower as frequency is increased. Quantitatively, the model accurately estimates both the width and depth of the noise notch. The agreement lends credence both to the predictive abilities of the model and to the hypothesis that internal wave effects on the ambient noise notch are important.

Additional details on the calculations: All results assumed the noise sources were radially distributed according to Eq. (45) with $r_{\min}=10$ m and $r_{\max}=50$ km. Varying $r_{\min}$ and $r_{\max}$ about these values had little effect on the calculated beamformer response. The sources were uniformly distributed in depth between 0.1 and 5.0 m. For noise sources far from the measurement array, the source depth is of little significance; in effect, these "initial conditions" are unimportant when the accumulated mode coupling induced by internal waves is substantial; see Eq. (43). The calculations re-

TABLE I. East China Sea model/data comparison. Width and depth of ambient noise notch as observed on two days compared to model prediction. Model assumes internal wave field of moderate strength.

	6 June data		7 June data		Model	
	Notch width (deg)	Notch depth (dB)	Notch width (deg)	Notch depth (dB)	Notch width (deg)	Notch depth (dB)
2 kHz	20.6±1.4	3.6	23.2±1.7	4.4	18.0	5.1
3 kHz	17.9±0.8	4.7	19.7±1.8	6.3	17.2	5.6
4 kHz	15.8±0.8	5.2	19.3±2.6	6.1	16.8	6.0

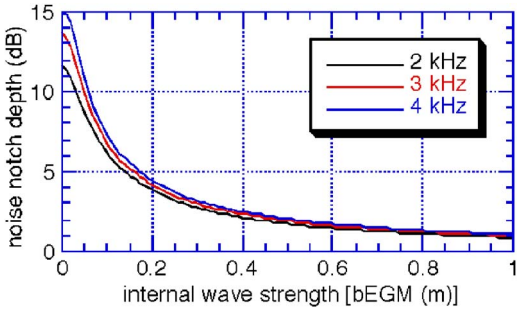


FIG. 9. (Color online) Effect of internal wave strength on the ambient noise notch depth.

tained $J=20$ internal wave modes. The number of acoustic modes M varied with frequency. At each frequency, M was set slightly greater than the number of modes nominally propagating in the water column; the extra sediment borne modes provide a necessary energy loss mechanism through coupling from the water borne modes.[41]

C. Effect of internal wave strength

As a sensitivity test, the depth of the noise notch was calculated as a function of the assumed internal wave strength. Figure 9 shows the results of the calculation at 2, 3, and 4 kHz. Keeping all other simulation parameters the same, the internal wave strength bE_{GM} was varied from 0 to 1.0 m.

Figure 9 lends credence both to the model and the underlying hypothesis that internal waves are responsible for the partial filling of the noise notch. If internal waves are neglected and $bE_{GM}=0$, the model predicts a noise notch that is too deep when compared to experimental observation. At the other extreme, assuming very strong internal waves, $bE_{GM}=1.0$ m, produces a noise notch that is too shallow. Assuming internal waves of reasonable strength, $bE_{GM}=0.135$ m, yields model results similar to the experimental data.

For the oceanographic data sets reviewed in Sec. II C, bE_{GM} varied between 0.08 and 0.5 m. Over this range of internal wave strength, Fig. 9 predicts that the depth of the noise notch could change by several decibels. The local strength of the internal waves could make the difference between whether or not a noise notch could be exploited in passive detection schemes.

IV. CONCLUDING REMARKS

A model has been developed for the ambient noise notch observed in shallow water at frequencies up to a few kilohertz. Assuming the presence of moderate-strength internal waves, the model accurately predicts both the width and depth of the noise notch observed in the East China Sea.

Detailed measurements of the internal wave field were not necessary to apply the model successfully; it sufficed to estimate the buoyancy frequency and to estimate the internal wave strength based on archival data. This is an encouraging result since it suggests that it may be possible to develop a predictive capability for the noise notch that would require only modest environmental information as input. Testing the model on data sets where there are synoptic acoustic noise and oceanographic internal wave measurements could develop further confidence in its validity. Analyzing longer sections of noise data might reveal periodicities that could be related to the time scales of the internal waves. Future work will involve additional data collection and model/data comparisons.

ACKNOWLEDGMENTS

This work was supported by ONR Code No. 321OA and SPAWAR. The authors thank Frank Henyey and Steve Reynolds of the University of Washington Applied Physics Laboratory and Patrick Ferat of the Johns Hopkins University Applied Physics Laboratory for several useful conversations. The anonymous reviewers also provided many helpful suggestions. Lewis Dozier furnished a copy of Ref. 22.

[1] A. V. Belousov and A. V. Furduev, "Estimation of the noise-level in the ocean in the angular interval of the refraction minimum," Sov. Phys. Acoust. **34**, 421–421 (1988).
[2] A. A. Aredov, N. N. Okhrimenko, and A. V. Furduev, "Anisotropy of the noise field in the ocean (experiment and calculations)," Sov. Phys. Acoust. **34**, 128–131 (1988).
[3] A. A. Aredov and A. V. Furduev, "Angular and frequency dependencies of the bottom reflection coefficient from the anisotropic characteristics of a noise field," Acoust. Phys. **40**, 176–180 (1994).
[4] M. J. Buckingham and S. A. S. Jones, "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 (1987).
[5] D. Tang, "Estimating shallow water bottom geo-acoustic parameters using ambient noise," in *Impact of Littoral Environmental Variability on Acoustic Predications and Sonar Performance*, edited by N. G. Pace and F. B. Jensen (Kluwer Academic, Dordrecht, 2002), pp. 147–154.
[6] C. H. Harrison and D. G. Simons, "Geoacoustic inversion of ambient noise: A simple method," J. Acoust. Soc. Am. **112**, 1377–1389 (2002).
[7] J. R. Short, "High-frequency ambient noise and its impact on underwater tracking ranges," IEEE J. Ocean. Eng. **30**, 267–274 (2005) [originally published in the US Navy J. Underwater Acoust. (1972)].
[8] N. O. Booth, R. Judd, and H. Bucker, "Measurement of vertical noise directionality with a mixed-polarity vertical array," IEEE J. Ocean. Eng. **28**, 537–543 (2003).
[9] R. Dashen and W. Munk, "Three models of global ocean noise," J. Acoust. Soc. Am. **76**, 540–554 (1984).
[10] C. Harrison, "Noise directionality for surface sources in range-dependent environments," J. Acoust. Soc. Am. **102**, 2655–2662 (1997).
[11] P. Ferat and J. Arvelo, "Mid- to high-frequency ambient noise anisotropy and notch-filling mechanisms," in *High Frequency Ocean Acoustics*, edited by M. B. Porter, M. Siderius, and W. A. Kuperman (AIP, New York, 2004), pp. 497–507.
[12] W. A. Kuperman and F. Ingenito, "Spatial correlation of surface generated noise in a stratified ocean," J. Acoust. Soc. Am. **67**, 1988–1996 (1980).
[13] W. M. Carey, R. B. Evans, J. A. Davis, and G. Botseas, "Deep-ocean vertical noise directivity," IEEE J. Ocean. Eng. **15**, 324–334 (1990).
[14] R. Evans and W. M. Carey, "Low-frequency array response to surface-generated noise," in *Proc. 5th Euro. Conf. Underwater Acoust.*, edited by M. E. Zakharia (EC, Brussels, 2000) pp. 693–698.
[15] T. C. Yang and A. Al-Kurd, "The internal wave effect on the noise vertical directionality in shallow water," in *OCEANS'96* (IEEE, New York, 1996), Vol. **1**, pp. 9–16.
[16] B. F. Kuryanov and B. I. Klyachin, "Application of the theory of radiation transport to problems of oceanic noise propagation," in *Problems of Oceanic Acoustics* (Nauka, Moscow, 1984), pp. 16–36.
[17] E. P. Kuznetsova, "Calculation of the directivity of noise in a shallow sea," Sov. Phys. Acoust. **35**, 627–630 (1989).
[18] B. G. Katsnelson and V. G. Petnikov, *Shallow-Water Acoustics* (Springer-Verlag, Berlin, 2002), Chap. 7.
[19] R. B. Evans, "Vertical directivity of ambient noise in the presence of internal waves," J. Acoust. Soc. Am. **115**, 2508 (2004).

[20]P. H. Dahl, R. Zhang, J. H. Miller, L. R. Bartek, Z. Peng, S. R. Ramp, J. X. Zhou, C. S. Chiu, J. F. Lynch, J. A. Simmen, and R. C. Spindel, "Overview of results from the Asian Seas International Acoustics Experiment in the East China Sea," IEEE J. Ocean. Eng. **29**, 920–928 (2004).
[21]L. B. Dozier and F. D. Tappert, "Statistics of normal mode amplitudes in a random ocean I. Theory," J. Acoust. Soc. Am. **63**, 353–365 (1978).
[22]L. B. Dozier, "A coupled-mode model for spatial coherence of bottom-interacting energy," in *Proceedings of the Stochastic Modeling Workshop*. edited by C. W. Spofford and J. M. Hayes (ARL-University of Texas, Austin, 1983) pp. 28-1–28-46.
[23]D. B. Creamer, "Scintillating shallow-water waveguides," J. Acoust. Soc. Am. **99**, 2825–2838 (1996).
[24]W. Munk, "Internal waves and small-scale processes," in *Evolution of Physical Oceanography*, edited by B. A. Warren and C. Wunch (MIT, Cambridge, MA, 1981), pp. 264–291.
[25]F. S. Henyey and A. Hoering, "Energetics of borelike internal waves," J. Geophys. Res. **102**, 3323–3330 (1997).
[26]S. M. Flatté (ed.), *Sound Transmission through a Fluctuating Ocean* (Cambridge Univ. P., Cambridge, 1979).
[27]K. B. Winters and E. A. D'Asaro, "Direct simulation of internal wave energy transfer," J. Phys. Oceanogr. **27**, 270–280 (1997).
[28]F. S. Henyey, D. Rouseff, J. M. Grochocinski, S. A. Reynolds, K. L. Williams, and T. E. Ewart, "Effect of internal waves and turbulence on a horizontal aperture sonar," IEEE J. Ocean. Eng. **22**, 270–280 (1997).
[29]M. Abromowitz and I. A. Stegun, *Handbook of Mathematical Functions* (NBS, Washington, DC, 1964), Eq. 23.1.18.
[30]R. C. Beardsley, K. H. Brink, S. R. Ramp, and R. Schlitz, "The Nantucket shoals flux experiment (NSFE 79). Part I: A basic description of the current and temperature variability," J. Phys. Oceanogr. **15**, 713–748 (1985).
[31]T. D. Dickey and A. J. Williams, "Interdisciplinary ocean process studies on the New England shelf," J. Geophys. Res. **106**, 9427 (2001).
[32]S. A. Reynolds and M. D. Levine, *DRI Internal Wave Simulations*. Tech. Memorandum APL-UW TM 5-04 (Applied Physics Laboratory, Seattle, 2005).
[33]D. Rouseff, "Effect of shallow water internal waves on ocean acoustic striation patterns," Waves Random Media **11**, 377–393 (2001).
[34]W. A. Kuperman and D. R. Jackson, "Ocean acoustics, matched-field processing and phase conjugation," in *Imaging of Complex Media with Acoustic and Seismic Waves*, edited by M. Fink, Top. Appl. Phys. **84**, 43–96 (2002).
[35]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, Orlando, 1980), Eq. 6.671.8.
[36]W. Kohler and G. Papanicolaou, "Wave propagation in a randomly inhomogeneous ocean," in *Wave Propagation in Underwater Acoustics*, edited by J. B. Keller and J. S. Papdakis (Springer-Verlag, New York, 1977), pp. 153–223.
[37]J. Yang, S. Zhou, J. X. Zhou, and J. F. Lynch, "Internal wave characteristics at the ASIAEX site in the east China Sea," IEEE J. Ocean. Eng. **29**, 1054–1060 (2004).
[38]J. W. Choi and P. H. Dahl, "Mid-to-high-frequency bottom loss in the east China Sea," IEEE J. Ocean. Eng. **29**, 980–987 (2004).
[39]S. R. Ramp, C-S Chui, F. L. Bahr, Y. Qi, P. H. Dahl, J. Miller, J. F. Lynch, R. Zhang, and J. Zhou, "The shelf-edge frontal structure in the central East China Sea experiment," IEEE J. Ocean. Eng. **29**, 1011–1031 (2004).
[40]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, New York, 1994), pp. 527–530.
[41]E. I. Thorsos, F. S. Henyey, W. T. Elam, S. A. Reynolds, and K. L. Williams, "Modeling shallow water propagation with scattering from rough boundaries," in *High Frequency Ocean Acoustics*, edited by M. B. Porter, M. Siderius, and W. A. Kuperman (AIP, New York, 2004), pp. 132–140.

A transformation of the environment eliminates parabolic equation phase errors

Irina I. Rypina,[a] Ilya A. Udovydchenkov, and Michael G. Brown
RSMAS/AMP, University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149

(Received 14 December 2004; revised 21 June 2006; accepted 22 June 2006)

It is shown that phase errors associated with the standard parabolic wave equation can be eliminated in a range-independent environment by appropriately transforming the environment. Solutions to the standard parabolic equation (PE) in the transformed environment are close to solutions to the Helmholtz equation in the original environment in the sense that ray and asymptotic mode contributions to the PE wave fields have no phase errors relative to their Helmholtz equation counterparts. Although phase errors are eliminated in the PE wave fields, complete asymptotic equivalence to the Helmholtz equation wave fields is not attained. Ray- and mode-based wave-field expansions are used to illustrate similarities and differences between PE and Helmholtz equation wave fields. Numerical simulations show excellent agreement between PE wave fields in the transformed environment and mode-based solutions to the Helmholtz equation in the true environment. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2227373]

PACS number(s): 43.30.Dr, 43.30.Cq [ADP] Pages: 1295–1304

I. INTRODUCTION

Solving the Helmholtz equation in realistic ocean models is a nontrivial task owing to the presence of range-dependent ocean structure. Three basic approaches to underwater acoustic modeling are used: ray expansions, normal mode expansions, and parabolic equation (PE) methods. Although ray-based models are fast and can easily handle range-dependence, the geometric approximation breaks down at low frequencies, caustics require special treatment, and computational difficulties, including those associated with ray chaos, arise at long range.[1] Normal mode expansions provide an exact solution to the range-independent problem and are convenient to use at lower frequencies where the number of propagating modes is moderate. Accounting for mode coupling in range-dependent environments is very challenging, however, especially at higher frequencies. These difficulties account for the popularity of PE methods which naturally handle range-dependence. PE methods were originally developed in the 1940s[2] and have been widely used in underwater acoustics since their introduction in the 1970s.[3] PE models neglect backscattered energy, but this is not a serious drawback in many underwater acoustic applications. A more serious shortcoming is that PE models introduce some geometric distortion to the wave field associated with the parabolic approximation. This geometric distortion, which can be quantified as phase errors,[4] can be significant.

In this paper it is shown that, by appropriately transforming a range-independent sound speed profile, phase errors associated with use of the standard parabolic wave equation are eliminated. Solutions to the standard PE in the transformed environment are close to solutions to the Helmholtz equation in the original environment in the sense that ray and asymptotic mode constituents of the PE wave fields have no phase errors relative to their Helmholtz equation counterparts. Use of the transformation eliminates PE phase errors for all propagation angles simultaneously. Validity of the transformation does not require small propagation angles or small sound speed variations. Our derivation of the environmental transformation that eliminates PE phase errors assumes range-independence, so it cannot be guaranteed that the transformation eliminates PE phase errors in range-dependent environments. However, the transformation is expected to reduce PE phase errors to a negligibly small level when applied to a sequence of profiles that define an environment with slow and/or weak range dependence. The transformation is most naturally applied as a preprocessing step prior to running an acoustic model, so no modification to an existing standard PE model is required to make use of the transformation. In underwater acoustic applications the transformation is probably best suited for use in long-range deep ocean applications, for reasons to be discussed in the following.

Before describing important background material and details of the transformation that eliminates PE phase errors, it is useful to point out why one should expect that such a transformation should exist. This explanation also provides some guidance as to how to find the desired transformation. The conceptual picture underlying the transformation that eliminates PE phase errors is shown in Fig. 1. This conceptual picture exploits, in a fundamental way, the action-angle description of ray motion in a range-independent environment. As described in the following, and indicated in Fig. 1, the Helmholtz equation (HE) ray equations in the true environment $c(z)$ can be written in action-angle form in terms of the Hamiltonian $H_{\mathrm{HE}}(I;c(z))$. The PE ray equations can also be written in action-angle form, but $H_{\mathrm{PE}}(I;c(z)) \neq H_{\mathrm{HE}}(I;c(z))$. We show in the following that it is possible to construct a different environment $\bar{c}(\bar{z})$ for which $H_{\mathrm{PE}}(I;\bar{c}(\bar{z}))=H_{\mathrm{HE}}(I;c(z))$. Indeed, the condition $H_{\mathrm{PE}}(I;\bar{c}(\bar{z}))$

[a]Electronic mail: irypina@rsmas.miami.edu

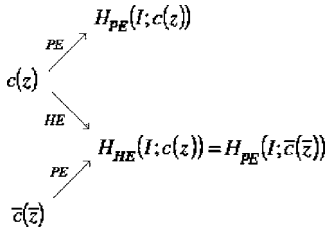


FIG. 1. Schematic diagram showing how the action-angle description of ray motion is exploited to define the desired sound speed transformation, from $c(z)$ to $\bar{c}(\bar{z})$.

$=H_{\mathrm{HE}}(I;c(z))$ defines the transformation from $c(z)$ to $\bar{c}(\bar{z})$. As described in the following, the equivalence of $H_{\mathrm{PE}}(I;\bar{c}(\bar{z}))$ and $H_{\mathrm{HE}}(I;c(z))$ guarantees that PE wave fields in $\bar{c}(\bar{z})$ have no phase errors relative to Helmholtz equation wave fields in $c(z)$. Hereafter $H_{\mathrm{HE}}(I;c(z))$ and $H_{\mathrm{PE}}(I;\bar{c}(\bar{z}))$ will be written $H(I)$. Explicit construction of $\bar{c}(\bar{z})$ is discussed in the following. The reason for writing the transformed environment $\bar{c}(\bar{z})$ rather than $\bar{c}(z)$ is that the domain of z generally changes under the transformation and source and receiver depths generally change under the transformation.

In Sec. II background material relating to the Helmholtz equation, PE equations, and the ray limits of these equations is reviewed. Included is a discussion of the action-angle description of ray motion. This material is then used in Sec. III to derive the transformation that eliminates standard PE phase errors. The effect of the transformation on ray travel times and geometric amplitudes is also discussed in Sec. III. In Sec. IV uniform asymptotic modal wave-field representations are considered: the PE wave field in the transformed environment is compared to the Helmholtz equation wave field in the original environment. Numerical results are presented and discussed in Sec. V. In the final section, our results are summarized and discussed.

II. BACKGROUND

A. Review of ray equations

Fixed-frequency (cw) acoustic wave fields satisfy the Helmholtz equation,

$$(\nabla^2 + k^2(\boldsymbol{r}))u = 0, \tag{1}$$

where $k^2(\boldsymbol{r})=\sigma^2/c^2(\boldsymbol{r})$. Here u is the acoustic pressure, σ is the acoustic wave angular frequency, and $c(\boldsymbol{r})$ is the sound speed. For simplicity we restrict our attention to two-dimensional problems $\boldsymbol{r}=(z,r)$, where z is depth and r is range; consistent with this assumption, azimuthal spreading factors are neglected throughout this paper. In the short wave limit or the so-called geometrical optics regime, when the acoustic wavelength $\lambda=2\pi/k$ is much smaller than the characteristic length scale of sound speed variations, the solution to Eq. (1) can be represented as a sum of rays, each representing a locally plane wave:

$$u(\boldsymbol{r},\sigma) = \sum_j A_j(\boldsymbol{r}) e^{i\sigma T_j(\mathbf{r})}. \tag{2}$$

Substituting the ray ansatz (2) into the Helmholtz equation (1), and equating terms of like order in σ, one obtains the eikonal and the transport equations,

$$(\nabla T)^2 = c^{-2}, \tag{3}$$

$$\nabla \cdot (A^2 \nabla T) = 0, \tag{4}$$

where for simplicity the subscript j on A and T has been omitted. Here $T=$const are surfaces of constant phase and $\nabla T=\boldsymbol{p}=(p_z,p_r)$ is the ray slowness vector. The solution to Eq. (3) can be constructed by solving a set of ray equations, which have Hamiltonian form. For directed wave propagation in which the spatial variable r increases monotonically following all rays of interest, r can be treated as the independent or time-like variable. In this case the ray equations can be written as a system of one-way Helmholtz ray equations,

$$\frac{dz}{dr} = \frac{\partial H}{\partial p}, \quad \frac{dp}{dr} = -\frac{\partial H}{\partial z}, \quad \frac{dT}{dr} = L = p\frac{dz}{dr} - H, \tag{5}$$

where

$$H(p,z,r) = -(c^{-2}(z,r) - p^2)^{1/2}. \tag{6}$$

Here $p \equiv p_z = \partial T/\partial z$, $H=-p_r$, and $dH/dr=\partial H/\partial r$. For notational simplicity the symbol p without a subscript is used to denote the z component of the slowness vector throughout this paper.

The ray equations (5), without specifying the form of Hamiltonian, correspond to a one-way wave equation of the form

$$i\sigma^{-1}\frac{\partial u}{\partial r} = H(\hat{p},z,r)u \tag{7}$$

with $\hat{p}=-i\sigma^{-1}\partial_z$. If $H(\hat{p},z,r)$ is taken to be $H_{\mathrm{HE}}(\hat{p},z,r)=-(c^{-2}(z,r)-\hat{p}^2)^{1/2}$ Eq. (7) is the one-way Helmholtz wave equation. Before proceeding we note that it is common at this point to introduce and focus on the equation satisfied by the envelope $\Psi(z,r)$ where $u(z,r)=e^{ik_0 r}\Psi(z,r)$ with $k_0=\sigma/c_0$ for some appropriately chosen constant c_0. Apart from a constant in $H(\hat{p},z,r)$ the equation satisfied by $\Psi(z,r)$ is the same as Eq. (7) satisfied by $u(z,r)$. For our purposes it is most natural to focus on Eq. (7) satisfied by $u(z,r)$ and define $H(\hat{p},z,r)$ consistent with that equation.

PE methods are based on choosing an operator $H_{\mathrm{PE}}(\hat{p},z,r)$ in Eq. (7) that is an approximation to $H_{\mathrm{HE}}(\hat{p},z,r)$. If $H_{\mathrm{PE}}(\hat{p},z,r)$ can be written as a sum of terms $K(\hat{p})$ (the kinetic energy term) and $U(z,r)$ (the potential energy term), the efficient split-step Fourier algorithm[5] can be used to numerically solve the equation. Since H_{PE} is an approximation to H_{HE} the solution to the parabolic wave equation is not identical to the solution to the one-way Helmholtz wave equation. The standard parabolic wave equation corresponds to

$$H_{\mathrm{PE}}(\hat{p},z,r)=\frac{c_0}{2}\hat{p}^2+U(z,r),\quad U(z,r)=-\frac{1}{c(z,r)}. \tag{8}$$

A more common choice of $U(z,r)$ in H_{PE} is $-1/c_0+(1-(c_0/c)^2)/(2c_0)$. Both this expression and our $U(z,r)$ in Eq. (8) are first-order accurate in the small parameter $\varepsilon=(c-c_0)/c_0$, but H_{PE} in Eq. (8), like the exact operator H_{HE}, has the advantage of reducing to $-1/c$ (independent of c_0) when $\hat{p}\to 0$. This simple property is important in the context of this paper because it allows the transformation derived in Sec. III to be applied in environments with multiple sound speed minima. Some of the details in Sec. III depend on the choice of $U(z,r)$ in H_{PE}; the formulas given apply to the choice $U=-1/c$, as in Eq. (8).

When the sound speed is a function of depth only, $c=c(z)$, the system of ray equations (6) constitutes an integrable Hamiltonian system. Under this assumption the phase-space trajectories are constrained to lie on closed invariant curves in the phase plane (p,z). The motion is thus periodic and returns to the same point in (p,z)-space after a characteristic ray cycle distance $R=2\pi/\omega$, where ω is the spatial "frequency" of the motion. The analysis of such a system is most convenient in terms of action-angle variables. These are conjugate variables (I,θ) such that I is constant following each trajectory and the conjugate coordinate θ increases by 2π after each complete period of motion. The new Hamiltonian depends only on I, $H=H(I)$. Details of the transformation $(p,z)\to(I,\theta)$, $H(p,z)\to H(I)$ are well known (see, e.g., Landau and Lifshitz;[6] Virovlyansky;[7] or Beron-Vera and Brown[8]). Relevant results are

$$p=\frac{\partial}{\partial z}\mathrm{G}(z,I), \tag{9a}$$

$$\theta=\frac{\partial}{\partial I}\mathrm{G}(z,I), \tag{9b}$$

where

$$G(z,I)=\int^{z}p(H(I),z')dz' \tag{10}$$

is the generating function for the transformation, and

$$I=\frac{1}{2\pi}\oint p(H,z)dz. \tag{11}$$

In the transformed system the ray equations are

$$\frac{d\theta}{dr}=\frac{\partial H}{\partial I}\equiv\omega(I),\quad \frac{dI}{dr}=-\frac{\partial H}{\partial\theta}=0 \tag{12}$$

and the corresponding travel time equation is

$$\frac{dT}{dr}=I\omega-H+\frac{d}{dr}(G-I\theta). \tag{13}$$

Note that we are making no notational distinction between H in $H(p,z)$ and $H(I)$. The reason is that the transformation from $H(p,z)$ to $H(I)$ leaves unchanged the numerical values of H and its physical interpretation as minus the r component of the slowness vector.

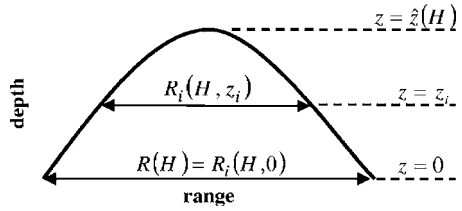


FIG. 2. Schematic diagram showing how the complete ray cycle distance $R(H)=R_i(H,0)$ and incomplete ray cycle distance $R_i(H,z_i)$ are defined for $z\geq 0$, the region above the sound channel axis. Note that H is a ray label, $\hat{z}(H)$ is defined by the condition $c(\hat{z})=-1/H$, and $R_i(H,\hat{z}(H))=0$.

The depth limits on the integrals defining I [Eq. (11)] and G [Eq. (10)] depend on the qualitative features of the sound speed profile. In environments with a single sound speed minimum it will be necessary in the following section to transform the regions above and below the sound channel axis separately to eliminate up/down ambiguity. With this in mind, we shall define a ray cycle to start and end at the sound speed minimum (taken to be $z=0$, for convenience) with a single turn at the midpoint (see Fig. 2). Then, for upper loops $(0\leq z\leq\hat{z}(H))$, explicit expressions for I and G are

$$I=\frac{1}{\pi}\int_0^{\hat{z}(H)}p(H,z)dz \tag{14}$$

and

$$G(z,I)=\pi I-\int_z^{\hat{z}(H)}p(H,z')dz', \tag{15}$$

where $p=0$ at $\hat{z}(H)$. For both choices of $H(p,z)$ [Eqs. (6) and (8)] considered in this paper $\hat{z}(H)$ is defined by the condition $c(\hat{z}(H))=-1/H$. Note that p changes from positive to negative at the upper turning depth $\hat{z}$, so G increases monotonically from 0 at $z=0$, to πI at $z=\hat{z}$, to $2\pi I$ at $z=0$. Note also that each time a ray completes a full cycle, θ increases by 2π and G increases by $2\pi I$, so the end-point correction term in Eq. (13), $G-I\theta$, oscillates about zero.

B. Complete and partial ray cycle distances, and the generating function

The complete ray cycle distance $R(H)$ for the system of ray equations (5) is defined as the distance in the horizontal direction that a ray travels during one complete cycle of motion,

$$R(H)=\oint dr=\oint\frac{dz}{dz/dr}=2\int_0^{\hat{z}(H)}\frac{dz}{\partial H/\partial p}. \tag{16}$$

Note that the bounds on the final integral are consistent with our earlier discussion. It is convenient to introduce the incomplete ray cycle distance (see Fig. 2),

$$R_i(H,z)=2\int_z^{\hat{z}(H)}\frac{dz}{\partial H/\partial p}, \tag{17}$$

where the lower limit of integration $0\leq z\leq\hat{z}(H)$. A simple connection between generating function and ray cycles can be obtained from Eq. (15) with use of Eqs. (12) and (17):

$$\frac{\partial G}{\partial H} = \pi \frac{\partial I}{\partial H} - \int_z^{\hat{z}(H)} \frac{\partial p}{\partial H} dz' = \frac{1}{2} R(H) - \frac{1}{2} R_i(H,z)$$

$$= \frac{1}{2} R_i(H,0) - \frac{1}{2} R_i(H,z). \tag{18}$$

Also from Eq. (15) it can be seen that

$$I(H) = \frac{1}{\pi} G(\hat{z}(H), I). \tag{19}$$

III. A TRANSFORMATION OF THE ENVIRONMENT ELIMINATES STANDARD PE PHASE ERRORS

A. A preliminary result

Inspection of the ray and travel time equations, written in terms of action-angle variables, Eqs. (12) and (13), shows that if $G(z,I)$ and $H(I)$ are the same for a PE and a Helmholtz system, then ray paths and travel times are identical in the two systems. This observation sharpens the heuristic argument given in Sec. I, which did not account for $G(z,I)$. The theorem that follows recasts the problem in a way that facilitates the construction of the transformed environment $\bar{c}(\bar{z})$. The construction of $\bar{c}(\bar{z})$ is discussed separately.

Theorem: Consider the systems described by

$$H_{\mathrm{HE}}(p,z) = -\sqrt{c^{-2}(z) - p^2} = -\frac{1}{c(\hat{z})} \tag{20a}$$

and

$$H_{\mathrm{PE}}(p,\bar{z}) = \frac{c_0}{2} p^2 - \frac{1}{\bar{c}(\bar{z})} = -\frac{1}{\bar{c}(\hat{\bar{z}})}, \tag{20b}$$

where $c(z)$, $z \geq 0$, and $\bar{c}(\bar{z})$, $\bar{z} \geq 0$, are monotonically increasing continuous functions, $c(0) = \bar{c}(0)$, and $\hat{z}(H_{\mathrm{HE}})$ and $\hat{\bar{z}}(H_{\mathrm{PE}})$ are turning depths. For both systems, Eqs. (14)–(19) define I, G, R, and R_i, and the ray and travel time equations are Eqs. (12) and (13). Assume that depths transform according to the rule

$$c(z) = \bar{c}(\bar{z}). \tag{20c}$$

Then $H(I)$ is the same for both systems and $G_{\mathrm{HE}}(z,I) = G_{\mathrm{PE}}(\bar{z},I)$ if and only if the sound speed transformation between c and $\bar{c}$ is constrained to satisfy

$$R_{i,\mathrm{HE}}(H_{\mathrm{HE}},z) = R_{i,\mathrm{PE}}(H_{\mathrm{PE}},\bar{z}), \quad \forall\, z \in [0,\hat{z}], \bar{z} \in [0,\hat{\bar{z}}]. \tag{20d}$$

Proof: Because $H_{\mathrm{HE}} = -1/c(\hat{z})$, $H_{\mathrm{PE}} = -1/\bar{c}(\hat{\bar{z}})$, $c(0) = \bar{c}(0)$, and Eq. (20c) is satisfied, the range of H values for both systems is the same. The $z=0$, $\bar{z}=0$ limit of Eq. (20d) reduces to $R_{\mathrm{HE}}(H) = R_{\mathrm{PE}}(H)$. For both systems $dH/dI = \omega = 2\pi/R$, so equivalence of $R(H)$ for the two systems guarantees that $H(I)$ for the two systems differ only by an integration constant. Because the range of H values for the two systems is the same, the integration constant is zero: note that $H(I) \to 1/c(0)$ as $I \to 0$ for both systems. For $z > 0$, $\bar{z} > 0$ the same argument, together with Eq. (18), dictates that $G_{\mathrm{HE}}(z,I)$ and $G_{\mathrm{PE}}(\bar{z},I)$ differ only by an integration constant. In the limit $z \to \hat{z}$ both G_{HE} and G_{PE} approach πI, so the integration constant is zero and $G_{\mathrm{HE}}(z,I) = G_{\mathrm{PE}}(\bar{z},I)$. The same arguments apply in reverse by differentiation with respect to I of $H(I)$ and $G(I,z)$.

Remarks: According to this theorem PE rays and travel times in the environment $\bar{c}(\bar{z})$ are identical to Helmholtz equation rays and travel times in the environment $c(z)$ if and only if a transformation from $c(z)$ to $\bar{c}(\bar{z})$ that satisfies condition (20d) exists. The problem of finding a transformation that satisfies condition (20d) will be discussed next. The condition $c(z) = \bar{c}(\bar{z})$ dictates that the sound speeds at the acoustic source and receiver do not change as part of the transformation. This dictates that, in general, source and receiver depths must change as part of the transformation. Stated somewhat differently, the theorem guarantees that, if the conditions of the theorem are met, PE rays and travel times from $(\bar{z}_s, 0)$ to $(\bar{z}_r, r)$ in the transformed environment $\bar{c}$ are identical to Helmholtz equation rays and travel times from $(z_s, 0)$ to (z_r, r) in the original environment c, where $c(z_s) = \bar{c}(\bar{z}_s)$, $c(z_r) = \bar{c}(\bar{z}_r)$.

B. Abel transform pairs and the sound speed profile transformation

Abel transform pairs provide the basis for the desired sound speed transformation. One form of an Abel transform pair is well known to seismologists, as this transform pair forms the basis of the famous Herglotz-Wiechert inversion formula. Consistent with our notation, this transform pair is[9]

$$R_i(H,z_i) = -2H \int_{z_i}^{\hat{z}(H)} \frac{dz'}{\sqrt{c^{-2}(z') - H^2}} \tag{21}$$

and

$$z(c) - z_i = \frac{1}{\pi} \int_{-1/c(z_i)}^{-1/c} \frac{R_i(H,z_i)}{\sqrt{H^2 - c^{-2}}} dH, \tag{22}$$

where $0 \leq z_i < \hat{z}(H)$ and $z_i \leq z(c) < \hat{z}(H)$. This transform pair arises naturally in a Helmholtz-equation-based description of ray motion. A similar transform pair applies to a standard PE-based description of ray motion:[9]

$$R_i(H,\bar{z}_i) = \sqrt{\frac{2}{c_0}} \int_{\bar{z}_i}^{\hat{\bar{z}}(H)} \frac{dz'}{\sqrt{H - U(z')}} \tag{23}$$

and

$$\bar{z}(U) - \bar{z}_i = \frac{1}{\pi} \sqrt{\frac{c_0}{2}} \int_{U(\bar{z}_i)}^{U} \frac{R_i(H,\bar{z}_i)}{\sqrt{U - H}} dH, \tag{24}$$

where $0 \leq \bar{z}_i < \hat{\bar{z}}$ and $\bar{z}_i \leq \bar{z}(U) < \hat{\bar{z}}(H)$. Note that no notational distinction is made between H in Eqs. (21), (22) and H in Eqs. (23), (24) as the equivalence of these quantities is central to the argument underlying the desired sound speed transformation. For the same reason, no notational distinction is made between R_i in Eqs. (21), (22) and R_i in Eqs. (23), (24).

Equations (21) and (22) define a valid transform pair for any $z_i \in [0,\hat{z}]$, and similarly Eqs. (23) and (24) define a valid transform pair for any $\bar{z}_i \in [0,\hat{\bar{z}}]$. An important question re-

mains, however: Are the reconstructed functions $z(c)-z_i$ and $\bar{z}(U)-\bar{z}_i$ independent of z_i and $\bar{z}_i$, respectively? Unfortunately, the answer is that, in general, the reconstructed functions depend on the choice of z_i and $\bar{z}_i$. This means that in general it is not possible to satisfy the conditions of the theorem for all z_i and $\bar{z}_i$ simultaneously. Thus, according to the theorem, exact equivalence between PE rays and travel times in the transformed environment and Helmholtz equation rays and travel times in the original environment is, in general, not achievable. With these comments in mind, we proceed after making the choice $z_i=0$ in Eqs. (21), (22), and $\bar{z}_i=0$ in Eqs. (23), (24). This choice guarantees the equivalence of $H(I)$ in the two systems, but does not, in general, satisfy the condition $G_{\mathrm{HE}}(z,I)=G_{\mathrm{PE}}(\bar{z},I)$. This compromise means that it is not possible, in general, to achieve complete asymptotic equivalence between PE and Helmholtz equation wave fields using our transformation. We show in the following, however, that the transformation does succeed in eliminating PE phase errors.

The desired transformation is described by Eq. (21) with $z_i=0$ and (24) with $\bar{z}_i=0$: using the true $c(z)$, $R(H)$ is computed using Eq. (21) from which $\bar{z}(U(\bar{c}))$ is constructed using Eq. (24). Inverting $\bar{z}(U(\bar{c}))$ gives $\bar{c}(\bar{z})$. [Note that Eqs. (22) and (23) are not needed. These equations could be used, for example, to find the Helmholtz equation equivalent to a sound speed structure used in a standard PE simulation.] The transformation is unique provided the monotonicity and continuity conditions stated in the theorem are satisfied. As noted earlier, this dictates that the regions above and below the sound channel axis are transformed separately. Environments with submerged ducts can be handled by transforming the regions internal to and external to the duct separately; this requires some care. To treat surface reflected energy $R(H)$ in Eq. (21) must be properly defined. For instance, in an environment with a deep sound speed excess that supports RR and RSR propagation, $R(H)$ in the H domain corresponding to RSR rays must be defined as twice the range integral over depth from the lower turning depth to the surface. Practical issues relating to use of the transformation are discussed in Sec. V.

Consider a purely refracted ray in a deep ocean environment with only one sound speed extremum, a minimum at the sound channel axis (at $z=\bar{z}=0$), and assume that the ray is emitted from an axial source. In such an environment the regions above and below the sound channel axis are transformed separately using Eqs. (21) and (24). Let $\Delta T(r;H)=T_{\mathrm{PE}}(\bar{z}(r;H))-T_{\mathrm{HE}}(z(r;H))$ denote the travel time error associated with the transformation following the ray whose Hamiltonian is H (the same value in both environments). The transformation guarantees that $\Delta T(r;H)$ has four zero crossings within each complete ray loop—one at each of two axis crossings, one at the upper turning depth $\hat{z}$ and one at the lower turning depth $\check{z}$. This behavior is illustrated in Fig. 4. Furthermore, the transformation guarantees that there is no secular (in r) growth of $\Delta T(r;H)$ for all values of the ray label H. In this sense, the transformation eliminates phase errors. This definition of the absence of phase errors may seem arbitrary. Note, however, that because any travel time error over a complete ray cycle will be N times as large after N cycles, travel time errors in a range-independent environment grow linearly in range on average. Thus, the only sensible way to define phase errors is to take the large r limit of $\Delta T(r;H)/r$. Using the transformation that we have described this limit is zero. An alternative argument that leads to the conclusion that phase errors are eliminated by the transformation is presented in Sec. IV. There it is shown that asymptotic modal phase errors are eliminated by the transformation, i.e., that PE modal phases in the transformed environment are identical to Helmholtz equation modal phase errors in the original environment. We emphasize that the properties just described are guaranteed for arbitrary steep rays, independent of the magnitude of the sound speed variations. Note, however, that, although phase error elimination holds generally, maximum travel time errors (recall Fig. 4) are expected to be smaller when sound speed variations and propagation angles are small. We have not attempted to quantify this dependence.

The effect of the transformation on geometric amplitudes is also of interest. The transport equation corresponding to both the one-way Helmholtz equation and the standard PE [both are special cases of Eq. (7)] can be written

$$\frac{\partial A^2}{\partial r}+\frac{\partial}{\partial z}\left(A^2\left.\frac{\partial H}{\partial p}\right|_{p=\partial T/\partial z}\right)=0. \tag{25}$$

Derivation of this equation for $H=H_{\mathrm{PE}}$, Eq. (8), is straightforward, while considerable care is required to derive this equation for $H=H_{\mathrm{HE}}$, Eq. (6). In the latter case the square root operator must be expanded in powers of $(c\hat{p})^2$ and Eq. (25) must be established by induction (see also Ref. 10). For both choices of H, the solution to Eq. (25) following a ray can be written

$$A^2(r)=A^2(r_0)\frac{dz(r_0)}{dz(r)} \tag{26}$$

where $dz(r_0)$ and $dz(r)$ are infinitesimal ray separations at fixed ranges. An interesting observation is that the one-way Helmholtz form of Eq. (26) is guaranteed to agree with the solution to the two-way Helmholtz transport equation (4) only in a range-independent environment. (This was first pointed out in Ref. 11.) With regard to the transformation we have described, it follows from Eq. (26) that PE geometric amplitudes in the transformed environment are identical to one-way Helmholtz geometric amplitudes in the original environment if $(d\bar{z}(r)/d\bar{z}(r_0))_{\mathrm{PE}}=(dz(r)/dz(r_0))_{\mathrm{HE}}$ at all points following a ray. This condition would be satisfied if the transformation were constrained to satisfy $G_{\mathrm{HE}}(z,I)=G_{\mathrm{PE}}(\bar{z},I)$. But the latter condition is not, in general, satisfied, so in general it is expected that $(d\bar{z}(r)/d\bar{z}(r_0))_{\mathrm{PE}}\neq(dz(r)/dz(r_0))_{\mathrm{HE}}$ and small differences in geometric amplitudes can be anticipated in the PE wave fields in the transformed environment.

In environments with density variations $\rho(z)$, a simple argument based on the two-way Helmholtz equation-based transport equation (4) can be used to determine how $\rho(z)$ is transformed. In the presence of density variations $\nabla^2 u$ in Eq. (1) is replaced by $\rho\nabla\cdot((1/\rho)\nabla u)$ and the transport equation (4) is replaced by $\nabla\cdot((A^2/\rho)\nabla T)=0$, whose solution depends

only on the ratio of the densities at the receiver and source. To ensure that this ratio be the same in the transformed PE system as in the original Helmholtz equation system for all possible source/receiver depth combinations, densities must transform according to the rule $\rho(z)=\bar{\rho}(\bar{z})$. Together with the condition $c(z)=\bar{c}(\bar{z})$, this dictates that the density that accompanies each sound speed be the same before and after transformation.

IV. ASYMPTOTIC MODE THEORY

To provide additional insight into the transformation that we have introduced, we now consider modal expansions of the wave field. To make our discussion concrete, we consider uniform asymptotic modal expansions of wave fields excited by a time-harmonic point source in an environment in which $c(z)$ has a single minimum. The PE modal expansion in the transformed environment $\bar{c}(\bar{z})$ is compared to the Helmholtz equation modal expansion in the original environment $c(z)$. These modal expansions in the far field have the form

$$u_{\mathrm{HE}}(z,r;\sigma)=\sum_{m=0}^{\infty}\phi_m^{\mathrm{HE}}(z;\sigma)\phi_m^{\mathrm{HE}}(z_s;\sigma)\exp(i\sigma p_m^{\mathrm{HE}}r) \tag{27}$$

and

$$u_{\mathrm{PE}}(\bar{z},r;\sigma)=\sum_{m=0}^{\infty}\phi_m^{\mathrm{PE}}(\bar{z};\sigma)\phi_m^{\mathrm{PE}}(\bar{z}_s;\sigma)\exp(i\sigma p_m^{\mathrm{PE}}r), \tag{28}$$

where z_s is the source depth and $\bar{z}_s$ is the transformed source depth, and the ϕ_m's satisfy

$$\frac{d^2\phi_m}{dz^2}+\sigma^2p^2\phi_m=0 \tag{29}$$

for the appropriate choice of p^2:

$$(p^{\mathrm{HE}})^2=c^{-2}(z)-(p_m^{\mathrm{HE}})^2, \tag{30}$$

$$(p^{\mathrm{PE}})^2=\frac{2}{c_0}(-p_m^{\mathrm{PE}}-U(\bar{z})). \tag{31}$$

Assume that $c(z)$ has a single minimum at $z=0$. [It follows that $\bar{c}(\bar{z})$ also has a single minimum at $\bar{z}=0$.] Then, consistent with our asymptotic analysis, we assume that the ϕ_m's satisfy the boundary conditions $\phi_m\to 0$ as $z\to\pm\infty$.

Explicit construction of a uniform asymptotic solution to Eq. (29) [which applies to both choices of p^2, Eqs. (30) and (31)] is discussed in Ref. 12. The solution is

$$\phi_m(z;\sigma)=(dS(z;p_m)/dz)^{-1/2}2^{-m/2}e^{-(m+1/2)S^2(z;p_m)}\times H_m((2m+1)^{1/2}S(z;p_m)), \tag{32}$$

where H_m are Hermite polynomials and $S(z;p_m)$ is a scaled depth coordinate which for $\check{z}(p_m)\le z\le\hat{z}(p_m)$ satisfies

$$\int_{-1}^{S}(1-\eta^2)^{1/2}d\eta=\frac{\pi}{2J(p_m)}\int_{\check{z}(p_m)}^{z}p(z')dz', \tag{33}$$

where

$$J(p_m)=\frac{1}{\pi}\int_{\check{z}(p_m)}^{\hat{z}(p_m)}p(z)dz \tag{34}$$

and p_m is constrained to satisfy the quantization condition

$$\sigma J(p_m)=m+\tfrac{1}{2},\quad m=0,1,2,\ldots\ . \tag{35}$$

For $z<\check{z}(p_m)$ and $z>\hat{z}(p_m)$ (in the evanescent tails of the modes) a modified form of the relationship (31) that defines $S(z;p_m)$ applies.

The structure of the modes and the modal phases in (27) and (28) will now be compared. We begin by considering the phases $\sigma p_m^{\mathrm{HE}}r$ and $\sigma p_m^{\mathrm{PE}}r$. Note that in both cases p_m is a discrete value of the r component of the slowness vector, $p_r(=-H)$, that is constrained by Eqs. (34) and (35). $J(H)$ defined in Eq. (34) is the sum of $I(H)$ for upper and lower ray loops. But $I(H)$ for both upper and lower loops is invariant under the transformation, so $J(H)$ is also invariant under the transformation. It follows that $p_m^{\mathrm{HE}}=p_m^{\mathrm{PE}}$, so the phases in expressions (27) and (28) are identical. This important observation supports our claim that the transformation that we have described eliminates PE phase errors. Asymptotic modal phases are simpler than ray phases inasmuch as the former depend only on properties of complete ray cycles (double loops) which are correctly constrained by our transformation.

The depth structure of the PE and HE modes has the same general form (32), so $\phi_m^{\mathrm{HE}}(z;\sigma)=\phi_m^{\mathrm{PE}}(\bar{z};\sigma)$ if $S_{\mathrm{HE}}(z;p_m)=S_{\mathrm{PE}}(\bar{z};p_m)$. But $S_{\mathrm{HE}}(z;p_m)=S_{\mathrm{PE}}(\bar{z};p_m)$ if and only if $G_{\mathrm{HE}}(z,I)=G_{\mathrm{PE}}(\bar{z},I)$ for both upper and lower ray loops and we have seen that our transformation does not, in general, satisfy the latter constraint. This means that while $\phi_m^{\mathrm{HE}}(z;\sigma)$ and $\phi_m^{\mathrm{PE}}(\bar{z};\sigma)$ are very similar, they are not identical; slight shifts in the location of zero crossings and local extrema of $\phi_m^{\mathrm{HE}}(z;\sigma)$ and $\phi_m^{\mathrm{PE}}(\bar{z};\sigma)$ are generally present, but their turning speeds are identical.

An argument similar to the one just given demonstrates the asymptotic equivalence of the HE- and standard PE-based modal phases in $c(z)$ and $\bar{c}(\bar{z})$, respectively, for environments that support modes with a single turning point [$c(z)$ increasing monotonically away from a pressure release surface at $z=0$]. For that problem, results similar to the ones just presented hold.[13] Modified forms of Eqs. (34) and (35) apply, and the modes are Airy functions of a stretched vertical coordinate $S(z)$ which is defined by a relationship similar to Eq. (33). As in the above-presented analysis, $I=I(p_m)$ satisfies a frequency-dependent quantization condition, and equivalence of HE and standard PE modal phases follows from equivalence of $H(I)$ before and after transformation. Again, the stretched vertical coordinate $S(z)$ is closely related to $G(z,I)$. Because our transformation does not enforce the constraint $G_{\mathrm{HE}}(z,I)=G_{\mathrm{PE}}(\bar{z},I)$ slight differences in mode structure of the above-noted type are, in general, present.

V. NUMERICAL RESULTS

In this section both ray and full-wave (normal mode and standard PE) simulations are presented which demonstrate the utility of the transformation that we have described. Four steps are involved. The first step is to transform the environ-

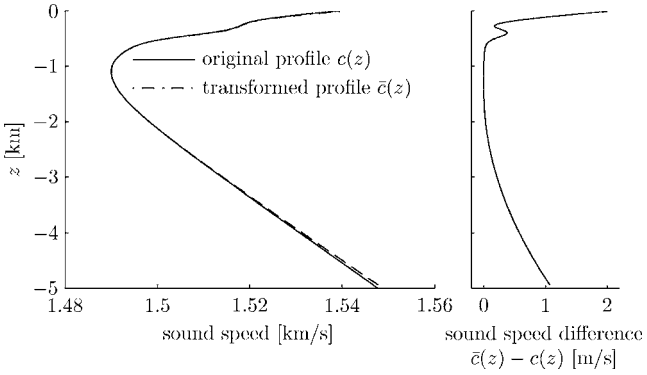


FIG. 3. Left panel: original and transformed sound speed profiles. Right panel: difference between original and transformed sound speed profiles.

ment using Eqs. (21) and (24). In this process it is important that $R(H)$ be defined using ray segments that are periodically repeated. Often this process requires piecewise construction of $R(H)$. Examples involving the deep ocean sound channel, secondary submerged ducts, and RSR paths were noted earlier. Transforming the environment generally changes the depth domain of the problem. The second step is transformation of source and receiver depths using $c(z)=\bar{c}(\bar{z})$; the sound speed at the source and receivers are unchanged, but their depths generally change. The third step is to perform PE modeling in the transformed environment. A fourth step is required if one wishes to present acoustic field plots on an undistorted depth grid; one must then undo the vertical distortion associated with the condition $c(z)=\bar{c}(\bar{z})$—and thereby simultaneously restore the original depth domain.

The simulations shown were performed in range-independent environments. Prior to transformation $c(z)$ was a perturbed canonical profile:[14]

$$c(z) = c_M(z) + dc \cdot \exp\left(-\frac{1}{2}\left(\frac{z_c - z}{z_w}\right)^2\right),$$

where $c_M(z)=c_a(1+\varepsilon(e^{\eta}-\eta-1))$ with $\eta=2(z-z_a)/B$. Here depth z decreases upward, $z_a=-1.1$ km is the sound channel axis depth, $B=1.0$ km is a thermocline depth scale, $c_a=1.49$ km/s is the sound speed on the channel axis, $\varepsilon=0.0057$ is a dimensionless constant, $z_c=-0.35$ km is the depth of the midpoint of the Gaussian perturbation, $z_w=0.1$ km is the width of the Gaussian perturbation, and $dc=0.008$ km/s is the maximum amplitude of the Gaussian perturbation. The original and transformed sound speed profiles are shown in Fig. 3.

Figure 4 shows travel time error associated with the transformation following a steep purely refracted ray path,

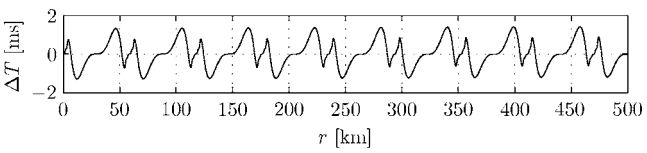


FIG. 4. In the environments shown in Fig. 3 travel time error, $\Delta T(r;H)=T_{\mathrm{PE}}(\bar{z}(r;H))-T_{\mathrm{HE}}(z(r;H))$, associated with the transformation following a ray as a function of range. The Helmholtz equation ray in the original environment and the PE ray in the transformed environment have the same ray label $H=-1/c(\hat{z})=-1/\bar{c}(\hat{\bar{z}})$.

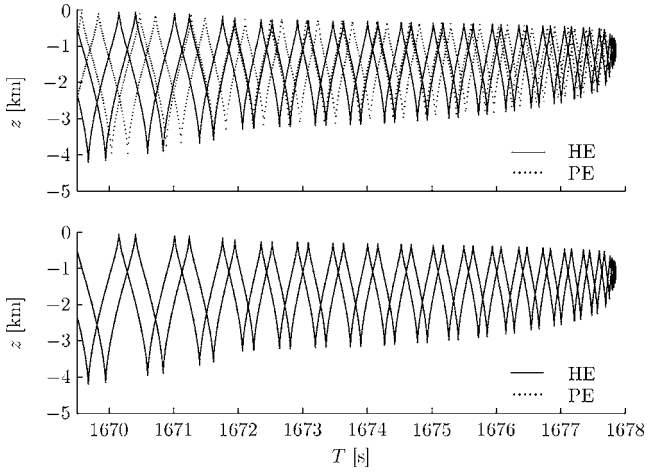


FIG. 5. In the environments shown in Fig. 3 ray travel times vs depth at a range of 2500 km for an axial point source. The upper subplot shows Helmholtz equation (solid line) and standard PE (dots) time fronts in the original environment. The lower subplot shows Helmholtz equation (solid line) time front in the original environment and standard PE (dots) time front after the use of the transformation.

$\Delta T(r;H)=T_{\mathrm{PE}}(\bar{z}(r;H))-T_{\mathrm{HE}}(z(r;H))$, as a function of range. Two rays with the same H value were propagated to 500 km in range, a Helmholtz equation ray in the original environment and the corresponding PE ray in the transformed environment. As explained in Sec. III, $\Delta T(r;H)$ has four zero crossings within each complete ray loop—one at each of two axis crossings, one at the upper turning depth, and one at the lower turning depth. Note that there is no secular (in range) growth of $\Delta T(r;H)$ and the maximum absolute value of ΔT is about 1 ms.

Figure 5 illustrates the effect of the transformation on the ray limit of a transient wave field in the above-described environment. Using a point source on the sound channel axis ray travel times at all depths are plotted at a range of 2500 km. This figure shows that in the original environment PE travel time predictions are significantly different ($O(0.1$ s) errors) from HE travel time predictions, but with use of the environmental transformation PE travel time predictions are in excellent agreement with HE travel time predictions. In Fig. 5 maximum travel time errors associated with the transformation are approximately 0.2 ms. These errors are much smaller than those shown in Fig. 4. The reason is that small errors in ray depth and travel time associated with the transformation largely cancel one another along the time front (Fig. 5).

Figures 6 and 7 illustrate use of the transformation in broadband full-wave wave-field simulations; PE simulations are compared to normal mode simulations that provide a very accurate solution to the Helmholtz equation in a range-independent environment. The parabolic equation model UMPE[15] was used for PE simulations; the normal mode model PROSIM[16] was used for normal mode simulations. Transient finite frequency wave fields at a range of 2500 km are shown in Figs. 6 and 7. In Fig. 6 a broadband pulse with center frequency 100 Hz and bandwidth 50 Hz was used; in Fig. 7 a broadband pulse with center frequency 50 Hz and bandwidth 50 Hz was used. (In both cases the stated band-

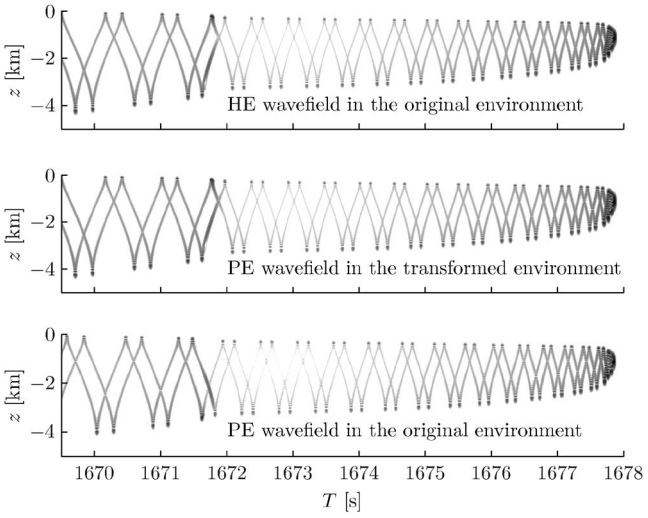


FIG. 6. In the environments shown in Fig. 3 wave-field intensity in the depth-time plane at a range of 2500 km for a broadband (f_0=100 Hz, Δf =50 Hz) axial source. A Helmholtz equation simulation in the original environment is shown in the upper subplot; a standard PE simulation after the use of the transformation of the environment is shown in the middle subplot; a standard PE simulation in the original environment is shown in the lower subplot. In all three subplots wave-field intensity is plotted on a logarithmic scale with a dynamic range of 37.5 dB.

width is the computational bandwidth; the effective bandwidth is about half the computational bandwidth. A Hanning weighting function of the spectrum was used in all the simulations shown.) With use of the transformation of the environment PE simulations are seen to be in excellent agreement with HE simulations. Note that even for a frequency as low as 50 Hz the transformation works well despite the fact that it was derived using ray-based arguments.

The reference sound speed c_0 used in the PE simulations shown in Figs. 4–7 was the axial sound speed 1.49 km/s. The same value was used in constructing the transformed environment using Eq. (24). It is important that the same value of c_0 be used when constructing the transformed environment and performing the subsequent PE simulations, but the wave fields produced are remarkably insensitive to the choice of c_0. Surprisingly, even a bad choice of c_0 (e.g., c_0 =0.33 km/s in our problem) produces good results. This behavior can be partially explained by the following observations. First, the no phase error property of the transformation holds for any c_0. Second, a choice of c_0 that is well outside the c domain in a given problem will cause a severe distortion of the depth domain after transformation, but this distortion is subsequently eliminated by enforcing the condition $c(z)=\bar{c}(\bar{z})$.

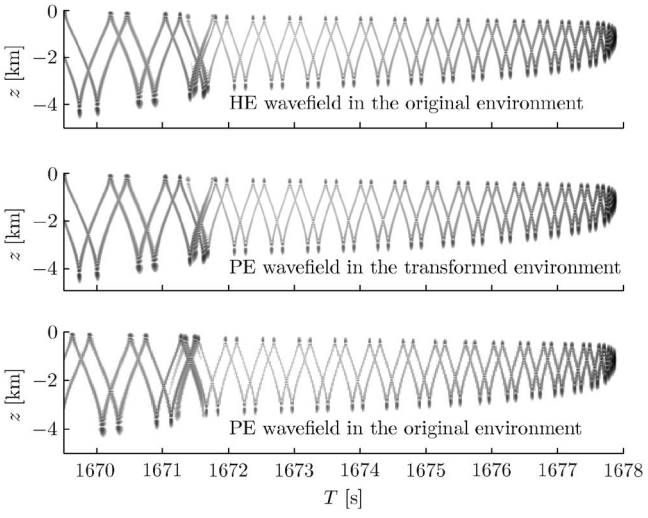


FIG. 7. Same as Fig. 6 except that f_0=50 Hz, Δf=50 Hz.

VI. DISCUSSION AND SUMMARY

In this paper we have shown that in a range-independent environment it is possible to eliminate phase errors associated with the standard parabolic wave equation by solving the PE equation in a suitably transformed environment. The phase error elimination property of the transformation is not restricted to small propagation angles or to environments with small sound speed variations. Phase errors are eliminated in the sense that asymptotic PE modal phases in the transformed environment are identical to Helmholtz equation modal phases in the original environment. The corresponding ray statement is that the transformation guarantees the equivalence of all complete ray loop travel times, thereby guaranteeing the absence of secular growth of travel time errors. In spite of these properties, the transformation described falls short of guaranteeing complete asymptotic equivalence of PE and Helmholtz equation wave fields. Subtle transformation-induced differences in both ray- and asymptotic mode-based wave-field representations have been identified and discussed.

Loosely speaking, we have shown that very accurate approximate solutions to the Helmholtz equation can be computed by solving the standard parabolic wave equation, after suitably transforming the environment. In addition to the errors associated with our transformation that we have discussed, it should be emphasized that equivalence between Helmholtz equation and parabolic equation solutions cannot be exact, even in a range-independent environment. In such an environment, the solution to the Helmholtz equation expressed as a modal expansion includes evanescent modes, but the corresponding modal expansion of the parabolic wave equation does not include evanescent modes. For this reason, equivalence of the two solutions must be limited to asymptotic equivalence, at best, which breaks down when $k_0 r$ is not large. In practice, this limitation is unimportant in almost all underwater acoustic applications. Also, in the presence of range dependence that is sufficiently strong that backscattering of energy is important, the parabolic-equation-based solution, which does not account for the backscattered energy, must deviate from the Helmholtz-equation-based solution. In range-dependent environments the transformation that we have described can be applied to each of a sequence of sound speed profiles. In such an environment, there is no guarantee that PE phase errors are eliminated. In environments with slow and/or weak range dependence, however, near complete phase error elimination should be expected. In deep ocean environments the slowly

varying in range assumption is not very restrictive because the relevant range scale is half the range of an upper or lower ray loop—typically 15 km or less.

Over the past 30 years many techniques have been developed to reduce PE phase errors. Particularly noteworthy in this regard are techniques based on the use of Padé approximations[17] to the square root operator $H(\hat{p},z,r)$. Given that other good PE phase error reduction techniques are available, it is natural to ask whether the transformation that we have described has any practical advantages over other techniques. The principal advantages of use of the environmental transform that we have described are: (1) the secular growth of phase errors is completely eliminated when $c=c(z)$, leading to accurate predictions at very long range; and (2) the transform is based on the use of the standard PE, which can be solved using the split-step Fourier algorithm. That algorithm is fast, accurate and preserves, in discrete form, the unitary structure of the equation, but has the disadvantage of not handling discontinuities in sound speed or density well. With these comments in mind, in underwater acoustic applications the transformation we have described is probably best suited for use in long-range deep ocean problems.

Another good PE phase error reduction technique is to make use of the c_0-insensitive approximation.[18] Like the work described here, that approximation involves transforming the environment and solving a parabolic equation in the transformed environment. In the c_0-insensitive approximation the simple stretching $d\bar{z}=\sqrt{c_0/c}dz$ is used in conjunction with the Thomson-Chapman[19,20] equation (TCPE). Interestingly, the arguments that we have used provide a means to derive the c_0-insensitive approximation. The TCPE Hamiltonian is

$$H_{\mathrm{TCPE}}(p,z,r)=-(c_0^{-2}-p^2)^{1/2}+\frac{1}{c_0}-\frac{1}{c(z,r)}. \tag{36}$$

As described in Sec. III, the objective is to find a transformation for which $H(I)$ is the same in HE and TCPE systems, and which satisfies $G_{\mathrm{HE}}(z,I)=G_{\mathrm{TCPE}}(\bar{z},I)$. Both objectives are achieved by requiring

$$\int_z^{\hat{z}}\sqrt{c^{-2}(z')-H^2}dz' = \int_{\bar{z}}^{\hat{\bar{z}}}\sqrt{c_o^{-2}-\left(H-\frac{1}{c_0}+\frac{1}{\bar{c}(\bar{z}')}\right)^2}d\bar{z}', \quad \forall\, z\in[0,\hat{z}],\bar{z}\in[0,\hat{\bar{z}}]. \tag{37}$$

Changing variables from z to c and $\bar{z}$ to $\bar{c}$, requiring that the integration domains coincide, and equating integrands leads to the conditions $c(z)=\bar{c}(\bar{z})$ and

$$d\bar{z}=\sqrt{\frac{\frac{2}{c}-\left(H+\frac{1}{c}\right)}{\frac{2}{c_0}-\left(H+\frac{1}{c}\right)}}dz\approx\sqrt{\frac{c_0}{c}}dz, \tag{38}$$

which is precisely the c_0-insensitive approximation. Corrections to the approximate expression $d\bar{z}=\sqrt{c_0/c}dz$ are $O(\varepsilon^2)$ where $\varepsilon=(c-c_0)/c_0$, which explains why this approximation works very well when sound speed variations are small.

There is also a similarity between the argument that we used to derive our exact environmental transformation and the argument that was used to derive the "optimal PE" approximation.[21] It is of interest to point out how the arguments underlying our work and optimal PE differ. In the derivation of optimal PE, PE ray paths $z_{\mathrm{PE}}(r)$ were constrained to coincide with ray paths of the Helmholtz equation $z_{\mathrm{HE}}(r)$ in a range-independent environment. That constraint was applied, however, without consideration of ray travel times and without forcing $z_{\mathrm{PE}}(r)$ and $z_{\mathrm{HE}}(r)$ to correspond to the same constant value of H. Because the H constraint was not imposed, the action-angle description [including both $H(I)$ and $G(z,I)$] for the two problems is generally different, and, for that reason, phase errors were not eliminated. In contrast, our constraint on $H(I)$ guarantees that phase errors are eliminated in the sense that we have described.

There may also be a connection between our work and the work described by DeSanto[22] and Thomson and Wood[23] where it is argued that PE phase errors can be eliminated as a postprocessing step, but we have not investigated this issue.

ACKNOWLEDGMENTS

We thank F. J. Beron-Vera, O. A. Godin, and A. G. Voronovich for the benefit of discussions relating to this work. This research was supported by Code 321OA of the Office of Naval Research and the National Science Foundation, Grant No. CMG-0417425.

[1] M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. A. Wolfson, and G. M. Zaslavsky, "Ray dynamics in long-range deep ocean sound propagation," J. Acoust. Soc. Am. **113**, 2533–2547 (2003).
[2] M. A. Leontovich and V. A. Fock, "Solution of the problem of electromagnetic wave propagation along the surface of the earth according to the method of parabolic equations," J. Phys. (USSR) **10**, 13–24 (1946).
[3] F. D. Tappert, "The parabolic approximation method," in *Wave Propagation and Underwater Acoustics*, Lecture Notes in Physics **70**, edited by J. B. Keller and J. S. Papadakis (Springer, Berlin, 1977).
[4] F. D. Tappert and M. G. Brown, "Asymptotic phase errors in parabolic approximations to the one-way Helmholtz equation," J. Acoust. Soc. Am. **99**, 1405–1413 (1996).
[5] R. H. Hardin and F. D. Tappert, "Applications of the split step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations," SIAM Rev. **15**, 423 (1973).
[6] L. D. Landau and E. M. Lifshitz, *Mechanics*, translated from the Russian by J. B. Sykes and J. S. Bell (Butterworth-Heinemann, Oxford, 1997).
[7] A. L. Virovlyansky, 'Ray travel times at long ranges in acoustic waveguides," J. Acoust. Soc. Am. **113**, 2523–2532 (2003).
[8] F. J. Beron-Vera and M. G. Brown, "Travel time stability in weakly range-dependent sound channels," J. Acoust. Soc. Am. **115**, 1068–1077 (2004).
[9] K. Aki and P. G. Richards, *Quantitative Seismology* (Freeman, San Francisco, 1980).
[10] O. A. Godin, "Parabolic approximation in the theory of the sound propagation in three-dimensionally heterogeneous media," Dokl. Phys. **45**, 367–371 (2000).
[11] O. A. Godin, "Reciprocity end energy conservation within the parabolic approximation," Wave Motion **29**, 175–194 (1999).
[12] M. G. Brown, J. Viechnicki, and F. D. Tappert, "On the measurement of modal group time delays in the deep ocean," J. Acoust. Soc. Am. **100**, 2093–2102 (1996).
[13] D. S. Ahluwalia and J. B. Keller, "Exact and asymptotic representations of the sound field in a stratified ocean," in *Wave Propagation and Underwa-*

ter Acoustics, Lecture Notes in Physics **70**, edited by J. B. Keller and J. S. Papadakis (Springer, Berlin, 1977).
[14]W. H. Munk, "Sound channel in an exponentially stratified ocean, with application to SOFAR, "J. Acoust. Soc. Am. **55**, 220–226 (1974).
[15]K. B. Smith and F. D. Tappert, "The University of Miami parabolic equation model," MPL Tech. Memor. No. 432, May 1993.
[16]F. Bini-Verona, P. L. Nielsen, and F. B. Jensen, "PROSIM broadband normal-mode model: A users's guide," SACLANTCEN SM-358 technical report, 1998.
[17]M. D. Collins, "A split-step Padé solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[18]F. D. Tappert, J. L. Spiesberger, and L. Boden, "New full-wave approximation for ocean acoustic travel time predictions," J. Acoust. Soc. Am. **97**, 2771–2782 (1995).
[19]D. J. Thomson and N. R. Chapman, "A wide-angle split step algorithm for the parabolic equation," J. Acoust. Soc. Am. **74**, 1848–1854 (1983).
[20]M. D. Feit and J. A. Fleck, Jr., "Light propagation in graded-index optical fibers," Appl. Opt. **17**, 3990–3998 (1978).
[21]D. H. Berman, E. B. Wright, and R. N. Baer, "An optimal PE-type wave equation," J. Acoust. Soc. Am. **86**, 228–233 (1989).
[22]J. A. DeSanto, "Relation between the solutions of the Helmholtz and parabolic equations for sound propagation," J. Acoust. Soc. Am. **62**, 295–297 (1977).
[23]D. J. Thomson and D. H. Wood, "A postprocessing method for removing phase errors in the parabolic equation," J. Acoust. Soc. Am. **82**, 224–232 (1987).

Experimental demonstration of iterative time-reversed reverberation focusing in a rough waveguide. Application to target detection[a)]

Karim G. Sabra,[b)] Philippe Roux, Hee-Chun Song,
William S. Hodgkiss, and W. A. Kuperman
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

Tuncay Akal
TUBITAK-MAN, Marmara Research Center, Earth and Marine Science Research Institute, Kocaeli, Turkey

J. Mark Stevenson
NATO Undersea Research Center, Viale San Bartolomeo, La Spezia, Italy

(Received 14 June 2005; revised 14 June 2006; accepted 20 June 2006)

For most shallow water waveguides, the backscattered energy measured in a monostatic configuration is dominated by ocean bottom reverberation. A selected time-gated portion of the measured reverberation signal is used to provide a transfer function between a time-reversal array and a corresponding range interval on the bottom. Ultrasonic and at-sea experiments demonstrate the focusing capabilities of a time-reversal array along the rough bottom interface using these reverberation signals only. The iterative time-reversal technique facilitates robust focusing along the ocean bottom, with little signal processing effort involved and *a priori* information of the environment. This allows for enhanced detection and localization of proud or buried targets in complex shallow water environments. A passive implementation of the iterative time-reversal processing is used to construct reflectivity maps, similar to a sonar map, but with an enhanced contrast for the strongest reflectors (or scatterers), at the water-bottom interface. Ultrasonic and at-sea experiments show that targets on the seafloor located up to 400 wavelengths from a time-reversal array are detectable in the presence of bottom reverberation.
 [DOI: 10.1121/1.2227377]

PACS number(s): 43.30.Gv, 43.30Hw, 43.60.Tj [DRD] Pages: 1305–1314

I. INTRODUCTION

Detection of objects on the seafloor is an application of active sonar that continues to witness significant commercial and military investment. This article investigates the potential use of time reversal with a transducer array for the detection problem in seafloor mapping. This article builds upon research in time-reversal arrays (TRA) described in the extensive overview of TRA applications (such as medical imaging, nondestructive testing, or underwater acoustics) published recently by Fink *et al.*[1]

In underwater acoustics, the robust focusing and pulse compression provided by time-reversal techniques may be exploited for active sonar.[2,3] Backscattering from the rough water-bottom interface can be used as a surrogate probe source.[4] For most shallow water waveguides, the backscattered energy in a monostatic configuration is dominated by ocean bottom reverberation. By selecting a time-gated portion of the reverberation signals (or backscattered signals) the transfer function between the TRA and a corresponding range interval on the bottom can be estimated. In this paper, this method is used to focus sound in the vicinity of the water-bottom interface and to improve the detection of targets lying on the seafloor or buried in the sediments. Since the goal here is to focus on the strongest scatterer located in the selected range interval, the iterative time-reversal process applied to a given time-gated window of backscattered signals provides a straightforward solution using only time-domain signals.[5–8] Hence, when the reflectivity contrast between the buried target and the bottom reverberation is sufficient, focusing energy on the target is accomplished with very little signal processing effort. Other related techniques have been developed in underwater acoustics for reverberation nulling[9] or applications of the DORT method to discriminate between scatterers.[10] But these techniques based on singular value decomposition require processing each frequency independently and thus do not yield easily broadband focusing signals for the TRA, which are necessary for optimal spatio-temporal focusing.

Repeating the iterative time-reversal process over a sliding time-gated window portion of the initial reverberation signals recorded in the oceanic waveguide enables the system to scan the bottom surface and search for local strong scatterers. The main result of this paper is a reflectivity map of the water-bottom interface, similar to a sonar map, but with an enhanced contrast for the strongest reflectors (or scatterers), which appear as bright spots on the reflectivity

[a)]Part of this work was presented at the 150th meeting of the Acoustical Society of America in Minneapolis, MN.
[b)]Electronic mail: ksabra@mpl.ucsd.edu

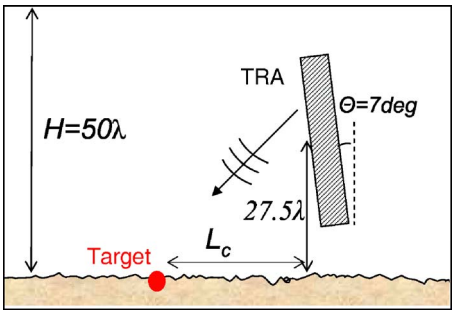


FIG. 1. (Color online) Ultrasonic waveguide experimental setup. The source-receive array is located on the left-hand side and its support can be oriented arbitrarily. The probe transducer was set at different ranges and depths above the bottom interface. One spherical target lays at a distance L_c from the base of the array.

map. Both ultrasonic tank data and at-sea data are used to demonstrate experimentally the feasibility of this technique.

In Sec. II, the experimental setups used for the ultrasonic tank experiments and the at-sea trials are presented. Section III presents the implementation of the iterative time-reversal processing when using reverberation focusing. Experimental results demonstrate that backscattered signals can be used to focus sound along the seafloor. In Sec. IV, the reverberation focusing technique is used to detect targets on the ocean bottom. Section V presents the conclusions of this analysis.

II. EXPERIMENTAL SETUP

A. Ultrasonic water tank experiment

The first set of ultrasonic experiments were conducted in a water tank (see Fig. 1) using a standard linear array of $M=32$ elements equally spaced by 1 mm with a central frequency $f_c=1.49$ MHz and a 75% frequency bandwidth. The waveguide has a uniform sound speed of $c_0 \approx 1486$ m/s, a depth $H=5$ cm, ($H \approx 50\lambda$) with a moving free surface. This setup is a scaled-down approximation of a typical oceanic if experiment by a factor of 1000. The bottom was covered with fine sand grains of diameter ranging from 10 to 100 μm, scaling to an oceanic bottom consisting of gravels i.e., highly reverberant. The targets were identical steel spheres (ball bearings) of 5-mm diameter (see Sec. IV).

The array makes an angle of 7° with respect to the vertical, and the middle element was at 27.5λ above the water-bottom interface to mimic practical tilted array geometry for transducers mounted at the bow of a ship along the hull. Each individual element has an aspect ratio of 12 mm in width by 0.5 mm in height. The narrow horizontal directivity of each transducer prevents reflections from the sidewalls of the waveguide. Additionally this 2D directionality is also representative of a monostatic sonar system with fine azimuthal resolution (see Sec. II B). The probing signal broadcast during the reverberation measurements was a 5-μs pulse that was obtained after time compression of a Linear Frequency Modulated (LFM) chirp of 40 μs spanning the whole frequency bandwidth of the transducers.

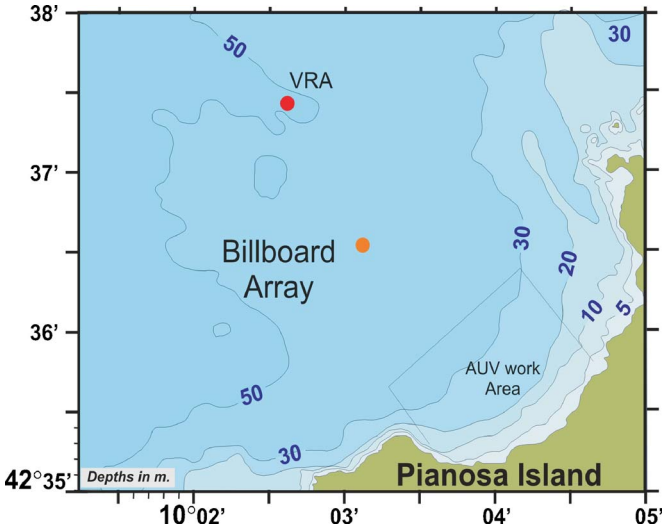


FIG. 2. (Color online) Bathymetry North West of Pianosa Island, off the West coast of Italy. Location of the billboard array is indicated.

B. Sea experiments

The at-sea experiment was conducted North West of Pianosa Island, off the West coast of Italy (see Fig. 2). The $M=28$ elements of the TRA, with a center frequency of 3.5 kHz and a bandwidth of 1 kHz, were assembled into a billboard configuration (roughly 8 by 1.5 m, see Fig. 3) that is a surrogate for existing sonar systems in a monostatic configuration. The azimuthal angular beam width of the billboard array was 16°. This improves target detection and localization along the seafloor (see Sec. IV) by sending directional acoustic power on a particular patch of the seafloor. Hence the signal-to-noise ratio of the backscattered reverberation measurements is enhanced. The ultrasonic experiments were designed to have a similar setup and seafloor ensonification as the at-sea experiments.

Technical features of the billboard array transducers and the electronic system have been described previously.[11] Figure 4 represents several sound speed profiles collected in the Pianosa area over the period JD204–JD210. These profiles were strongly downward refracting with a thermocline transition around 20 m. The billboard array was suspended from the *R/V Alliance* below the thermocline with the first horizontal subarray of the array approximately located at a depth of 24.5 m. The bathymetry was relatively range independent (around 50 m) in the vicinity of the billboard array location (see Fig. 2). The top layer of the ocean bottom principally was made of sand and fine gravel.

III. TIME-REVERSED REVERBERATION FOCUSING IN A ROUGH WAVEGUIDE

A. Data processing

Following Prada *et al.*,[6] the received backscattered signals can be expressed as a function of the transmitted signals using the interelement impulse response matrix $\mathbf{K}$ of the M elements of the TRA. Each element $k_{l,m}(t)$ of the matrix $\mathbf{K}$ is the output of the element number l ($1 \leq l \leq M$), when the input element number m ($1 \leq l \leq M$) is a temporal Dirac function. The knowledge of $\mathbf{K}$ allows for the description of

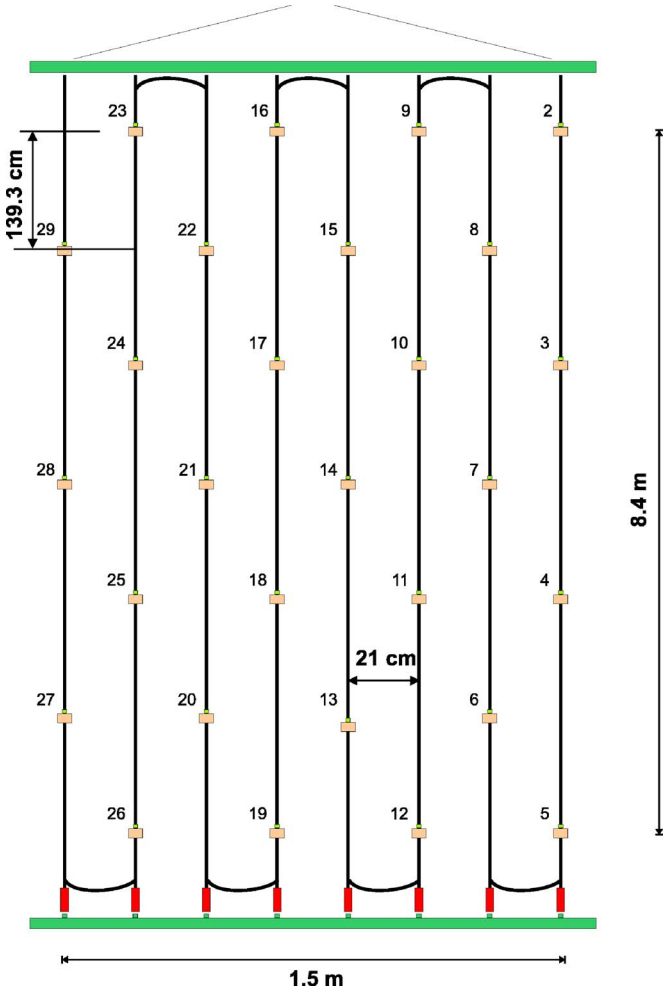


FIG. 3. (Color online) Billboard array geometry and element spacing are indicated.

any transmit-receive sequences of the time-reversal process using reverberation focusing in a rough waveguide described in Fig. 5. In Fig. 5(a), a TRA composed of M elements is used to initially ensonify the ocean floor (transmit mode), e.g., transmitting a broadside illumination to maximize the reverberation return level over the ambient noise level. The resulting reverberation time series recorded on each element m $(1 \leq m \leq M)$ [receive mode, see Fig. 5(b)] are

$$R_m^{(0)}(t) = \sum_{l=1}^{M} k_{l,m}(t). \quad (1)$$

A selected time-gated portion of the reverberation signals (referred to as *reverberation window*) $R_m^{(0)}(T_c - \delta t/2 \leq t \leq T_c - \delta t/2)$ of length δt around a two-way travel time T_c is used to provide a transfer function between the TRA and a corresponding range interval on the bottom.[4] In an oceanic waveguide with small bathymetry changes, this selected reverberation window corresponds to a spatial cell on the seafloor of resolution $\delta r = \delta t c_0/2$ at a range $L_c = T_c c_0/2$, where c_0 is the average sound speed in the water column. At longer ranges due to dispersion effects and large bathymetry changes, additional measurements and numerical modeling of the acoustic paths are necessary to estimate a more accurate relationship between δr and δt.[12]

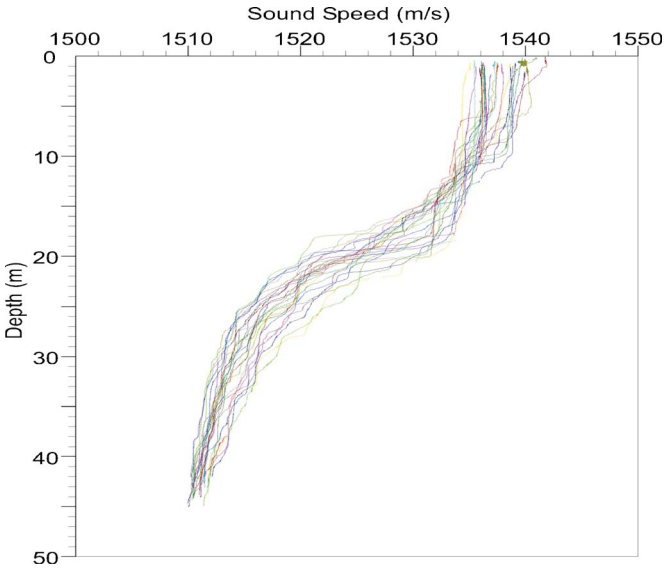


FIG. 4. (Color online) CTD casts collected in the vicinity of the billboard array location on JD204-JD210. Despite some temporal variability of the sound speed profile, these profiles generally were strongly downward refracting with a thermocline transition around 20 m.

The signals in the selected reverberation window are time-reversed [Fig. 5(c)] and broadcast back (transmit mode) Fig. 5(d) to focus along the sea floor at the corresponding range L_c. Finally, enhanced backscattered originating from the focusing region is recorded on the TRA [Fig. 5(e)]. Each time-reversal procedure involves two successive back and forth propagations between the spatial cell on the ocean floor and the TRA. The measured reverberation time series on channel m after completion of the time-reversal process is

$$R_m^{(1)}(t) = \sum_{l=1}^{M} k_{l,m}(t) \otimes R_l^{(0)}(T_c - \delta t/2 \leq t \leq T_c + \delta t/2), \quad (2)$$

where the symbol $\otimes$ indicates a correlation operation. Note that singular value decomposition formalism previously developed in the frequency domain[4] is not necessary since the main source of backscattered energy is only the bottom reverberation. In this article, a simple and direct time-domain implementation of the time-reversal process is used. The time-domain formulation has the advantage, in particular for broadband signals, of fully taking into account the coherence of the reverberation signals across the entire frequency bandwidth and not just at the center frequency. Furthermore, using an M-element TRA instead of a single transmit-receive element[13,14] has the advantage of providing an additional coherent array gain of $10 \log_{10}(M)$ at the focusing location on the ocean floor. In general, this coherent array gain along with the selection of the reverberation window is crucial in improving the detection of a target (i.e., a point scatterer) over the reverberation signal (i.e., an extended target). Indeed if the target strength is weak, the echoes from the target will likely be dominated by the contribution of whole rough bottom reverberation in the backscattered signals recorded after the initial broadside illumination. By selecting a short reverberation window, the influence of ocean reverberation is limited to the close vicinity of

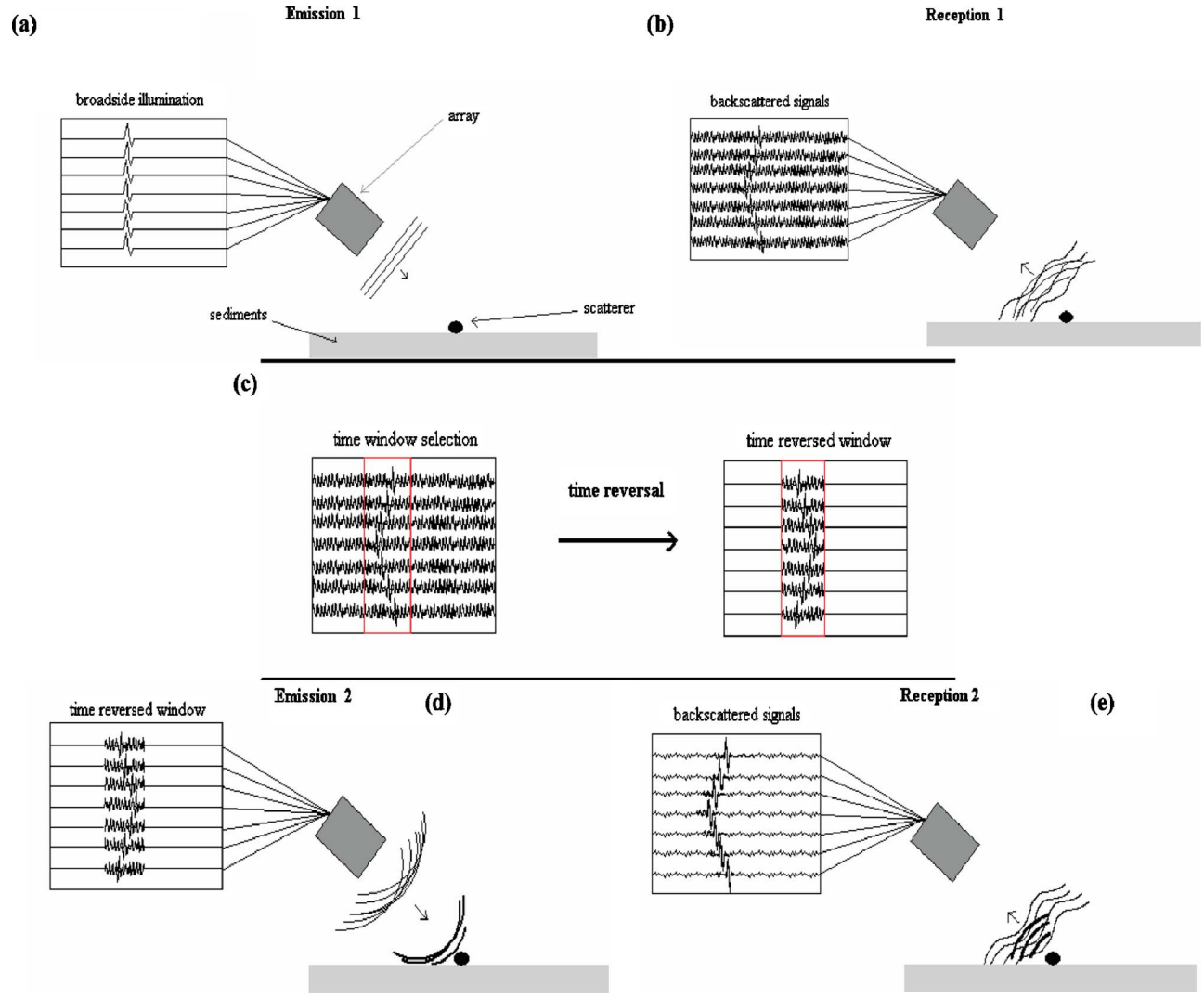


FIG. 5. (Color online) Schematic of the implementation of the reverberation focusing technique using a TRA in a monostatic configuration.

the target. Then if the reflectivity contrast between the target and the bottom reverberation is sufficiently high locally (e.g., a steel target lying on a sandy bottom), the backscatter locally will be higher when focusing at the target range L_c.

Iterating the time-reversal process provides a simple means to enhance focusing on the strongest reflector[6–8] and a solution for separating the echoes of the targets from the reverberation of the ocean bottom. Using Eq. (2) recursively, the reverberation time series recorded on the TRA at the end of ith iterations of the iterative time-reversal process (ITR) (i.e., after two set of back and forth propagations) are $R_m^{(2i+1)}(t)$ (i.e., the odd power of the recursive index i) ($i \geq 0$).[6] However, a direct implementation of the-iterative time-reversal process for multiple reverberation windows, i.e., selecting several different values of T_c, can be very lengthy at sea since it requires recording long backscattered signals each time. Array motion during the acquisition time might further degrade the results. Thus a completely active implementation of all steps of the ITR is likely to be impractical for mapping the ocean bottom.

One practical solution during sea trials was to first acquire the complete interelement impulse response matrix **K** by successively firing each element of the TRA and recording the backscattered signals on the whole TRA after a round-trip propagation between the TRA and the ocean bottom.[6] The time-reversal iterations can then be implemented passively using Eq. (2) recursively by simply computing correlations of the $k_{l,m}(t)$ signals with the selected time-gated window of the reverberated signals obtained from the previous iterations. The acquisition of the transfer matrix **K** is the only active part of the processing. In the case of low signal-to-noise ratio, it is necessary to acquire a new realization of the transfer matrix **K** for each iteration of this passive implementation of the ITR. Otherwise the ambient noise recorded along with the backscattered signals in the matrix **K** would always remain the same and would then be autocorrelated at each iteration of Eq. (2), which may bias the ITR process.

B. Ultrasonic water tank experiments

This section investigates experimentally the reverberation focusing technique presented in Sec. III A, using the ultrasonic water tank setup described in Sec. II A. The iterative time-reversal process is implemented passively based on

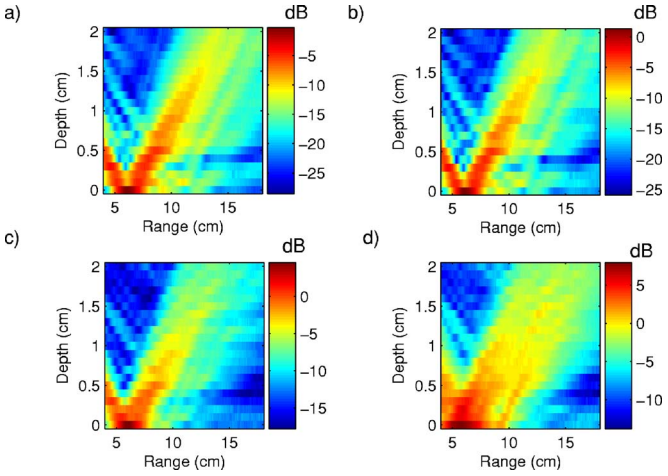


FIG. 6. (Color online) Spatial resolution of reverberation focusing. The intended focusing range $L_c \approx 6$ cm and the duration of the Hanning reverberation window is (a) $\delta t = 5$ μs, (b) $\delta t = 10$ μs, (c) $\delta t = 40$ μs, and (d) $\delta t = 120$ μs. The four plots have the same dB scale.

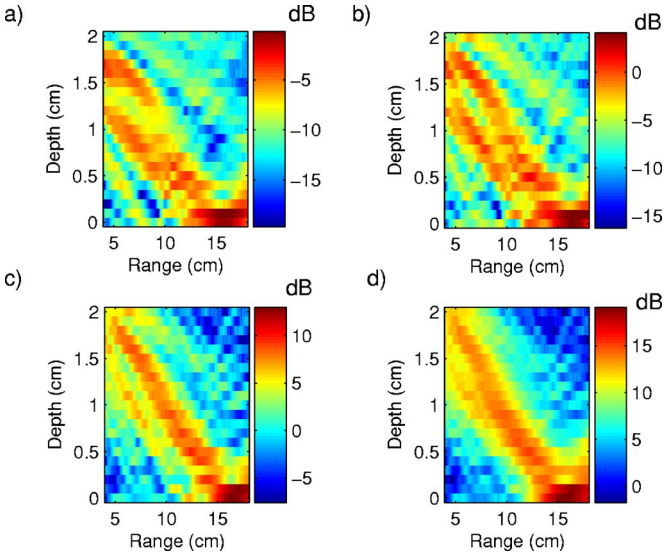


FIG. 7. (Color online) Same as Fig. 6 but for $L_c \approx 16$ cm.

the measurement of the interelement impulse response matrix $\mathbf{K}$ for the $M=32$ elements of the ultrasonic TRA [see Eqs. (1) and (2)]. In order to investigate the spatial resolution of the focal spot along the water-bottom interface, the time-domain Green's function $G_l(R_f, Z_f; t)$ was also measured between each of the lth elements of the TRA and an omnidirectional probe transducer scanning the water column at a range R_f and a depth Z_f in the vicinity of the intended focal spot along the water-bottom interface (see Fig. 1). Based on the formalism of Sec. III A [see Eqs. (1) and (2)], the resulting retrofocused field $B^{(2i+1)}(R_f, Z_f; t)$ at the hydrophone location after the ith iteration of the time-reversal process is

$$B^{(2i+1)}(R_f, Z_f; t) = \sum_{l=1}^{M} G_l(R_f, Z_f; t) \otimes R_l^{(2i)}(T_c - \delta t/2 \le t \le T_c + \delta t/2), \tag{3}$$

where T_c denotes the center of the reverberation time window used for the time-reversal process and δt is its duration.

Figure 6 illustrates the spatial resolution of the focal spot along the water-bottom interface for Hanning reverberation windows of increasing duration $\delta t = 5$, 10, 40, and 120 μs. The normalized retrofocused energy, defined as the integration of $[B^{(2i+1)}(R_f, Z_f; t)]^2$ over the recording time window [see Eq. (3)], is displayed in depth versus range. A strong focus of the backpropagated energy is observed near the bottom at a range $L_c \approx 6$ cm, corresponding to a two-way travel time $T_c = 72$ μs (see Sec. III A) from the TRA with a 6-dB width of the focus along the bottom interface on the order of 2.5 cm for $\delta t = 10$ μs [see Fig. 6(b)]. The retrofocusing range $L_c \approx 6$ cm is relatively close to the TRA (around 1.2 times the water depth). Increasing the length of the reverberation window (up to $\delta t < 20$ μs in this configuration) allows us to better capture the transfer function between the TRA and a corresponding spatial cell on the seafloor, thus improving the focusing. However, if the value of δt gets too large (i.e., $\delta t > 40$ μs here), the focal spot gets broader and larger sidelobes appear [see Figs. 6(c) and 6(d)]. Indeed, for long reverberation windows, the distribution of scattering sources along the interface becomes too wide and thus the resulting backscattered signals differ from the case of a point scatterer on the bottom. Hence the correspondence between the selected time-gated reverberation window and a specific spatial cell on the seafloor at a range $L_c \approx T_c c_0/2$ is no longer unique.

Figure 7 displays the distribution of normalized retrofocused energy, similarly to Fig. 6, when retrofocusing at a longer range $L_c \approx 16$ cm (i.e., $T_c = 210$ μs). The spatial resolution of the focal spot is worse (higher sidelobes) when compared to Fig. 6 with a 6-dB width of the focal spot along the bottom interface now in the order of 4 cm. Since the waveguide dispersion effects are more pronounced at larger range L_c, the backscattered signals originating from the same spatial cell of length δr along the seafloor are further spread in time at the TRA location and might not be all correctly captured within the duration δt of the reverberation window. Based on the results of Figs. 6(a) and 7(a), the parameters $\delta = 5$ μs and 6 cm $\le L_c \le 17$ cm yield good spatial focusing for this TRA geometry and will be used for the target detection experiment (see Sec. IV).

Figure 8 displays the backscattered energy $E_{\mathrm{av}}(t)$ recorded by each of the $M=32$ elements of the TRA after either the initial broadside transmission [Fig. 8(a)] or the time-reversal operation [Figs. 8(b) and 8(c)]. The backscattered energy $E_{\mathrm{av}}(t)$ is obtained by incoherently averaging the envelope of the signals $R_l^{(i)}(t)$ [see Eqs. (1) and (2)] where

$$E_{\mathrm{av}}^{(i)}(t) = \sum_{l=1}^{M} (|\mathrm{Hilbert}(R_l^{(i)}(t))|)^2. \tag{4}$$

A coherent average of the backscattered signals may also be used instead, but it was found to be less robust, in practice, to environment fluctuations and array motion during the at-sea trials. The backscattered energy $E_{\mathrm{av}}^{(i)}(t)$ (measured after the ith time-reversal iteration) was normalized by the corresponding ambient noise level measured for $t > 400$ μs. After the initial broadside transmission [Fig. 8(a)], the back-

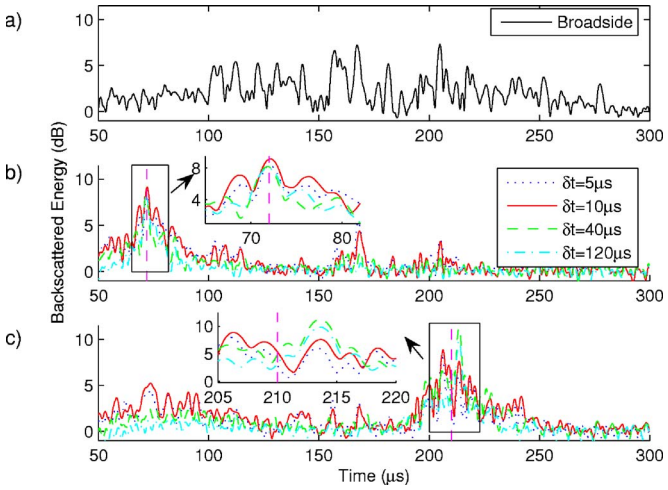


FIG. 8. (Color online) Backscattered reverberation time-series at the TRA location. (a) After initial broadside illumination. (b) After focusing at a range $L_c \approx 6$ cm, i.e., $T_c = 72$ μs. Results for four reverberation windows are indicated (corresponding to the values used in Fig. 7). (c) Same as (b) but for $L_c \approx 16$ cm, i.e., $T_c = 210$ μs.

scattered energy comes from, to a large extent, the rough bottom of the waveguide.

Figures 8(b) and 8(c) show the backscattered energy $E_{\mathrm{av}}(t)$ recorded after focalization for values of the parameters $L_c = 6$ cm and $L_c = 16$ cm, respectively, and for values of δt similar to Figs. 6 and 7. Most of the backscattered energy is now coming from bottom reverberation in the vicinity of L_c since the maxima remains centered around the focusing time T_c. The strong return at $T_c = 75$ μs in Fig. 8(b) [or $T_c = 215\mu$ in Fig. 8(c)] illustrates the coherent gain across the array obtained by using the time-reversal process. Increasing duration δt of the reverberation window can yield higher values of the backscattered retrofocused energy $E_{\mathrm{av}}(t)$ [see Fig. 8(c)]. In this case the measurements of $E_{\mathrm{av}}(t)$ at the TRA location indicate that higher energy is sent along the seafloor since a longer reverberation window is used. But the results on Figs. 6 and 7 show that a high return on the TRA does not necessarily imply a good focusing along the seafloor when the selected reverberation windows get too long. Thus for practical application of the time-reversed reverberation focusing technique, the choice of δt will depend on the acoustic propagation characteristics of the waveguide (e.g., dispersion effects) and TRA orientation. Furthermore, the parameter δt could be estimated from simple broadband numerical modeling of the oceanic waveguide between the TRA and a range L_c.[12]

C. At-sea experiments

The time-reversed reverberation focusing technique was investigated at sea in an area northwest of Pianosa Island (see Sec. II B), using a TRA deployed in a billboard configuration (see Fig. 3). The azimuthal rotation of the suspended billboard array around its main axis was monitored with a compass and did not exceed one or two degrees in azimuth for quiet sea conditions over a time period on the order of 20 min. Thus the billboard array illuminates a constant area on the seafloor (at least for short ranges) during this time period. The time-reversed reverberation focusing technique was implemented *actively*, meaning that the time-reversed reverberation window (centered on T_c and having a duration δt, see Sec. III A) was effectively broadcast back by the TRA in the ocean. In order to enhance the signal-to-noise ratio for the measured reverberation time series over the ambient noise level, a LFM signal of duration $T = 100$ ms or $T = 1$ s (for long range focusing) over a bandwidth $B_\omega = 3-4$ kHz is used. The LFM signal was broadcast by all TRA elements at full power (approximately 180 dB) to provide an initial broadside illumination of the seafloor.

The time-delay between the initial broadside transmission and the focusing transmission was on the order of 7 min due to broadcasting and processing constraints. The data processing consisted in time compressing by cross correlating the reverberation time series with the initial LFM signal, selecting the desired reverberation window (see Sec. III A) and then convolving the reverberation window with the LFM signal before the following transmission. The time compression by the chirp provided a gain proportional to $\log(TB_\omega)$ for the measured reverberation signals.

The billboard array consists of seven horizontal subarrays of four elements each (see Fig. 3). The recorded backscattered signals after the ith iterations are noted $R_{l,m}^{(i)}(t)$, where l is the stave index $1 \le l \le 7$, and m is the element index on each stave $1 \le m \le 4$. In order to benefit from the azimuthal resolution of the billboard array, the backscattered signals are first summed coherently for the four elements of each horizontal subarray (i.e., broadside beamforming in the horizontal plane). The averaged backscattered energy is then computed by summing the envelope of the broadside beamformer output signal obtained for each of the seven staves:

$$E_{\mathrm{av}}^{(i)}(t) = \sum_{l=1}^{7} \left(\left| \mathrm{Hilbert}\left(\sum_{m=1}^{4} R_{l,m}^{(i)}(t) \right) \right| \right)^2. \tag{5}$$

Figure 9(a) displays the backscattered energy $E_{\mathrm{av}}^{(0)}(t)$ ($i = 0$) recorded after the initial broadside transmission. The large increased reverberation return around 3.1 s is related to reflection from the rapid bathymetry change located along the curved part of the coast of Pianosa Island (see Fig. 2) around (42°N, 10°04E). Other strong reverberation returns probably are related to strong bathymetry changes or the tip of the island [e.g., around (42°N, 35.5, 10°03.5E)]. Ambient noise level can be determined from the plateau for recording time $t > 3.5$ s, i.e., around 125 dB (after time compression with the chirp signal).

The average backscattered energy $E_{\mathrm{av}}^{(1)}(t)$ ($i = 1$) recorded after focusing along the ocean bottom using an a time-gated window of $\delta t = 60$ ms of the initial broadside reverberation centered around three increasing recording times T_c is shown in Figs. 9(b)–9(d). As expected, the retrofocused backscattered energy has a maximum at the expected round-trip travel time T_c, between the billboard array and the portion of the ocean floor of interest, which results from coherent gain across the array after time reversal. This demonstrates the focusing capabilities of the billboard array along the ocean

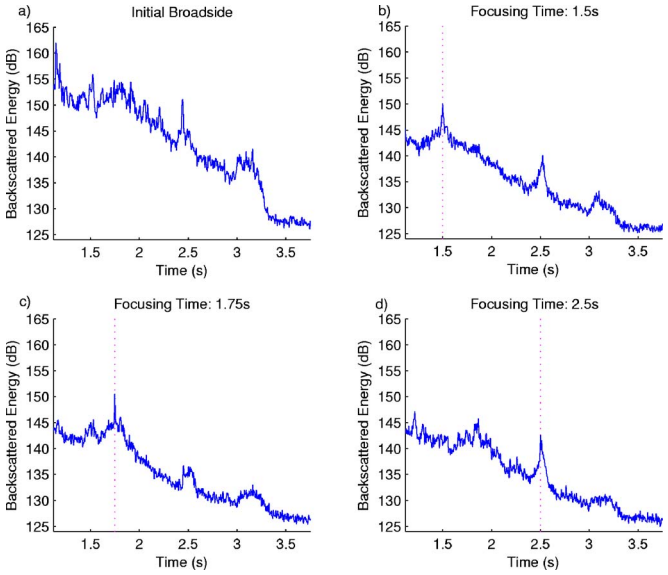


FIG. 9. (Color online) Active reverberation focusing recorded on JD208. (a) Average recorded backscattered energy from broadside transmission of a 3–4 kHz LFM of signal of duration 1 s. (b) Average recorded backscattered energy after reverberation focusing on the ocean bottom using a time-gated window of the initial broadside reverberation centered on $T_c=1.5$ s. (c) Same as (b) but $T_c=1.75$ s and (d) same as (b) but $T_c=2.5$ s. The duration of the time-window reverberation was $\delta t=60$ ms for all cases. Note that the retrofocused backscattered energy is maximized at the expected roundtrip travel time T_c (indicated by a dashed-line), between the billboard array and the portion of the ocean floor of interest. The average backscattered energy $E_{av}^{(1)(t)}$ $(i=1)$ is computed from Eq. (5). The delay between broadside transmission and the focusing transmission was on the order of 7 min.

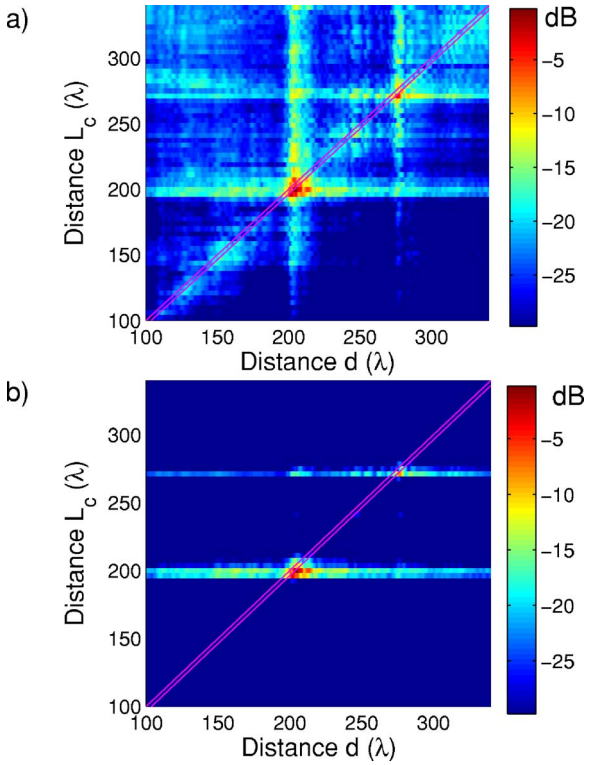


FIG. 10. (Color online) Reflectivity maps for two targets located at $L_c=210\lambda$ and $L_c=275\lambda$ away from the TRA. These maps were computed over sliding time-gated Hanning window portion of $\delta_t=5$ μs of the reverberation signals (delimited by the dashed lines along the main diagonal) for (a) the time-reversal process or (b) one iteration, of the time-reversal process. The horizontal axis is the recording time (i.e., the round trip propagation) converted to the distance. $d=tc_0/2$ ($c_0=1486$ m/s) between the TRA to the bottom. The vertical axis is the average intended focusing distance along the bottom being scanned by the TRA corresponding to the selected time-gated window. Each plot has its own dB scale. The targets appear as bright spots at $L_c=210\lambda$ and $L_c=275\lambda$ along the main diagonal. Note the contrast improvement achieved for the targets after using one iteration of the time-reversal process.

floor are robust even with a delay of 7 min between broadside transmission and the focusing transmission.

IV. APPLICATION TO TARGET DETECTION

A. Constructing a Reflectivity map of the rough bottom: data processing

By stacking the averaged backscattered energy measured at the TRA location $E_{\text{av}}^{(i)}(t)$ [see Eq. (4) or (5)] for increasing focusing time T_c (e.g., see Fig. 8 or 9), a *reflectivity map* of the water-bottom interface can be constructed, similar to a sonar map, but with an enhanced contrast for the strongest reflectors (or scatterers), which appear as bright spots on this reflectivity map (see Fig. 10). In simple TRA geometry and environment, the recording time t corresponding to the round trip propagation from the TRA to the water-bottom is converted to an equivalent distance $d=tc_0/2$, where c_0 is the averaged speed of sound in the waveguide. Similarly, the center of the reverberation window T_c is related to the focusing range $L_c=T_cc_0/2$ and one then can simply locate the strong reflectors (or scatterers).

B. Reflectivity maps in the ultrasonic rough waveguide

The same ultrasonic waveguide configuration described in Secs. II A and III B is used in this section. Using $c_0=1486$ m/s and $f_c=1.5$ MHz (center frequency), the average wavelength λ was on the order of $\lambda\approx 1$ nm. Figure 10 shows the reflectivity map obtained when two 5-mm steel spheres were placed two-thirds buried at the water-bottom interface at respectively $L_c=210\lambda$ and $L_c=275\lambda$ away from the TRA (where the distance was measured from the center located 27.5λ of the array to the target). Based on the results of Fig. 6(a), the parameter $\delta=5$ μs was selected as the length of the time-reversed Hanning window of backscattered signals since it yields a good spatial resolution for this TRA geometry. The reflectivity maps represent the variations of the averaged backscattered energy by the TRA, $E_{\text{av}}^{(2i+1)}(d,L_c)$ i.e., the local reflectivity of the bottom after the ith iterations of the time-reversal process $(i>0)$, in unit of wavelengths. To compensate for cylindrical range spreading, the averaged backscattered energy $E_{\text{av}}^{(2i+1)}(d,L_c)$ was multiplied by a factor $(1/L_c)^{2i+1}$, which approximates the total geometric energy decay after successive back and forth propagation between the TRA and the seafloor at an average range L_c. The values were also normalized by their maxima and each reflectivity map has its own color scheme and the reflectivity contrast ranges from 0 to 30 dB. The vertical axis represents the distribution of the various average focusing ranges L_c (from 100λ to 320λ). The horizontal axis is the equivalent distance $d=tc_0/2$ from the TRA to the waveguide bottom.

Both diagonals and off-diagonal terms of the reflectivity

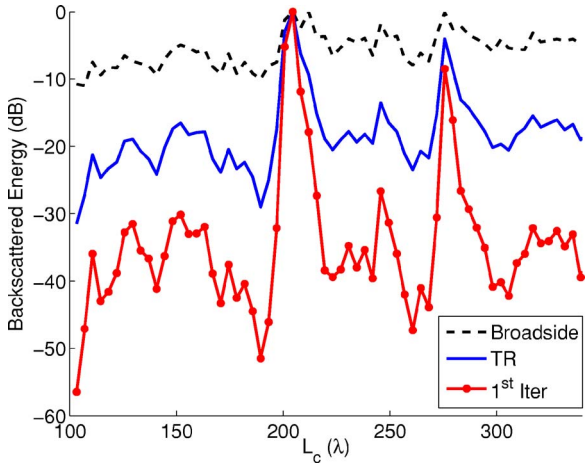


FIG. 11. (Color online) Comparison of the diagonal of the reflectivity maps (see Fig. 10) to a simple broadside transmission for target detection. For instance, in the vicinity of the first target located around $L_c=210\lambda$, the reflectivity contrast is on the order of 6 dB for the broadside transmission, 20 dB after time reversal, and above 38 dB after one iteration of the time-reversal process.

maps can help for detection, localization, and classification of the target. For each reflectivity map, the strong scatterers appear as bright spots in the main diagonal delimited by the time-gated window interval $(L_c-\delta tc_0/2\leq d\leq L_c+\delta tc_0/2)$ (indicated by dashed lines) in Fig. 10. The brightness of each spot depends on the target strength. The off-diagonal terms of the reflectivity map indicate coupling paths between multipaths of one scattered or different scatterers when focusing along the seafloor. In the case of more complex targets, the off-diagonal terms may also indicate coupling between the different structural echoes and thus may be used to improve classification of the target. The initial illumination of the water bottom interface was provided by a broadside transmission from the TRA. For this particular TRA geometry, the reflectivity maps were not varying significantly if a different illumination was used (e.g., fewer elements or different waveforms) as long as sufficiently uniform illumination over a large section of the bottom interface was achieved (not shown here).

After one iteration of the passive time-reversal process, the reflectivity contrast between the bright spot amplitudes and the surrounding bottom reverberation amplitude increases as seen by comparing Fig. 10(a) (standard time reversal) to Fig. 10(b) (first iteration). Additionally, Fig. 11 demonstrates the efficiency of the iterative time-reversal process by comparing the diagonal of the reflectivity maps to a simple broadside transmission for target detection. For instance, in the vicinity of the first target located around $L_c=210\lambda$, the reflectivity contrast is on the order of 6 dB for the broadside transmission, 20 dB after time reversal, and above 38 dB after one iteration of the time-reversal process. However, when the number of iterations gets too large, the performance does not improve significantly (not shown here) since after each iteration the signals are filtered by the limited bandwidth of the transducers, thus reducing the available frequency bandwidth and degrading the spatial resolution of focusing.[5,7] Thus one iteration of the time-reversal process appears sufficient for this particular TRA geometry and experimental setup.

C. Reflectivity maps in shallow water

The same experimental setup described in Secs. II A and III B is used in this section. The passive iterative time-reversal technique was applied to detect a target composed of 12 small glass spheres of 50-cm diameter deployed on the seafloor. The target strength of each sphere is weak in the frequency range of the billboard array illumination (3–4 kHz) and thus detection with a conventional sonar system would be problematic in this shallow water environment. The billboard array azimuth was 340°N. Based on GPS measurements the distance between the billboard array position near the bow of the *R/V Alliance* and the target location was estimated to lay at a bearing angle of 351°N and at a distance of 204.5 m with respect to the center of the billboard array. This distance corresponds to a round-trip travel time $T_c=2L_c/c_0=0.27$ s, using an average sound speed below the thermocline of $c_0=1515$ m/s (see Fig. 4).

The measurement of the interelement response matrix was done by firing each of the seven horizontal subarrays (i.e., four elements used in broadside transmission) successively using a 100-ms LFM pulse. Data processing was similar to the procedure described in Sec. III C. Four independent realizations of the interelement response matrices were collected in total over a time period of 3 min and were used to construct the backscattered signals for the iterative time-reversal process [see Eqs. (1) and (2)]. This ensures that the ambient noise recorded along with the backscattered signals in the matrix $\mathbf{K}$ is different for each acquisition and thus yields a negligible output in the computed backscattered signals after successive cross correlation [see Eq. (2)].

Figure 12 represents the reflectivity maps computed first after passive time-reversal alone and after one iteration using a sliding reverberation window of $\delta t=20$ ms. A strong reflector appears as a bright spot in the main diagonal at a range $L_c=203$ m, which is in very good agreement with the estimated range from GPS measurements (i.e., 204.5 m). Thus this reverberation focusing technique appears robust with respect to environmental fluctuations. Detection of the target is enhanced by using the iterative time-reversal process over the conventional broadside return as illustrated in Fig. 13, which compares the averaged backscattered energy after the initial broadside transmission to the diagonal of the two reflectivity maps in Fig. 12. Indeed the target is not visible in the broadside return but the reflectivity contrast, in the vicinity of the estimated target location $L_c=203$ m, is on the order of 5 dB after time reversal and above 9 dB after one iteration of the time-reversal process.

Using the azimuthal resolution of the billboard array provides additional confirmation of the target detection and localization. The interelement response matrix can be measured by steering the array axis 46° away from its axis using simple time-delay beamforming during the initial illumination each of the seven horizontal subarrays (i.e., each of the four elements now broadcasts a beam-steered LFM pulse ac-

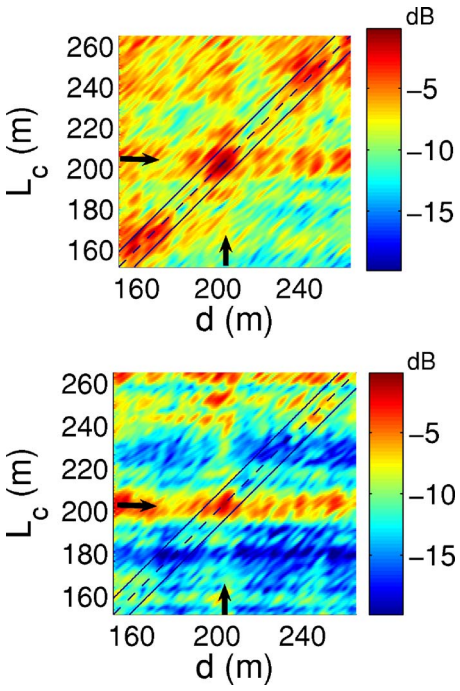


FIG. 12. (Color online) Reflectivity maps for a target, located 203 m away from the billboard array. These maps were computed over the sliding time-gated Hanning window portion of δ_t=20 ms of the reverberation signals (delimited by the dashed lines along the main diagonal) for (a) the time-reversal process or (b) one iteration of the time-reversal process. The horizontal axis is the recording time (i.e., the round trip propagation) converted to the distance. $d=tc_0/2$ (c_0=1515 m/s) between the TRA and the seafloor. The vertical axis is the average intended focusing distance along the bottom being scanned by the TRA corresponding to the selected time-gated window. Each plot has its own dB scale. Note the contrast improvement achieved for the strongest reflectors after using one iteration of the time-reversal process. The target location is the brightest spot at L_c=203 m.

cordingly). The effective look direction of the array is now centered at 294°N and the array is illuminating a different search sector. Figure 14 represents the reflectivity maps computed first after passive time reversal alone and after one iteration using a sliding reverberation window of δt=20 ms, using same processing as for Fig. 12. No distinct bright spots appear now along the main diagonal of the reflectivity map in Fig. 14 in comparison to Fig. 12, meaning that no strong reflector was detected in this search sector. This illustrates the target search capabilities over various sector angles of this simple adaptive sonar system using the azimuthal resolution of the billboard array and the (passive) iterative time-reversal process.

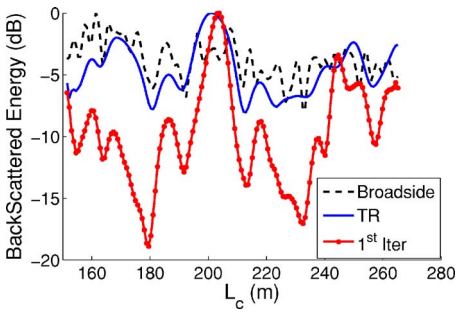


FIG. 13. (Color online) Comparison of the initial broadside return to the diagonal of the reflective map after iterative time reversal.

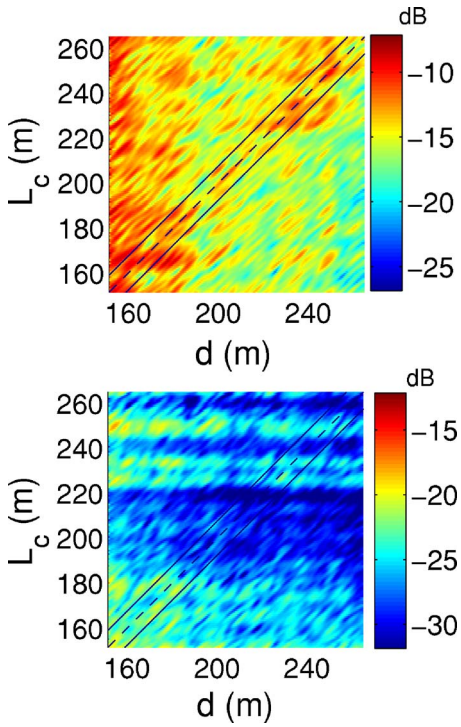


FIG. 14. (Color online) Same as Fig. 12 but the billboard main lobe was 46° away from its axis using simple time-delay beamforming during the initial illumination. For each plot, the normalization for dB scale is identical to Fig. 12. No distinct bright spots appear now along the main diagonal of the reflectivity map in comparison to Fig. 12.

V. CONCLUSIONS

A reverberation focusing technique was presented that allows a time-reversal array to retrofocus energy along the seafloor interface using only measured backscattered signals and minimal *a priori* environmental information. Reflectivity maps of the water-bottom interface were constructed. Iterative time-reversal provides a simple solution for self-adaptive focusing on strong reflectors (i.e., scattering targets) located on the ocean bottom without relying on predictive or modeling capabilities of the environment and of the target of interest. For sufficient reflectivity contrast between the target and the bottom reverberation, target detection was enhanced after one iteration of the time-reversal process. These robust focusing properties are crucial for mapping large and uncharted areas with little signal processing effort involved.

Flush-buried targets appeared clearly, while more deeply buried targets may not be as easily detected without the use of more advanced classification schemes. Furthermore, the temporal length δ_t of the time-gated reverberation window, as well as the number of time-reversal iterations used, should be adjusted given the target range L_c. Mapping the ocean bottom for successive locations of the TRA (e.g., along a ship track) can provide several reflectivity maps of the environment, which could then be used to create an extended synthetic aperture for the TRA. This may help to distinguish a target with constant backscattering return from strong local clutter on the bottom (false alarms), as well as to differentiate among several potential buried targets. Finally a bistatic implementation of the passive iterative time-reversal technique can be developed (i.e., where the source and receiver

role are performed by two different transducer arrays) and might provide bistatic detection enhancement for the case of targets designed to be stealthy in regard to monostatic sonar systems.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research under Contract No. N00014-01-0043-D-7.

[1]M. Fink, D. Cassereau, A. Derode, C. Prada, P. Roux, M. Tanter, J. Thomas, and F. Wu, "Time-reversed acoustics," Rep. Prog. Phys. **63**, 1933–1995 (2000).
[2]D. R. Jackson and D. R. Bowling, "Phase conjugation in underwater acoustics," J. Acoust. Soc. Am. **89**, 171–181 (1991).
[3]W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. R. Jackson, "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **103**, 25–40 (1998).
[4]J. F. Lingevitch, H. C. Song, and W. A. Kuperman, "Time reversed reverberation focusing in a waveguide," J. Acoust. Soc. Am. **111**, 2609–2614 (2002).
[5]C. Prada, M. Fink, and F. Wu, "The iterative time-reversal mirror: a solution to self focusing in pulse echo mode," J. Acoust. Soc. Am. **90**, 1119–1129 (1991).
[6]C. Prada, J. Thomas, and M. Fink, "The iterative time-reversal process: Analysis of the convergence," J. Acoust. Soc. Am. **97**, 62–71 (1995).
[7]H. C. Song, W. A. Kuperman, W. S. Hodgkiss, T. Akal, and C. Ferla, "Iterative time reversal in the ocean," J. Acoust. Soc. Am. **105**, 3617–3184 (1999).
[8]G. Montaldo, M. Tanter, and M. Fink, "Revisiting iterative time reversal processing: Application to detection of multiple targets," J. Acoust. Soc. Am. **115**, 776–784 (2004).
[9]H. C. Song, W. S. Hodgkiss, W. A. Kuperman, P. Roux, and T. Akal, "Experimental demonstration of adaptive reverberation nulling using time reversal," J. Acoust. Soc. Am. **118**, 1381–1387 (2005).
[10]C. Gaumond, D. Fromm, J. Lingevitch, R. Menis, G. Edelmann, D. Calvo, and E. Kim, "Application of dort to active sonar," in Proc. Oceans 2004 (IEEE, New York, 2004), pp. 2230–2235.
[11]W. Hodgkiss, J. Skinner, G. Edmonds, R. Harris, and D. Ensberg, "A high-frequency phase conjugation array," in Proc. Oceans 2001, Hawaii (IEEE, New York, 2001).
[12]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, in Computational *Ocean Acoustics*, edited by R. T. Beyer (AIP, Melville, NY, 2000).
[13]L. Pautet, A. Tessei, P. Guerrini, and E. Pouliquen, "Target echo enhancement using single-element time reversal mirror," IEEE J. Ocean. Eng. **30**, 4912–920 (2005).
[14]T. Folegot, J. de Rosny, C. Prada, and M. Fink, "Adaptive instant record signals applied to detection with time reversal operator decomposition," J. Acoust. Soc. Am. **117**, 3757–3765 (2005).

A passive fathometer technique for imaging seabed layering using ambient noise

Martin Siderius
HLS Research Inc., 12730 High Bluff Drive, Suite 130, San Diego, California 92130

Chris H. Harrison
NATO Undersea Research Centre, Viale S. Bartolomeo 400, 19138 La Spezia, Italy

Michael B. Porter
HLS Research Inc., 12730 High Bluff Drive, Suite 130, San Diego, California 92130

(Received 14 October 2005; revised 20 June 2006; accepted 22 June 2006)

A passive acoustics method is presented that uses the ocean ambient noise field to determine water depth and seabed sub-bottom layering. Correlating the noise field measured by two sensors one can recover a function that closely resembles the two-point Green's function representing the impulse response between the two sensors. Here, a technique is described that is based on noise correlations and produces what is effectively a passive fathometer that can also be used to identify sub-bottom layers. In principle, just one or two hydrophones are needed—given enough averaging time. However, by combining the cross correlations of all hydrophone pairs in a vertical array a stronger signature can be obtained and this greatly reduces averaging time. With a moving (e.g., drifting) vertical array, the resulting algorithm yields both a map of the bottom depth (passive fathometer) and the locations of significant reflectors in the ocean sub-bottom. In this paper, the technique is described and illustrated using numerical simulations. Results are also shown from two experiments. In the first, ambient noise is taken on a fixed array in the 200–1500 Hz frequency band and the second experiment uses a drifting array in the 50–4000 Hz band. [DOI: 10.1121/1.2227371]

PACS number(s): 43.30.Nb, 43.30.Wi, 43.30.Gv, 43.30.Pc [DRD] Pages: 1315–1323

I. INTRODUCTION

Passive techniques that exploit the ocean ambient noise field are useful when active sonar is not practical or feasible. Situations include operations in areas where sonar is prohibited due to, for example, environmental restrictions. In this paper, a technique is described that uses ambient noise correlations to determine the acoustic travel time from hydrophones in the water column to the seabed. This provides a measure of the absolute depth of both the water-sediment interface (a fathometer) and the sub-bottom layers. Vertical beamforming is used to limit the contributions from distant noise sources while emphasizing those directly overhead; this greatly reduces the averaging time required to extract coherent arrivals. A simplified derivation of the noise correlation function is included to illustrate how coherent arrivals from the noise field are used for the passive fathometer processing.

In recent years, several new techniques have been proposed to exploit the ocean ambient noise field for sonar and seismic applications. Harrison and Simons showed that the ratio of the upward to downward directionality of the noise field is the incoherent bottom reflection coefficient, and they measured it by beamforming on a vertical array.[1] That technique was extended to derive sub-bottom layering with a drifting array by reconstructing the reflection loss phase using spectral factorization.[2,3] Roux, Kuperman, and the NPAL Group demonstrated how wave fronts can be extracted from the ocean noise field using horizontally separated hydrophones.[4] Their work was inspired by the developments by Weaver and Lobkis[5] and the conjecture put forward by Rickett and Claerbout:[6] "By cross correlating noise traces recorded at two locations on the surface, we can construct the wave field that would be recorded at one of the locations if there was a source at the other." The wave fronts reconstructed by Roux, Kuperman, and the NPAL Group showed that, in fact, cross correlations between two receivers resembled that from a directional source to a receiver with the directionality dependent on the characteristics of the noise sources. A more detailed derivation of the angularly shaded, two-point Green's function obtained from ocean noise correlation functions was developed by Sabra *et al.*[7] In that paper it was shown that the coherent arrivals are primarily due to the noise sources located in the end fire direction to the hydrophones being cross correlated. Given sufficient averaging time, the cross correlation produces the eigenray arrivals between the two hydrophones.

The work described here exploits the same noise cross-correlation phenomenon. However, closely spaced hydrophones vertically separated are used to take advantage of both cross correlations between sensors and beamforming. This allows for short averaging times, on the order of 30 s, to extract the coherent arrivals. This combination makes it possible to estimate both the water depth and sub-bottom layering from the ambient noise correlation function. In Sec. II, a simplified theoretical description is developed that includes a method of images construction to elucidate the nature of arrivals that are extracted from the noise correlation function.

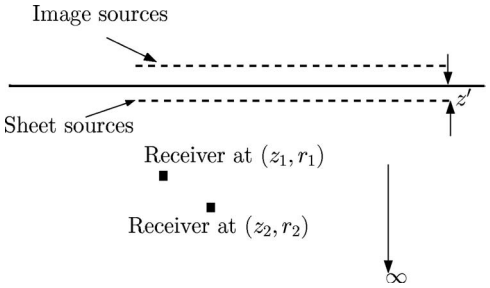


FIG. 1. Geometry for the half-space problem where the perfectly reflecting surface gives rise to image sources at $z=-z'$.

In Sec. III, numerical simulations illustrate the technique under known conditions. Finally, in Sec. IV results are shown from experiments using a fixed vertical array and a drifting vertical array.

II. THEORETICAL DESCRIPTION

Over the last several decades, a number of theoretical approaches have been developed to describe the ocean ambient noise field. In Buckingham,[8] a normal mode approach was used to develop a model for ambient noise in shallow water waveguides. Harrison[9] developed a ray-based approach which is particularly advantageous for broadband computations and for higher frequencies. Here, the wave approach taken by Kuperman and Ingenito[10] is adopted. This approach is used together with the method of images to construct the Green's functions and is similar to that taken by Sabra *et al.*[7] The special geometry of vertically separated receivers used for the passive fathometer processing simplifies the analysis that is presented here. The purpose of these derivations is to demonstrate, for geometries that can be easily treated analytically, how the noise correlations between two receivers can closely resemble the response from a source to receiver.

The sound generated from wind action on the surface is modeled as an infinite sheet of point sources located just below the surface at depth z'. The geometry for the sheet source and receivers is shown in Fig. 1. The derivation and notation follows Refs. 10 and 11. To start, (assuming cylindrical symmetry) the cross spectral density is written as an integral over all source directions,

$$C(\omega,R,z_1,z_2) = \frac{8\pi^2 q^2}{k^2(z')} \times \int_0^\infty [g(k_r,z_1,z')g^*(k_r,z_2,z')]J_0(k_rR)k_r dk_r. \tag{1}$$

This derived cross-spectral density $C(\omega)$ (at frequency ω), between receivers at depths z_1 and z_2 is written in terms of the (range-independent) Green's functions with horizontal wave-number k_r (the * indicates the complex conjugation), $k=\omega/c_w$, and water sound speed c_w. The quantity q is included to give proper scaling due, for instance, to various wind speeds. The Bessel function (J_0) is required for receivers separated in range by $R=r_1-r_2$. For the vertically separated receivers used for the passive fathometer process $J_0(k_rR)=1$.

It is worthwhile to note the similarity in form between the cross-spectral density (1) and the pressure field from a point source at z_1 to a receiver at depth z_2

$$P(\omega,R,z_1,z_2) = \int_0^\infty g(k_r,z_1,z_2)J_0(k_rR)k_r dk_r. \tag{2}$$

The integral in Eq. (2) produces the usual pressure field as a function of range and depth due to a point source. Wavenumber integration methods have been developed which evaluate this integral and are described in Refs. 11–13. An important difference between Eqs. (1) and (2) is that the cross-spectral density is an ensemble average and the pressure field is deterministic. With measured data, the averaging time needed can be an important consideration for noise processing techniques.

A. Calculating the Green's function using the method of images

The similarity between Eqs. (1) and (2) depends on the extent to which the product of Green's functions in Eq. (1) behaves like the single Green's function in Eq. (2). That is, to what extent does the noise correlation behave as a source-receiver pair? To gain insight into this question it is worthwhile to begin with simple Green's functions constructed using the method of images.

The environment is assumed to be a fluid halfspace (i.e., no seabed) with the noise source located at z'. The pressure release surface gives rise to a sheet source of images at $-z'$. The geometry is shown in Fig. 1.

The Green's function to the receiver at z_1 from sources at z' and its image at $-z'$ can be written in terms of the horizontal wave-number k_r

$$g(k_r,z_1,z') = \left[\frac{e^{ik_z|z_1-z'|}}{4\pi ik_z} - \frac{e^{ik_z(z_1+z')}}{4\pi ik_z}\right], \tag{3}$$

where the vertical wave number is defined as $k_z=\sqrt{k^2-k_r^2}$. Since the sound sources are very near the surface (within a fraction of a wavelength) the receivers can safely be assumed deeper and the magnitude sign in the exponential can be omitted. Denoting $g_1=g(k_r,z_1,z')$ and $g_2=g(k_r,z_2,z')$, the term in square brackets in Eq. (1) is

$$g_1g_2^* = \left[\frac{e^{ik_z(z_1-z')}}{4\pi ik_z} - \frac{e^{ik_z(z_1+z')}}{4\pi ik_z}\right] \times \left[\frac{e^{ik_z(z_2-z')}}{4\pi ik_z} - \frac{e^{ik_z(z_2+z')}}{4\pi ik_z}\right]^*. \tag{4}$$

After some manipulation this can be written

$$g_1g_2^* = \frac{1}{(2\pi|k_z|)^2}[e^{i(k_zz_1-k_z^*z_2)}\sin(k_zz')\sin(k_z^*z')]. \tag{5}$$

This is the result for the most common case of a surface of dipole sources (monopoles very near the pressure release sur-

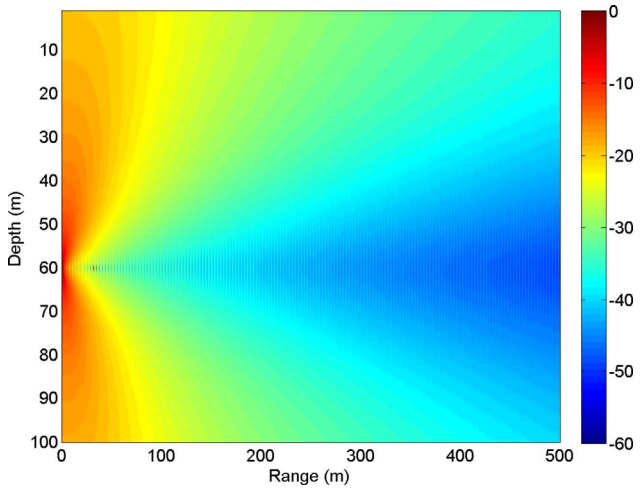


FIG. 2. Cross spectral density, normalized on a decibel scale $[10\log C(\omega)]$ given by Eq. (6). The source appears to be at the location of the receiver at z_1 but is not a point source but has dipole-like shading.

face). This is essentially the result obtained by Cron and Sherman[14] for the deep ocean with limited seabed reflections (the equivalence of this form to the earlier work was derived by Kuperman and Ingenito).[10] The cross-spectral density is written considering only real k_z

$$C(\omega,R,z_1,z_2)=\frac{2q^2}{k^2(z')}\times\int_0^k\left[\frac{e^{ik_z(z_1-z_2)}\sin^2(k_zz')}{|k_z|^2}\right]J_0(k_rR)k_rdk_r. \tag{6}$$

The term in square brackets looks similar to a single, free-space Green's function but *not* originating from a surface source but rather for a source located at z_1 and received at z_2. Further, rather than a true, free-space Green's function there is an extra $\sin^2(k_zz')$ term that gives a dipole-like shading. The cross-spectral density is, therefore, expected to look similar to the pressure field from a shaded free-space point source. A plot of the cross-spectral density given by Eq. (6) is shown in Fig. 2. In the figure, the first noise receiver is fixed at $z_1=60$ m, $r_1=0$ m and this is correlated with receivers at $z_2=0-100$ m, $r_2=0-500$ m (for 500 Hz), note that the source *appears* at (z_1,r_1).

In the previous description k_z was assumed real which makes the expression easier to interpret but is only correct for $k_r<k$. This is only a minor approximation since there is an exponential decay that occurs for $k_r>k$ (i.e., the evanescent part of the wave-number spectrum). Consider expanding the first term in square brackets in Eq. (5)

$$e^{i(k_zz_1-k_z^*z_2)}=e^{i\Re\{k_z\}(z_1-z_2)}\times e^{-\Im\{k_z\}(z_1+z_2)}, \tag{7}$$

where $\Re\{k_z\}$ and $\Im\{k_z\}$ indicate the real and imaginary components of k_z. This indicates that for the part of the wave-number spectrum where k_z has an imaginary component $g_1g_2^*$ will decay exponentially in depth.

Next, the surface and bottom boundaries are included. This will result in an infinite number of image sources, however to show the pattern and for illustration purposes, just a single bottom image and a single surface image are included in this discussion. The geometry is shown in Fig. 3 for a water depth of H. Including the first bottom image yields a superposition of the source at z' and the image at $2H-z'$, that is

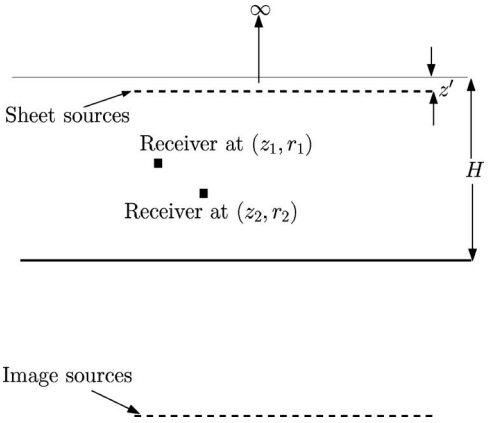


FIG. 3. Geometry for the half-space problem where the reflecting bottom gives rise to image sources at $z=2H-z'$ in addition to the surface image.

$$g_1=\left[\frac{e^{ik_z(z_1-z')}}{4\pi ik_z}-\frac{e^{ik_z[(H-z_1)+(H-z')]}}{4\pi ik_z}\right]. \tag{8}$$

Similarly for g_2

$$g_2=\left[\frac{e^{ik_z(z_2-z')}}{4\pi ik_z}-\frac{e^{ik_z[(H-z_2)+(H-z')]}}{4\pi ik_z}\right]. \tag{9}$$

The product $g_1g_2^*$ is

$$g_1g_2^*=\frac{2}{(4\pi|k_z|)^2}\{\cos[k_z(z_1-z_2)]-\cos(k_z[(H-z_1)+(H-z_2)])\}. \tag{10}$$

The first term in square brackets looks like a point source term between z_1 and z_2. The second term looks similar to an image source from the bottom bounce.

Including both the images from the bottom and the surface, the receiver at z_1 has a Green's function

$$g_1=\left[\frac{e^{ik_z(z_1-z')}}{4\pi ik_z}-\frac{e^{ik_z(z_1+z')}}{4\pi ik_z}-\frac{e^{ik_z[(H-z_1)+(H-z')]}}{4\pi ik_z}\right]. \tag{11}$$

Similarly, the Green's function at z_2

$$g_2=\left[\frac{e^{ik_z(z_2-z')}}{4\pi ik_z}-\frac{e^{ik_z(z_2+z')}}{4\pi ik_z}-\frac{e^{ik_z[(H-z_2)+(H-z')]}}{4\pi ik_z}\right]. \tag{12}$$

The product $g_1g_2^*$ is

$$g_1g_2^*=\frac{1}{(2\pi|k_z|)^2}\left\{e^{ik_z(z_1-z_2)}\sin^2(k_zz')+\frac{1}{4}e^{-ik_z(z_1-z_2)}-\frac{1}{2}\cos(k_z[(H-z_1)+(H-z_2)])+\frac{1}{2}\cos(k_z[(H-z_1)+(H-z_2)-2z'])\right\}. \tag{13}$$

This is a fairly complicated expression, but the first two

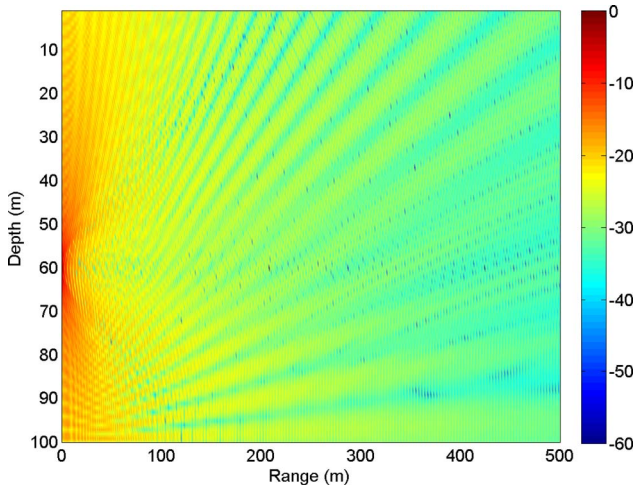


FIG. 4. Cross spectral density, normalized on a decibel scale $[10 \log C(\omega)]$ for the case of perfectly reflecting boundaries (including just the first bottom and surface images). The source appears to be at the location of the receiver at z_1 along with an image giving rise to the inverted Lloyd mirror pattern.

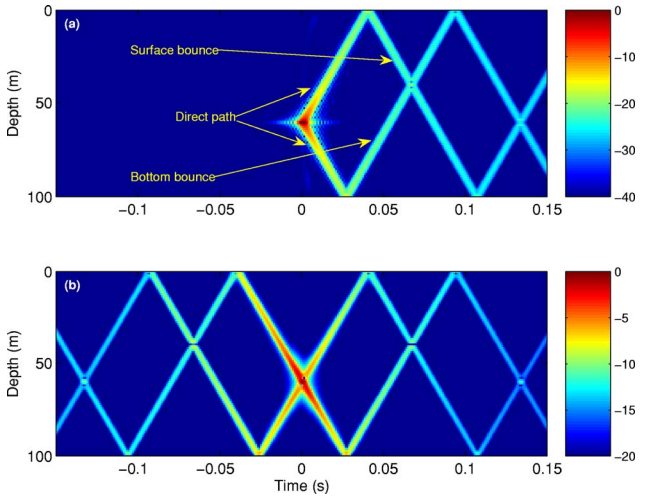


FIG. 5. Panel (a) shows the response on a 100-m vertical array from a point source at 60-m depth (all colocated in range). In panel (b) the noise cross correlations are shown using the reference phone at 60 m correlated with the other hydrophones in the array. The negative part of the time series is due to the cross-correlation process. In both (a) and (b) the envelope of the response is normalized and put on a decibel scale.

terms in curly brackets correspond to the shaded point source between z_1 and z_2. The last two terms correspond to an image of the dipolelike "source" (i.e., the bottom bounce). In other words, to the receiver at z_2 it appears like a shaded source at z_1 and a bottom bounce image. When processing for bottom depth and sub-bottom layering, however, the source directionality will have very little impact since the interest is in the vertically received bottom bounces. Inserting Eq. (13) into Eq. (1) the field can be calculated and this is shown in Fig. 4 for $z_1=60$ m and frequency of 500 Hz. The figure is consistent with the terms in Eq. (13) with the appearance of an inverted Lloyd mirror pattern.[11] Note that for the derivation, k_z is assumed real but for the numerical simulation results shown in Fig. 4, the complex values are included. Numerically, it is not difficult to include the higher order images (i.e., use the complete Green's functions) and this will be done for the time-domain solutions in the following sections.

III. TIME-DOMAIN PROCESSING FOR THE PASSIVE FATHOMETER

Next, consider the cross-spectral density in the time domain where the fathometer and sub-bottom profiler are most useful. Using Fourier synthesis, the frequency domain solution is transformed to the time-domain τ according to

$$c(\tau,z_1,z_2)=\frac{1}{2\pi}\int_{-\infty}^{\infty} C(\omega,z_1,z_2)e^{-i\omega\tau}d\omega. \tag{14}$$

Using the synthesis Eq. (14), the frequency domain correlation of the ambient noise field between two (vertically separated) receivers at z_1 and z_2 is transformed to a time series. Note that with vertically separated receivers there is no R dependency and the Bessel function in Eq. (1) disappears. According to the previous analysis this is expected to look similar to a source at z_1 and a receiver at z_2 with some slight differences in the source directionality as described.

To illustrate, consider a vertical array in a water depth of H=100 m with perfectly reflecting boundaries. The array spans the water column with receivers every meter from 1 to 100 m. To orient the reader, the response on the array due to a true point source located at 60-m depth is shown in panel (a) of Fig. 5. The vertical array and source are all colocated at range 0. The time-domain signal is generated by computing the response from 1 to 512 Hz with 1 Hz sampling and using Fourier synthesis to produce a 1 s time series at each receiver. A Hanning window is applied to the spectrum before transforming in order to reduce time-domain ringing. The direct paths are those that reach the receivers first and are indicated in the figure. Arriving later in time are the bottom and surface bounces that are also indicated. In panel (b) of Fig. 5, the noise cross-correlation time-domain responses are shown. The reference hydrophone z_1 is at 60 m depth and cross correlated with the other receivers in the array. That is, the cross-spectral density is formed between z_1 and all the other receivers at each frequency (from 1 to 512 Hz with 1 Hz sampling). Fourier synthesis is used in exactly the same way as for the true source to produce a 1 s time series at each receiver. For the fathometer application, the interest is in arrival times, and for this the two signals in panels (a) and (b) of Fig. 5 are practically the same for positive time.

Some differences between the cross-correlation time series and that from a true source become evident when a more realistic seabed is substituted for the perfectly reflecting bottom boundary. These differences will not impact the processing for bottom depth and layering as will be demonstrated in the next sections with simulations and data. However, it is important to note these differences to explain some of the features that appear in the intermediate results. The simulation will use the same geometry as for the perfectly reflecting boundaries of Fig. 5 but with an acoustic half-space seabed with 1550 m/s sound speed, density of 1.5 g/cm^3, and attenuation of 0.2 dB/λ (decibels per wavelength). The Green's functions and integral in Eq. (1) are evaluated using the OASES program.[12,15] In panel (a) of Fig. 6 it can be seen

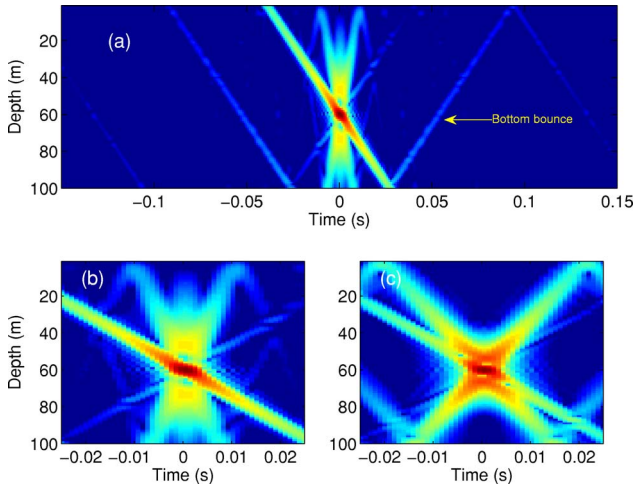


FIG. 6. In panel (a) a realistic (lossy) seabed is used with the same geometry as for panel (b) of Fig. 5. Note, the wavefronts near time-lag zero are propagating at a more horizontal angle relative to the bottom bounce (the M-shape near time zero). In panel (b) is a zoom of (a) showing the details of the wavefront. The propagation angle of the wavefront in (b) corresponds to the critical angle of the seabed. In (c) the seabed sound speed is changed to 1750 m/s and the steeper propagation angle of the wavefront near time-lag zero can be seen.

that the bottom bounce is similar to the case of the perfectly reflecting boundary with the exception of the higher loss on the bottom bounce. One of the differences with this simulation occurs near time lag zero. A zoomed in display in panel (b) shows a wave front arriving more horizontally than the bottom bounce (the M-shape near time zero). The angle of this wave front corresponds to the critical angle of the 1550 m/s seabed. To compare, in panel (c) the seabed sound speed is changed to 1750 m/s and the angle of the wave front has increased. These wave fronts that occur at the critical angle suggest this process is detecting the head wave. These wave fronts, near time-lag zero, are at angles that are different from the bottom bounce and will have insignificant impact on the passive fathometer processing. They may, however, be important for identifying the critical angle of the seabed if they can also be detected using measured data. They will not be considered further in this paper as the focus is on the bottom and sub-bottom returns.

In addition to the wave fronts arriving near time lag zero there are additional features that also occur near time-lag zero. These too, will also have a negligible effect on the final passive fathometer processing but help in identifying features that can be seen in some of the intermediate results. To understand, consider a coherent "click" (or impulse) generated on the surface from various noise events. Again, take the reference receiver $z_{\mathrm{Ref}}=60$ m and two separate receivers $z_1=25$ m and $z_2=95$ m. These coherent sounds will have a direct arrival on each receiver followed by bottom and sub-bottom arrivals (bounces from the seabed). The timing of these impulses depends on receiver depth and are shown in Fig. 7. Since z_1 is above the reference, the correlation of the two for positive lag time (indicated in the figure as $z_{\mathrm{Ref}}{}^{*}z_1$) shows the initial weak arrivals followed later by a perfect reproduction of the bottom and sub-bottom arrivals. For $z_{\mathrm{Ref}}{}^{*}z_2$, the positive lags show the perfect reproduction of the impulses received on z_2 (no faint arrivals). The main point is that regardless of the relative position of the receiver, the correlation processing produces the coherent bottom and sub-bottom arrivals. There can be a complicated pattern near time-lag zero that includes these weaker direct path and other echoes but these will be of no consequence in the processing as will be described in the next section.

A. Processing vertically separated hydrophones for bottom depth and sub-bottom layering

The previous sections outlined what is expected when the measured noise is correlated between two hydrophones and is transformed into the time domain. The resulting time series is very similar to what is produced from fathometers and sub-bottom profiling sonar systems. That is, a monostatic source/receiver pair. As mentioned, there are some differ-

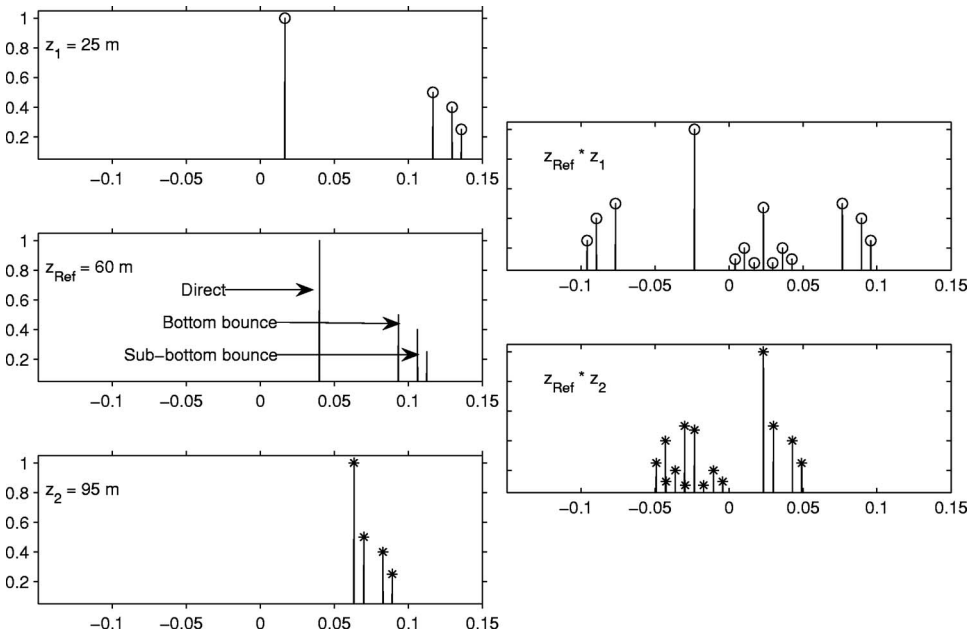


FIG. 7. The impulses received from a coherent "click" from the surface. On the left the received impulses are shown for receivers $z_1=25$ m (top left), $z_{\mathrm{Ref}}=60$ m (middle left), and $z_2=95$ m (bottom left). The right side shows the correlation of the reference with the other two channels, $z_{\mathrm{Ref}}{}^{*}z_1$ (top right) and $z_{\mathrm{Ref}}{}^{*}z_2$ (bottom right). Considering only positive time, the receiver above the reference shows faint arrivals before reproducing the bottom and sub-bottom returns. The receiver below the reference shows a perfect reproduction of the bottom and sub-bottom returns without the faint arrivals.

ences between a true source/receiver and the noise correlation processing. The noise correlation processing produces a time series with some directionality and there are arrivals before time-lag zero. Another difference includes arrivals near time-lag zero due to a variety of effects as described previously. However none of these have a significant impact on the ambient noise fathometer/sub-bottom profiler. The only caveat is that the time-series synthesis must be long enough to prevent wrap around. The time series length is set by the selected frequency sampling in the cross-spectral density and, in practice, can be set arbitrarily small.

Next, consider a more realistic simulation consisting of a layered seabed and a realistic array which will be used for the beam-forming part of the processing. The array has the same characteristics as one that is used for the measured data analysis in Sec. IV and has 32 vertically separated hydrophones located between depths of 70 to 75.58 m (0.18 m spacing). The seabed is made up of a top 10 m layer with sound speed of 1550 m/s, density of 1.5 g/cm^3, and attenuation of 0.06 dB/λ. Below that is a 5 m layer with sound speed of 1600 m/s, density of 1.65 g/cm^3, and attenuation of 0.2 dB/λ. The half-space below both sediment layers has sound speed of 1700 m/s, density of 1.65 g/cm^3, and attenuation of 0.2 dB/λ. To simulate the noise field, Eq. (1) is evaluated using the OASES program.[12,15]

The first fathometer processing step is to correlate each receiver in the array with each of the others to form the cross-spectral density at all frequencies (in this simulation frequencies between 50 and 4000 Hz are included). The time series is derived from the frequency domain correlation between receivers at z_n and z_m using Eq. (14). The bottom bounce absolute travel times will correspond to the travel time from the reference receiver to the seabed plus the travel time back to each of the other receivers in the array. Regardless of the receiver chosen as the reference the bottom bounce arrivals will appear delayed from the bottom of the array to the top. These arrivals can all be combined by applying the correct time delay to align the receptions. The alignment is accomplished through delay and sum beamforming (in the end-fire direction)

$$\tilde{C}_n(\tau) = \frac{1}{M}\sum_{m=0}^{M-1} C(\tau + m\Delta\tau, z_n, z_m), \tag{15}$$

where

$$\Delta\tau = \frac{\Delta z_h}{c_w}. \tag{16}$$

The time delay between receivers is $\Delta\tau$ and Δz_h is the receiver separation and c_w is the water sound speed.

With receiver 1 (shallowest at 70 m) as the reference, the first bottom reflection arrives on channel 32 (at depth 75.58 m) at time

$$\frac{(100-70)+(100-75.58)}{1500} = 0.0363 \text{ s}, \tag{17}$$

and arrives at channel 1 at 0.04 s. The 32 channels are combined with the appropriate delay to form a single time trace. Next, channel 2 is used as the reference and the 32 receivers are beamformed again. Since channel 2 is the reference, the bottom bounce arrivals arrive earlier by $\Delta\tau$ (since the effective source is now closer to the seabed). The process is repeated until 32 time traces are formed. This first stage of beamforming produces an arrival structure for each reference hydrophone $n=1\cdots32$ and is illustrated in panel (a) of Fig. 8. In panel (a) each row is the result using a different reference channel which causes the delayed arrivals from the seabed. The bottom and sub-bottom reflections are visible with travel times relative $n=1$ (receiver at 70 m depth) since all beamforming is done relative to this receiver.

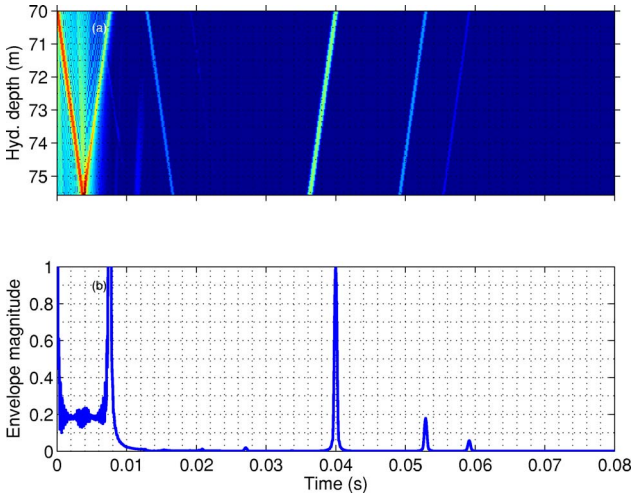


FIG. 8. Panel (a): Stacked, delay, and sum beamforming [i.e., $\tilde{C}_n(t)$ with $n=1\cdots32$]. Note, there is a bottom reflection corresponding to the water-seabed interface and two sub-bottom reflections corresponding to the two layers. Time series magnitudes are shown on a decibel scale with a range of 30 dB. Panel (b): the second stage of beamforming [i.e., beamforming the sequences shown in panel (a) and given by the expression for $r(t)$]. This is on a normalized linear scale where the envelope has been taken.

The time series shown in panel (a) of Fig. 8 is combined again using delay and sum beamforming. This final step aligns the bottom and sub-bottom arrivals into a single time series $r(\tau)$

$$r(\tau) = \frac{1}{N}\sum_{n=0}^{N-1} \tilde{C}_n(\tau + n\Delta\tau). \tag{18}$$

This is shown in panel (b) of Fig. 8 where the bottom and sub-bottom returns are clearly visible. The absolute water depth can be determined from the two-way travel times taken from depth of receiver $n=1$. In this case the first arrival is at 0.04 s corresponding to $(30\times2)/1500$ s. The sub-bottom layers are also two-way travel times relative to the previous bottom reflection. The next arrival occurs after an additional $(10\times2)/1550$ s, or at 0.0529 s. The final arrival is later by and additional $(5\times2)/1600$ s, or at 0.0592 s. There are residual peaks near the zero lag of the time series that can be ignored in the processing since these peaks are predictable from the length of the array. In this case the peaks occur between time zero and time 0.0074 s (twice the travel time of the length of the array due to the two stages of beamforming).

In these simulations, the cross-spectral density that was calculated is idealized and assumes an infinite averaging time. Long averaging is probably not practical for application as a survey tool since the integration times would need to be relatively short. The beamforming greatly assists the shortened averaging time with measured data by emphasizing the coherent part of the noise field coming from directly over the vertical array. As will be seen with the experimental data in the next section, with short averaging times the beamforming is critical to observing the bottom arrivals.

IV. EXPERIMENTAL RESULTS

Two measured data examples of the fathometer processing will be presented. The first example uses measured data of opportunity. The experiment had an active source (10 km from the receiver array) so the ambient noise was carefully windowed from the time series. This experiment had a fixed array with carefully measured array and water depths so it verifies the processing in a known environment. The second example is taken from a more practical scenario with the vertical array drifting over a varying bathymetry and sub-bottom. These arrays are both electronically quiet having electronic noise floor below sea-state zero.

A. Stellwagen Bank: ASCOT01

In 2001, the NATO Undersea Research Centre conducted the ASCOT-01 experiment near the Stellwagen Bank off the Northeast coast of the United States in a site with 101 m water depth.[16] This was primarily a geo-acoustic inversion experiment so sound sources were nearly continuously transmitting with the exception of about 0.5 s of data at the end of each file. Since the vertical array was fixed, these 0.5 s snapshots could be averaged over many snapshots producing, effectively, about a 30 s average. The sound source was located 10 km away and the ambient noise data was believed to be free from multipath and reverberation from the projector. The fixed and well measured geometry of ASCOT-01 provides a useful check of the bottom return timing.

A total of 33 elements with 0.5-m spacing array were used with the top hydrophone measured to be 52.25 m from the seabed. The frequency band considered is 200–1500 Hz. In Fig. 9, panel (a) shows the output after the first stage of beamforming. The bottom bounce is weak yet visible. From the single-phone correlations, the bottom returns were too weak to be visible without beamforming. In panel (b) of Fig. 9, the second stage of beamforming is applied and the bottom bounce is clearly visible. The peak near 0.07 s, in panel (b) puts the estimate of the distance to the bottom at 52.5 m, well within the experimental error on the measured hydrophone distance of 52.25 m. In this case, the array reference was the shallowest channel at 52.25 m which is near the midwater depth, and if a surface bounce were present, it would nearly interfere. To break the symmetry, the array beamforming was shifted to the deepest hydrophone and this is shown in panel (c) of Fig. 9. As predicted, in neither case is there evidence of a surface bounce. This may primarily be due to the beamforming deemphasize this return, but was also seen to be weaker than might be expected from a true source as discussed previously. Sub-bottom returns are also present, but these could not be verified for correctness.

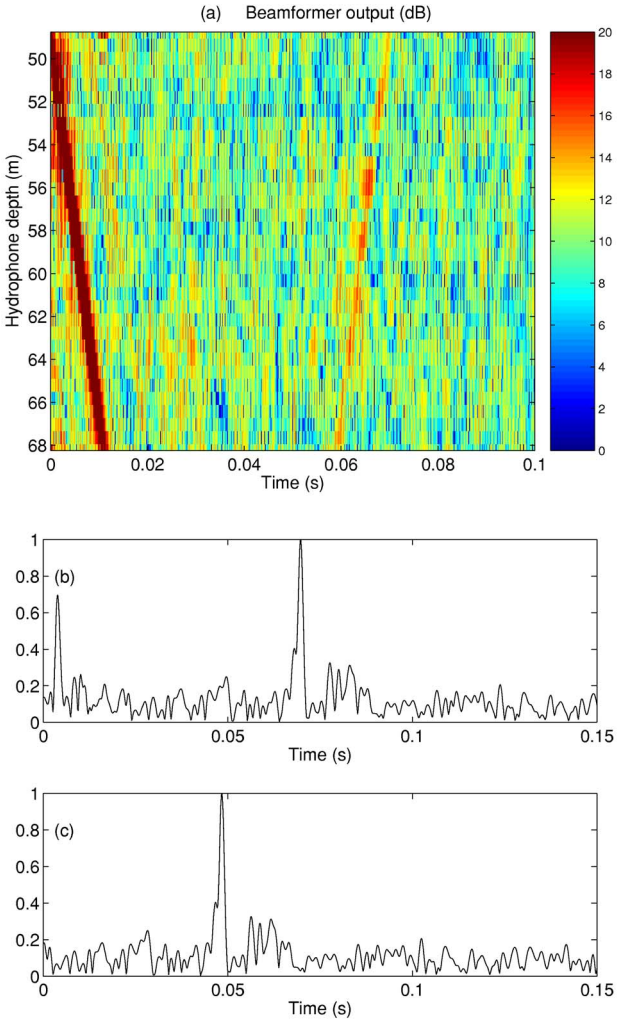


FIG. 9. Panel (a) shows the ASCOT-01 ambient noise processing after the first stage of beamforming. Near time zero, the direct arrival is visible arriving at later times as the wave front progresses down the array. In panel (b) is the second stage of beamforming showing the bottom bounce at around 0.07 s. Panel (c) shows the same processing with the array center shifted to the deepest hydrophone.

B. Strait of Sicily: Boundary 2003

The second experimental example is taken from a controlled set of (directional noise) data that were collected (by the second author) on a drifting array during the NATO Undersea Research Centres Boundary 2003 experiment. (This is the same data set as described in Ref. 3 where it was converted to sub-bottom layers by a different process.) The drifting array has 32 hydrophones spaced at 0.18 m (design frequency of 4.2 kHz). The wind varied during the experiment but was, on average, approximately 15 kn. The array was not equipped with a Global Positioning System receiver but was tracked using surface radar. At the time of the experiment, the depth of the array was not a critical factor and was therefore not measured carefully. However, it was reported that the hydrophones were to be kept less than about 80 m but

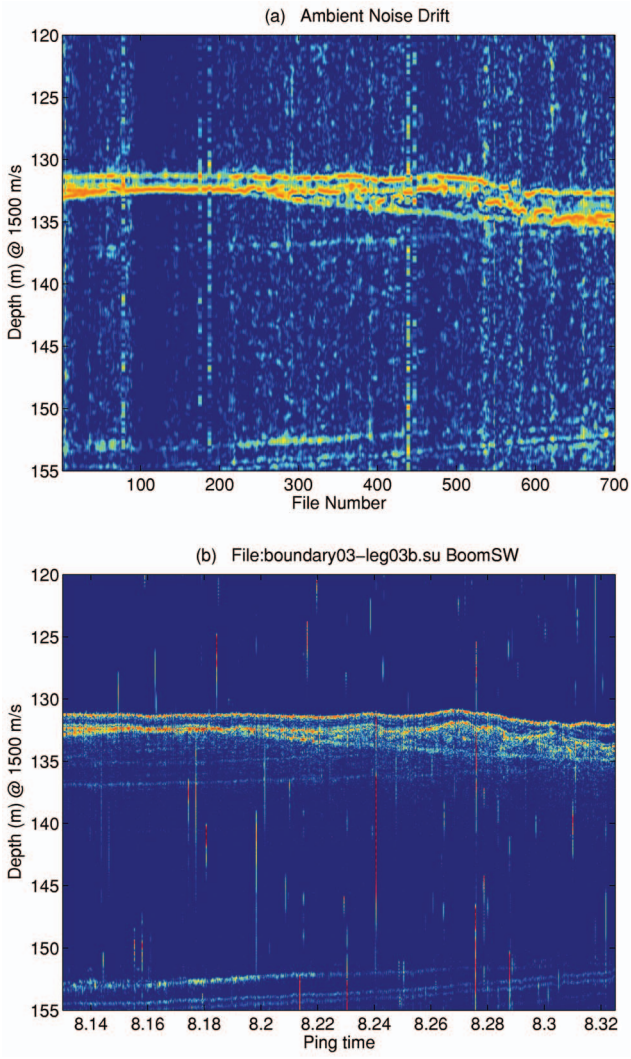


FIG. 10. In panel (a) the ambient noise fathometer processing is used and in panel (b) approximately the same track using a towed Uniboom sub-bottom profiler. The y axis is two-way travel times converted to depths using 1500 m/s sound speed.

were probably between 70 and 80 m. The signals were sampled at 12 kHz and each channel was transformed to the frequency domain using 16384 point fast Fourier transforms (FFTs) (or around 1.4 s of data). Approximately 70 s of data were averaged to form the cross-spectral density and produce a single fathometer time trace (over the 50–4000 Hz frequency band).

Following the array drift, seismic reflection data were collected to image the sub-bottom layers. The seismic reflection data was collected by towing a Uniboom source (5813B seismic boomer) with a ten element towed array behind the NATO *R/V Alliance*. This sonar is designed to measure both the bathymetry and the strongest reflectors from the seabed. It was only possible to approximate the drifting array tracks with the Uniboom tracks.

In Fig. 10, the ambient noise [panel (a)] and Uniboom [panel (b)] processed data are shown. For these displays. the data envelope of the time series are taken and put on a decibel scale. There is a 12 dB dynamic range in the color scale. Since the array depth and position were not known exactly, some alignment of the ambient noise and Uniboom data were made with the data itself. The depth of the array was taken as 73.5 m for the entire track. The range of the array along the track was allowed to slide a few hundred meters. However, a single range correction was used for the entire track. There are features in both (a) and (b) that are similar as far down as 25 m into the seabed. The map of the bathymetry from the noise [panel (a)] also closely resembles that from the Uniboom survey.

V. SUMMARY AND CONCLUSIONS

Using the coherent components of the ocean ambient noise, a passive fathometer and seabed layer imaging technique has been described. In theory, this might be done with a single receiver, however this may require averaging times too long to be practical. Using a vertical array of receivers, beamforming on enfire was used to emphasize the coherent arrivals coming from directly above the array. Beamforming allows averaging the noise over a short time which then resembles random clicks coming from the surface. The processing technique has been illustrated with simulations, however, these simulations assumed infinite averaging time. Experimental data shows the processing also works with reasonable averaging times of about one minute. The first example showed a fixed array with carefully measured geometry and the bottom bounce path arriving at exactly the predicted time. The second example used a drifting array that showed both bathymetry and sub-bottom layering. A sub-bottom profiling sonar was taken along the same path as the drift to validate the results from the passive fathometer processing.

ACKNOWLEDGMENTS

This research was sponsored by the Office of Naval Research Ocean Acoustics Program. Additional support was provided by the Office of Naval Research PLUSNet Program.The authors would like to thank the NATO Undersea Research Centre for the experimental data support. In particular, Jurgen Sellschopp and Peter Nielsen for the ASCOT01 experiment and to Peter Nielsen for the Boundary 2003 experiment. Thanks also to Paul Hursky for many valuable discussions.

[1]C. H. Harrison and D. G. Simons, "Geoacoustic inversion of ambient noise: A simple method," J. Acoust. Soc. Am. **112**(4), 1377–1389 (2002).
[2]C. Harrison, "Sub-bottom profiling using ocean ambient noise," J. Acoust. Soc. Am. **115**(4), 1505–1515 (2004).
[3]C. Harrison, "Performance and limitations of spectral factorization for ambient noise sub-bottom profiling," J. Acoust. Soc. Am. **118**(5), 2913–2923 (2005).
[4]P. Roux, W. A. Kuperman and the NPAL Group, "Extracting coherent wave fronts from acoustic ambient noise in the ocean," J. Acoust. Soc. Am. **116**(4), 1195–2003 (2004).
[5]R. L. Weaver and O. I. Lobkis, "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301 (2001).
[6]J. Rickett and J. Claerbout, "Acoustic daylight imaging via spectral factorization: Helioseismology and reservoir monitoring," Report 100, Stanford Exploration Project (unpublished).
[7]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**(1), 164–174 (2005).

[8]M. J. Buckingham, "A theoretical model of ambient noise in a low-loss shallow water channel," J. Acoust. Soc. Am. **67**(4), 1186–1192 (1980).
[9]C. H. Harrison, "Formulas for ambient noise level and coherence," J. Acoust. Soc. Am. **99**(4), 2055–2066 (1996).
[10]W. A. Kuperman and F. Ingenito, "Spatial correlation of surface generated noise in a stratified ocean," J. Acoust. Soc. Am. **67**(6), 1988–1996 (1980).
[11]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, Inc., New York, 1994).
[12]H. Schmidt, "SAFARI: Seismo-acoustic fast field algorithm for range independent environments. User's guide," SR-113, SACLANT Undersea Research Centre, La Spezia, Italy (unpublished).
[13]H. Schmidt and F. B. Jensen, "A full wave solution for propagation in multilayered viscoelastic media with application to Gaussian beam reflection at fluid-solid interfaces," J. Acoust. Soc. Am. **77**, 813–825 (1985).
[14]B. F. Cron and C. H. Sherman, "Spatial-correlation functions for various noise models," J. Acoust. Soc. Am. **34**, 1732–1736 (1962).
[15]H. Schmidt, "OASES version 3.1 user guide and reference manual," http://acoustics.mit.edu/faculty/henrik/oases.html, Department of Ocean Engineering Massachusetts Institute of Technology, Cambridge, MA (unpublished).
[16]J. Sellschopp, NATO Undersea Research Centre, Vaile S. Barolomeo 400, 19138 La Spezia, Italy, ASCOT-01, SACLANTCEN CD-ROM, 49, 2001.

Joint estimation of water column and sediment acoustic properties from broadband towed array data using modal inverse method

Subramaniam D. Rajan[a)]
Scientific Solutions, Incorporated, 99 Perimeter Road, Nashua, New Hampshire 03063

G. V. Anand[b)] and P. V. Nagesh
Indian Institute of Science, Bangalore, India, 560012

(Received 28 March 2006; revised 7 June 2006; accepted 18 June 2006)

Broadband data, collected on a towed horizontal array of length 2375 m during a geophysical survey, are used to perform joint inversion for the water column and sediment acoustic parameters using modal inverse technique [S. D. Rajan, J. F. Lynch, and G. V. Frisk, J. Acoust. Soc. Am. **82**, 998–1017 (1987)]. To perform the inversion based on modal wave numbers, we first obtain the pressure field as a function of range at a discrete set of frequencies from the broadband data. The sources of errors and their impact on the estimation of modal wave number from such data are discussed. Due to the possibility of large errors in the estimation of the modal wave numbers of the higher-order modes, the inversion was performed using the modal wave numbers of the dominant lower-order modes. The resolution estimates indicate that the wave speeds in the water column and top three sediment layers have resolution lengths that are approximately 1.5 times the layer thicknesses. With respect to deeper sediment layers, the resolution lengths are large, as only lower-order modes were used in inversion. These results are consistent with *a posteriori* variance estimates. The model from modal inverse can serve as the initial model for nonlinear methods, such as those based on matched field processing [P. V. Nagesh, S. D. Rajan, and G. V. Anand, Conference on Underwater Acoustics Measurements: Technologies and Results, Heraklion, Crete, Greece (2005)]. [DOI: 10.1121/1.2221558]

PACS number(s): 43.30.Pc [AIT] Pages: 1324–1334

I. INTRODUCTION

A method for estimating sediment compressional wave speed profiles from the wave numbers of propagating modes in a shallow water waveguide has been reported in the literature.[1] The applicability of this approach has been demonstrated in a number of field experiments.[3,4] In these experiments, the data were collected on a synthetic aperture horizontal array using a harmonic point source. In the analysis presented here, we use broadband data collected on a towed horizontal array by a survey agency conducting a geophysical survey in for an oil company. Due to security and other considerations, the location of the experiment and the sound speed profile in the water column during the experiment was not provided to us. An approximate estimate of the water column depth was, however, provided. We, therefore, investigate joint inversion for water column and sediment compressional wave speeds and their thicknesses by modal inverse method. Even though the theoretical framework for modal inverse has been presented in earlier papers,[1,3,4] its application to broadband data collected on a horizontal array has not yet been studied.

The goal of this paper is to examine the feasibility of using data that are routinely collected by oil companies for estimating the geoacoustic properties of near-surface sediment layers in a shallow water environment. While the analysis conducted by the seismic industry is directed toward defining the layering structure of the sediments (i.e., locating the reflectors/interfaces), we seek a description of the sediment in terms of the wave speeds in the layers and their thicknesses. The horizontal array used in the survey was approximately 2.4 km long. The array had 96 hydrophones with an interelement distance of 25 m, and was towed at a depth of 6 m. The source, an air gun array, was hung from the ship that was towing the array. The center of the source array was at a depth of 6 m. The distance from the source to the closest receiver on the hydrophone array was 200 m. The approximate depth of the water column, as indicated by the survey agency, was 35 m. During the experiment, some of the receivers were blanked out and no data were collected at these receivers. The sampling rate was 500 Hz, and each channel recorded 3 s of data after each transmission. The data acquisition system was designed to trigger the transmitter and the receivers simultaneously. The spectrum of the signal received by the hydrophones on the array indicated that the signal had considerable energy in the frequency interval from 15 to 50 Hz. In this paper, we use the broadband data to first obtain the modal wave numbers at a set of frequencies, and then use these data to jointly estimate the water column properties and sediment compressional wave speed profiles. We consider data acquired from three shots (Shot 5, Shot 50, and Shot 95).

[a)]Electronic mail: rajan@gte.net
[b)]Electronic mail: anandgv@ece.iisc.ernet.in

The signal at each receiver is a time series $p(t,r_n)$ where p is the pressure field, t is the time, and r_n is the range of the nth receiver on the horizontal array. By performing the discrete Fourier transform of the received signal with respect to t, we obtain the pressure field $p(\omega,r_n)$ as a function of range at a discrete set of frequencies. The data, thus transformed into the frequency domain, are used to estimate the water column and sediment acoustic properties.

It can be shown that in a range independent circular symmetric ocean, the pressure field $p(r,\omega)$ at a range r, due to a harmonic point source of radian frequency ω and the wave number spectrum $G(k_r,\omega)$ at the same frequency, are related by the equation[5]

$$G(k_r,\omega)=\int_0^\infty p(r,\omega)J_0(k_r r)rdr, \tag{1}$$

where k_r is the horizontal wave number and J_0 is the Bessel function of zero order. The inverse of this relation, as shown below, yields the pressure field from the wave number spectrum

$$p(r,\omega)=\int_0^\infty G(k_r,\omega)J_0(k_r r)k_r dk_r. \tag{2}$$

In the far field, the argument $k_r r$ has large values, and we can use the asymptotic approximation

$$J_0(k_r r)\approx\frac{\exp[i(k_r r-\pi/4)]}{\sqrt{2\pi k_r r}},\quad k_r r\gg 1, \tag{3}$$

for the Bessel function. Insertion of this approximation in Eq. (1) yields the following relationship for the wave number spectrum.

$$G(k_r,\omega)=\frac{\exp(-i\pi/4)}{\sqrt{2\pi k_r}}\int_0^\infty \sqrt{r}p(r,\omega)\exp(-ik_r r)dr. \tag{4}$$

For any given ω, the wave number spectrum $G(k_r,\omega)$ has peaks whose locations on the wave number axis are the modal wave numbers of the propagating modes at that ω. The modal wave numbers of the propagating modes in a shallow water waveguide are dependent on the acoustic and geometric parameters of the shallow water environment. Rajan *et al.*[1] describes a scheme wherein modal wave numbers obtained from a field experiment can be used to estimate the sediment acoustic properties. We adopt this procedure to estimate the water column and sediment acoustic properties here.

II. ESTIMATION OF MODAL WAVE NUMBERS

The data were collected on a horizontal array towed by a ship carrying the source. This experimental setup was not designed to collect data readily adaptable for estimating the wave number spectrum at various frequencies. As indicated in Sec. I, the acquired data are transformed to obtain the pressure field as a function of range at a given frequency. In using these data to obtain the modal wave numbers, the following issues need to be addressed:

1. Aliasing due to inadequate spatial sampling of the pressure field,
2. Errors in the estimation of modal wave numbers due to errors in range and source/receiver depths,
3. Doppler effect due to array motion,
4. Estimation errors due to the short aperture of data, and
5. Noise in the broadband data.

A. Inadequate sampling of the field

The wave number spectrum at a given frequency can be estimated by using a discretized version of Eq. (4). In order to avoid aliasing, the acoustic field has to be sampled according to the Nyquist criterion, i.e., at intervals of $\lambda/2$ or less, where λ is the lowest wavelength of interest. A much denser sampling ($\lambda/4$) is normally adopted. Hence, at a frequency of 25 Hz and a minimum compressional wave speed of 1500 m/s in the region of interest, the field has to be sampled at intervals of 15 m. The data acquired on the horizontal array have a sampling interval of 25 m, with some of the receivers blanked out.

We now demonstrate that the undersampled field data can be interpolated to give the required sampling rate, without significant error in amplitude and phase, using simulated data similar to the field data. In order to generate the simulated data, we consider a horizontal array with 100 receivers with an interelement spacing of 25 m. A source is placed in the end-fire position at a distance of 250 m from the first receiver on the horizontal array. This configuration simulates the setup used in the collection of the field data that is the subject of study in this paper. The source signal is assumed to be a Gaussian pulse with a central frequency of 35 Hz and a band width of 50 Hz. The signal received at each of the 100 receivers is obtained using the pulse propagation module in OASES.[6] The shallow water environment used in computing the simulated data was representative of the environment where the field data were collected. Further, we assumed that the properties of the medium are range independent. The simulated pressure field, $p_{\mathrm{sim}}(t,r_n)$, is a time series and represents the pressure field at the nth receiver at range r_n. We process these data, as described in Sec. I, and obtain the pressure field $p_{\mathrm{sim}}(\omega,r_n)$ as a function of frequency and range.

As the receiver spacing is 25 m, we have to sample the pressure field on a much finer grid in order to obtain the wave number spectrum over a range of frequencies. The field $p_{\mathrm{sim}}(\omega,r_n)$ is interpolated to give $p_{\mathrm{sim}}(\omega,r_{\mathrm{int}})$, where r_{int} is a set of ranges at a much higher sampling rate so as to avoid aliasing. The interpolation is performed using the cubic spline routine. Since the phase rapidly changes, we slowed down the phase by multiplying the field by the term $\exp(ikr_n)$ before performing the interpolation. After performing the interpolation, the correct phase was restored by multiplying the interpolated field by the term $\exp(-ikr_{\mathrm{int}})$. To check whether this scheme introduces error in the amplitude and phase of the interpolated field, we compute the pressure field $p_{\mathrm{test}}(\omega,r_{\mathrm{int}})$ using the procedure used to obtain $p_{\mathrm{sim}}(\omega,r_n)$, except that we now use a horizontal array where the element spacing corresponds to r_{int}. The magnitude and

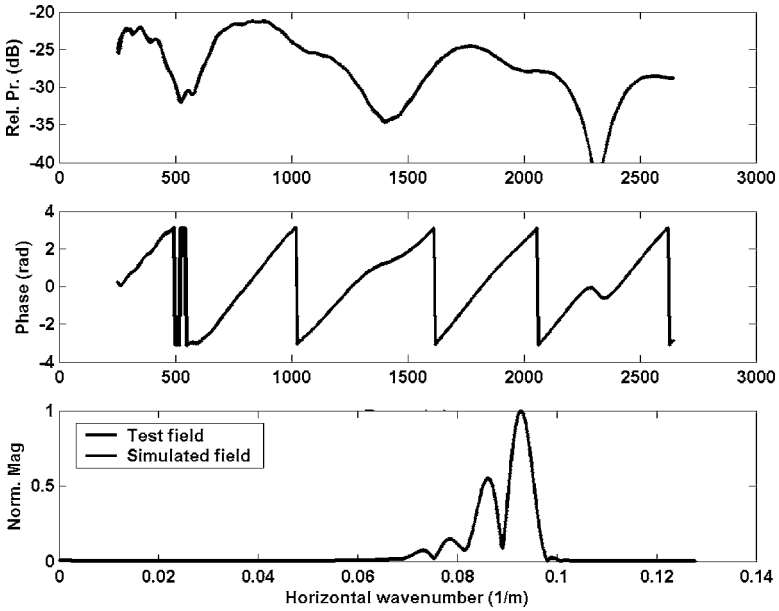


FIG. 1. The top panel shows the comparison between the magnitude of the field obtained with an interelement spacing of 6 m (dots) and the field computed with an interelement spacing of 25 m, and interpolated to a grid spacing of 6 m (firm lines). The source frequency was 25 Hz. The middle panel shows the comparison of the slowed down phase in the two cases. The results from the two cases are identical and, hence, only one line is seen in the plots. The bottom panel shows the wave number spectrum for the two cases, and here again the agreement is excellent.

phase of the pressure field $p_{\text{sim}}(\omega, r_{\text{int}})$ are compared with the amplitude and phase of $p_{\text{test}}(\omega, r_{\text{int}})$ in Fig. 1. These fields were computed at a frequency of 25 Hz. We note that the agreements between the two fields are excellent. A comparison between the wave number spectra obtained by the two methods is also shown in Fig. 1. Similar interpolations performed on an acoustic field at higher frequencies (up to 100 Hz) also showed negligible error arising from interpolation. These simulations show that broadband data, obtained at a set of receivers in a horizontal array can, by suitable processing, be converted into the field due to a harmonic source on a horizontal array, with adequate spatial sampling, in order to obtain the wave number spectrum.

B. Impact of errors in receiver range and depth

We have assumed that the horizontal array is perfectly straight with no distortion. Any distortion of the array will result in errors in the receiver position with respect to its range and depth. In this section, we investigate the effect of errors in receiver positions on the estimation of modal wave numbers. Since we assume circular symmetry for the ocean, lateral displacement of the array is ignored.

Let there be small random errors Δr and Δz in the range and depth of the receivers, whose nominal (assumed) range and depth are r and z_r, respectively. The estimate of the wave number spectrum, obtained from acoustic measurements made on the perturbed array, is given by

$$\hat{G}(k_r,\omega) = \frac{\exp(-i\pi/4)}{\sqrt{2\pi k_r}} \int_{r_0}^{R+r_0} \sqrt{r} p(r+\Delta r, z_r+\Delta z;\omega) \times \exp(-ik_r r)dr, \tag{5}$$

where R is the aperture of the horizontal array and r_0 is the distance between the source and the nearest receiver on the array. The acoustic pressure $p(r,z_s,z_r)$ for a given source and receiver location is the sum of contributions from the set of propagating modes and is given by the expression[7]

$$p(r,z_s,z_r) = \frac{(1+i)}{4\sqrt{\pi}} \sum_{n=1}^{N} \frac{1}{\sqrt{k_n r}} \phi_n(z_s)\phi_n(z_r)\exp(ik_n r), \tag{6}$$

where N is the total number of propagating modes, k_n is the wave number of the nth mode, $\phi_n(z_s)$ and $\phi_n(z_r)$ are the values of the nth mode function at the source and receiver depths, and r is the range to the receiver from the source. Hence, the estimate of the wave number spectrum is given by the expression

$$\hat{G}(k_r) = \frac{(1+i)}{4\sqrt{\pi}} \sum_{n=1}^{N} \phi_n(z_s)\phi_n(z_r+\Delta z) \int_{r_0}^{r_0+R} \frac{1}{\sqrt{k_n k_r}} \times \exp[ik_n(r+\Delta r)]\exp(-ik_r r)dr. \tag{7}$$

Assuming that the perturbation Δr is a Gaussian random variable with zero mean and variance σ^2, we have $E[\exp(ik_n\Delta r)]=\exp(-k_n^2\sigma^2/2)$, where $E[.]$ denotes the expectation or statistical average. If the depth perturbations Δz is small with zero mean, we have, ignoring the higher-order terms,

$$E[\phi_n(z_r+\Delta z)] \approx E[\phi_n(z_r)+\Delta z\phi_n'(z_r)] = \phi_n(z_r). \tag{8}$$

Hence, assuming that the range perturbation Δr and the depth perturbation Δz are statistically independent, we have

$$E[\hat{G}(k_r)] = \frac{(1+i)}{4\sqrt{\pi}} \sum_{n=1}^{N} \frac{\exp(-k_n^2\sigma^2/2)}{\sqrt{k_n k_r}} \phi_n(z_s)\phi_n(z_r) \times \int_{r_0}^{r_0+R} \exp[i(k_n-k_r)r]dr. \tag{9}$$

Therefore, small perturbations of the receiver range and depth result in a reduction in the estimated amplitude of the averaged wave number spectrum, and do not alter the location of the peaks, i.e., the modal wave numbers remain unchanged. While this is true for the averaged wave number spectrum, where the average is done over a large number of realizations, the field data have a single realization and there-

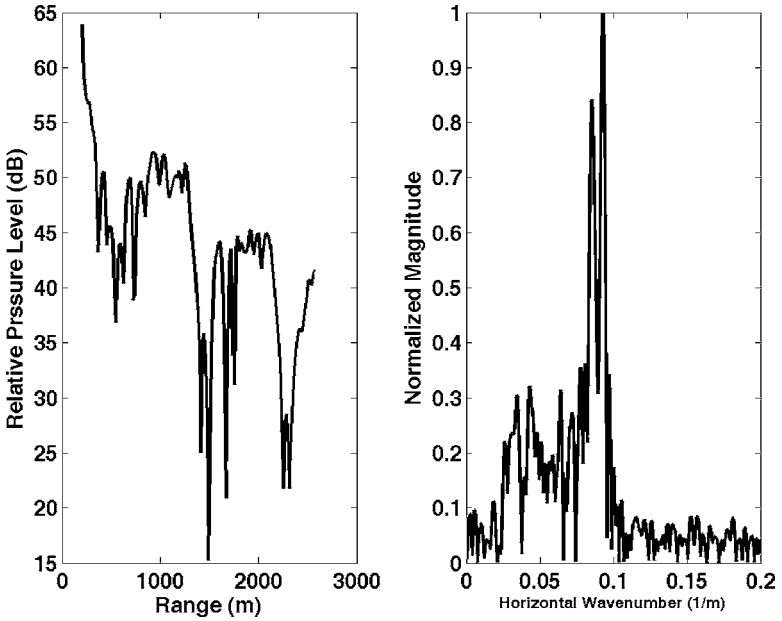


FIG. 2. The left panel shows the field at 25 Hz determined from Shot 50 data. The right panel shows the wave number spectrum. The wave number spectrum shows considerable energy beyond the water wave number. This results in the oscillatory structure seen in the pressure field. See the similarity between this and the plots in Fig. 3.

fore have phase errors that cause errors in the estimated modal wave numbers. Simulations indicate that the effect of an array perturbation on the estimation of modal wave numbers is negligible in the case of dominant lower-order modes at the frequencies of interest. Even in the case of weak higher-order modes, the errors in the estimation of modal wave numbers are small.

C. Errors due to source and receiver motion

In determining the wave numbers of the propagating modes, two sources of errors due to source/receiver motion have to be taken into account. The ship that tows the horizontal array is moving at an approximately constant speed. The source is suspended from the same ship that tows the receiver array. At the instant of source activation, we can consider both the source and the receivers to be stationary. Subsequently as the ship moves it does not affect the signal already transmitted, and hence we need to consider only the effect of receiver motion. Due to the receiver motion, the range to the receiver—at the time the signal arrives at the receiver—will be less than the nominal range based on the number of elements in the array and the interelement distance. The distance of each receiver, at the time the signal arrives at that receiver, is approximately the range at zero time (the time of shot activation) minus the distance traveled by the receiver, which is the ship speed multiplied by the time taken by the signal to reach the receiver.

The second correction to be applied is the Doppler correction, which is the impact of the receiver motion on the horizontal wave number. The corrected wave number is given by the expression $k_n^*=k_n[1+(v_r)/v_{gn}]$,[8,9] where k_n^* is the corrected wave number for the nth mode, v_r is the receiver speed, v_{gn} is the group velocity of the nth mode, and k_n is the wave number for a stationary receiver. The wave number values obtained from the experiment are to be corrected using the above expression in order to obtain the values corresponding to a stationary array. The frequency that the receiver senses will also be shifted due to its motion, and the shifted frequency ω^* is given by the approximate expression $\omega^* \approx \omega[1+v_r/c_w]$, where c_w is the sound speed in the water column. An exact expression for the frequency shift is found in Ref. 9.

D. Processing short aperture data

If the wave number spectrum is obtained using Eq. (1), it will not be possible to resolve modes whose separation in the wave number space is less than $2\pi/L$, where L is the aperture of the array. The number of propagating modes and their wave numbers depend on the frequency of the acoustic source, the acoustic properties of the water column and sediment layers, and the layer thicknesses. Since we are seeking to determine these very acoustic properties of the shallow water environment, which are unknown, we examined the structure of the wave numbers over a range of frequencies (15 Hz–50 Hz) using the ocean model used to generate the simulated data. As we had indicated, the ocean model used to generate the simulated data is representative of the environment where the field data were collected. The lower-order modes in these cases were well separated and resolvable. The higher-order modes were however irresolvable. In the event that some of the modes are not well resolved, high-resolution methods that use schemes, such as the autoregressive spectral estimation method,[10] MUSIC,[11] and ESPRIT,[11] can be employed.

E. Noise in broadband data

Finally, we study the effect of noise in the data. For this purpose, we use the data simulated as described in Sec. II A and add a small Gaussian distributed noise to the signal at each receiver. The noise is assumed to be uncorrelated between the receivers. The amount of noise added was such as to mimic the behavior observed in the field data, namely the oscillatory nature of the magnitude of the field—as shown in Fig. 2. In Fig. 3, the effect of noise in the data is shown. The left panel shows the pressure field for noise-free and noisy

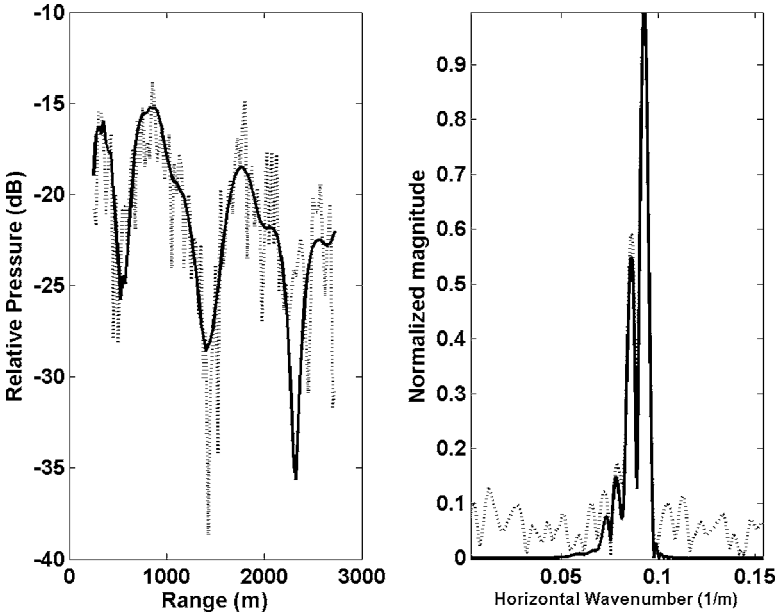


FIG. 3. The left panel indicates the field when the horizontal array data are corrupted by noise. This is compared with the field from the noise-free case. The right panel shows the wave number spectra for the two cases.

simulated data. On the right panel, we compare the wave number spectrum obtained with the noisy data with the spectrum for the noise-free data. We note that the lower-order modes, which are strong, can be easily detected in the wave number spectrum obtained from the noisy data. However, the higher-order modes have errors in their magnitude and location. Further, the wave number spectrum is also noisy. These simulations show the difficulty in estimating the wave numbers of the weak higher-order modes when noise is present in the data.

III. INVERSION METHOD

We now briefly describe the method for estimating the water column and sediment properties from modal wave numbers. This approach has been described extensively in the literature.[1,3,4] The wave numbers of the propagating modes in the waveguide are obtained from the wave number spectrum. Let k_n be the wave number of the nth mode, as obtained from the field measurement of the pressure field. Let us assume that the unknown quantities are the compressional wave speed in the water column and the sediment layers. All other parameters of the sediment layers and the water column are assumed known. We now assume that the compressional wave speed profile, representing the water column and the sediment layers, is given by $C_b(z)$. This initial estimate is normally obtained from archival data or other sources of information. Let this initial estimate be related to the true, but unknown, compressional wave speed profiles by the equation $C(z)=C_b(z)+\Delta C(z)$, where $\Delta C(z)$ is a small quantity which is the difference between the initial estimate of the compressional wave speed profile and the true profile. By solving the forward problem, we determine modal wave numbers for the initial ocean model. Let these be k_{bn}. In a range independent ocean environment, the difference, Δk_n, between the wave numbers obtained from the field measurements and the wave numbers corresponding to the assumed ocean model are related to $\Delta C(z)$ by the integral equation.[1]

$$\Delta k_n = \frac{-1}{k_{bn}} \int_0^\infty \frac{\omega^2 \Delta C(z)}{C_b^3(z)\rho(z)} |\phi_{bn}(z)|^2 dz. \tag{10}$$

In Eq. (10), which is obtained by a linearizing process, $\phi_{bn}(z)$ is the eigenfunction of the nth mode for the assumed ocean model, and ρ is the known density of the sediment layers. By solving this integral equation, we can obtain the quantity $\Delta C(z)$ and, hence, the true compressional wave speed profile. Since the integral equation is nonlinear, the solution is obtained iteratively. This approach has been extended to include inversion for sediment density profile and the attenuation profile.[12,13]

Let us now consider an ocean model in which the horizontally stratified medium is represented as a set of homogeneous layers, i.e., the compressional wave speed in each layer is a constant. Let there be M layers, where Layer 1 represents the water column and the remaining $M-1$ layers represent the sediment. We can then write Eq. (10) as

$$\Delta k_n = \frac{-1}{k_{bn}} \left[\frac{\omega^2 \Delta C_1}{C_{b1}^3 \rho_1} \int_0^{z_1} |\phi_{bn}(z)|^2 dz + \cdots + \frac{\omega^2 \Delta C_M}{C_{bM}^3 \rho_M} \int_{z_{M-1}}^{z_M} |\phi_{bn}(z)|^2 dz \right]. \tag{11}$$

In Eq. (11), the quantities ρ_m and C_{bm} are the density and the initial estimate of the compressional wave speed in the mth layer, respectively, and ΔC_m is the correction to the compressional wave speed in the mth layer that we seek to determine. For a set of N modal wave numbers, Eq. (11) can be represented by the matrix equation

$$\mathbf{\Delta k} = \mathbf{A} \mathbf{\Delta C}, \tag{12}$$

where $\mathbf{\Delta k}$ is a vector containing the wave number differences for N modes, $\mathbf{\Delta C}$ is a vector of length M containing the corrections to the compressional wave speed for the M layers, and $\mathbf{A}$ is an $N \times M$ matrix whose element A_{nm} is given by the expression

$$A_{nm} = \frac{-1}{k_{bn}} \frac{\omega^2}{C_{bm}^3 \rho_m} \int_{z_{m-1}}^{z_m} |\phi_{bn}(z)|^2 dz. \quad (13)$$

The element A_{nm} represents the change in the modal wave number of the nth mode due to unit change in the compressional wave speed in the mth layer. It is apparent from this formulation that the matrix $\mathbf{A}$ and the vector $\mathbf{\Delta C}$ can be augmented to include estimation of other parameters. For example, if we need to estimate the water column thickness, we can augment the matrix by one column with each element of this column representing the sensitivity of the corresponding modal wave number to changes in the water column thickness, and the vector $\mathbf{\Delta C}$ augmented to include Δh the correction to be applied to the water column thickness. The sensitivity of the modal wave number to changes in the water column thickness can be obtained using the expression derived in Ref. 14.

We have noted earlier that noise in the data creates noise in the wave number spectrum, which affects the values of the wave numbers of the higher-order modes. The inversion was therefore performed using the wave numbers of the dominant modes, viz. the lower-order modes, which are much less affected by noise in the data. The wave numbers of the dominant modes at 18 Hz, 20 Hz, 25 Hz, and 30 Hz were used as data. The integral Eq. (10) was replaced by a set of equations, as shown below, to incorporate multifrequency inversion:

$$\Delta k_n(\omega_p) = \frac{-1}{k_{bn}(\omega_p)} \int_0^\infty \frac{\omega_p^2 \Delta C(z)}{C_b^3(z)\rho(z)} |\phi_{bn}(z, \omega_p)|^2 dz;$$
$$n = 1, \cdots, N, \quad p = 1, \cdots, P. \quad (14)$$

This set of integral equations is reduced to a matrix equation of the form

$$\begin{bmatrix} \mathbf{\Delta k}(\omega_1) \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{\Delta k}(\omega_\mathbf{P}) \end{bmatrix} = \begin{bmatrix} \mathbf{A}(\omega_1) \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{A}(\omega_\mathbf{P}) \end{bmatrix} \begin{bmatrix} \mathbf{\Delta c}_1 \\ \cdot \\ \cdot \\ \mathbf{\Delta c}_\mathbf{M} \end{bmatrix}, \quad (15)$$

where the left-hand side is a vector made up of P subvectors $\mathbf{\Delta k}(\omega_\mathbf{p})$. Each of these subvectors represents the differences in wave numbers at frequency ω_p. On the right-hand side, the matrices $\mathbf{A}(\omega_\mathbf{p})$ are the kernels of the integral equations corresponding to the different frequencies. The vector on the right-hand side is a vector of the corrections to the compressional wave speed of the layers, and this is determined by solving the matrix Eq. (15).

The following assumptions are made in performing the inversion:

1. The sediment layers are assumed to be fluid, and
2. The density and attenuation of the sediment layers are assumed to have a constant value in each sediment layer and in the termination half-space. These values are assumed to be known. The acoustic field in the water column is most sensitive to the compressional wave speed of the sediments and water column, and its sensitivity to sediment density and attenuation is small in comparison. Therefore, errors in the assumed values of density and attenuation of the sediment layers will not affect the estimates of the sediment compressional wave speed.

IV. INVERSION RESULTS

A. Estimate of compressional wave speed profiles and water column thickness

As stated earlier, the acoustic parameters of the water column were not provided by the survey agency that collected the data. Only an approximate value of the water column thickness was provided. It was therefore necessary to consider the water column sound speed and the water column thickness as unknown parameters in the inversion process.

Initial estimates of the sound speed in the water column and the water column thickness were made assuming a Pekeris model for the ocean. Modal wave numbers of Mode one at frequencies from 80 to 100 Hz were obtained from Shot 50 data. The inverse problem was formulated with three unknown parameters; namely, the sound speed in the water column, the water column thickness, and the compressional wave speed in the half-space representing the sediments. At these frequencies, the influence of the mode function is mostly restricted to the water column; and this influence decays as we move away from the water sediment interface. This preliminary estimate gave the water column sound speed as 1540 m/s, the water column thickness as 36.9 m and the sound speed in the sediment as 1700 m/s.

The initial ocean model, therefore, consisted of a water column that was 36.9 m thick with the sound speed in the water column set as 1540 m/s. This is followed by three homogeneous sediment layers of thickness 35 m, in which the compressional wave speed was set at 1700 m/s. The density and compressional wave attenuation in the sediment layers were assumed to be 1.7 gm/cc and 0.1 dB/acoustic wavelength. The half-space beyond the layers was assumed to have a compressional wave speed of 2100 m/s. This value was based on the phase speeds of the propagating modes at the frequencies considered for analysis. The values of density and attenuation in the half-space were also set at 1.7 g/cm^3 and 0.1 dB/acoustic wavelength.

The inversions were then carried out for the compressional wave speed in the water column and sediment layers, and for the thickness of the water column. The wave numbers of Modes 1 and 2 at frequencies of 18 Hz, 20 Hz, and 25 Hz, and Mode 1 only at 30 Hz, were used as data. The thicknesses of the sediment layers were excluded from the estimation process, because the inversion process with low-frequency data yields only the depth-averaged compressional wave speed over the assumed layer thicknesses. The sensitivity of modal wave numbers to changes in the sound speed in the water column and the sediment layers were obtained from the expressions in Eq. (13). The sensitivity of the modal wave number to changes in water column depth was obtained numerically.

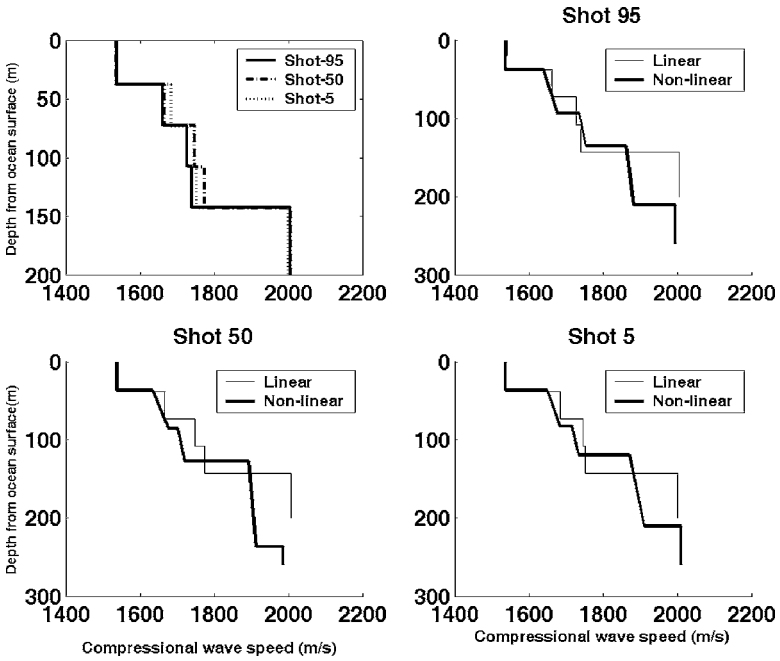


FIG. 4. The top left panel shows the compressional wave speed profiles determined from the three shot data. The remaining panels show the comparison between the profiles obtained from linear inverse method based on the modal wave numbers and the mean value of layer speeds determined through Gibb's sampling (Ref. 2).

The models obtained by the linear inverse for the three shot data are shown in the top left panel of Fig. 4. There are no *in situ* measurements of sediment properties to validate the compressional wave speed models obtained from modal inverse. We, therefore, evaluate this model by comparing:

(a) The model with that obtained from a nonlinear method,
(b) The pressure field predicted by the model with the field obtained from the measured field data, and
(c) The wave number spectra from the model with the spectra obtained from the measured field data.

For comparison, the other three panels in Fig. 4 show the compressional wave speed models obtained from the nonlinear method based on matched field processing,[2] together with the models from linear inverse. The values in the figure for the wave speeds, as determined by a nonlinear method, are the mean values obtained by Gibb's sampling.[2] It is seen that the estimates of the water column sound speed and thickness, obtained by the linear method, agree well with the estimates from matched field processing. The compressional wave speed in the top three sediment layers follow the trend seen in the model from matched field processing. To make the comparison meaningful, the mean values and the deviation estimates from the Gibb's sampling results were depth averaged over the layer intervals in the linear model, and are presented in Table I. It can be seen in Table I that the modal inverse estimates of the compressional wave speed, in the sediment layers up to a sediment depth of 105 m, are within two times the standard deviation from the mean values predicted by Gibb's sampling for models from Shots 50 and 95. However, in the case of Shot 5, similar agreements between the estimates exist only up to a sediment depth of 70 m.

The fields predicted by the model from modal inverse are compared with the field obtained from the field measurement in Fig. 5. We note that the field from the measured field data has a lot of structure. We have shown, by using synthetic data (Fig. 3), that the structure in the measured field is the result of noise in the data. The model-predicted field follows the general trend of the measured field at the lower frequencies. In these cases, the fields are dominated by the lower-order modes, and the wave numbers of these modes were used in the inverse. A comparison of the wave number spectra obtained from field measurements at 20 Hz and the model predicted spectrum is shown in Fig. 6. These plots show good agreement with respect to Modes 1 and 2.

Finally, the correlation between the measured field and

TABLE I. Comparison of wave speeds in the sediment layers. In the nonlinear method (Ref. 2) (matched field processing), both the thickness and the wave speeds of the sediment layers were estimated. In the linear inverse method based on modal wave numbers, the sediment layers were assumed to be 35 m thick and with homogeneous properties. To make the comparison meaningful, the profiles obtained by the nonlinear method were depth averaged to estimate the wave speeds in the layering adopted by the linear method.

Depth into sediment (m)	Shot 5			Shot 50			Shot 95		
	Mean	Deviation	Model	Mean	Deviation	Model	Mean	Deviation	Model
0–35	1664	13.5	1684	1656	13.7	1667	1658	14.2	1662
35–70	1720	25.9	1745	1705	24.5	1748	1694	20.0	1727
70–105	1836	23.8	1754	1804	26.1	1774	1779	27.1	1749
105–175	1890	19.1	2008	1892	18.7	2008	1898	18.6	2005

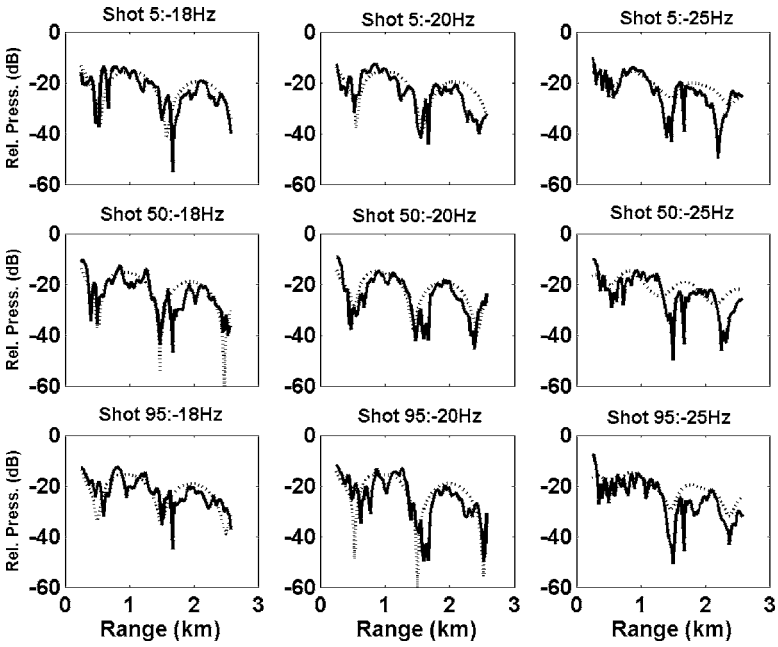


FIG. 5. The magnitude of the pressure field at three different frequencies obtained from the field data (full lines) are compared with the fields predicted by the models (dotted lines) for Shots 5, 50, and 95.

the predicted field are given in Table II. The correlation between the measured field and the predicted field is given by the expression

$$C(\mathbf{P}_{\text{meas}},\mathbf{P}_{\text{predict}}) = \frac{|\mathbf{p}_{\text{meas}}^{\mathbf{H}}\mathbf{p}_{\text{predict}}|}{\sqrt{\mathbf{p}_{\text{meas}}^{\mathbf{H}}\mathbf{p}_{\text{meas}}}\sqrt{\mathbf{p}_{\text{predict}}^{\mathbf{H}}\mathbf{p}_{\text{predict}}}}, \tag{16}$$

where $\mathbf{p}_{\text{meas}}=[p_{\text{meas}}(1)\cdots p_{\text{meas}}(N)]^T$ is the vector containing the measure field at N receivers and $\mathbf{p}_{\text{predict}}=[p_{\text{predict}}(1)\cdots p_{\text{predict}}(N)]^T$ is a vector containing the model-predicted fields at the same set of receivers. The superscripts T and H denote transpose and conjugate transpose, respectively.

The correlation has fairly high values for lower frequencies, and reduces appreciably for 30 Hz. As already mentioned, the data used in the inversion included the wave number of Mode 1 only at the frequency of 30 Hz. Hence, the models obtained with these data are not likely to have adequate resolution to predict the field of 30 Hz. This has resulted in a low correlation value at this frequency.

B. Resolution and variance of estimates

1. Resolution

The linearized version of the inverse problem represented by Eq. (15) is of the form

$$\mathbf{Gm} = \mathbf{d} + \boldsymbol{\nu}, \tag{17}$$

where $\mathbf{G}$ is a matrix containing the kernel, $\mathbf{m}$ is the model to be estimated, $\mathbf{d}$ is the data vector, and $\boldsymbol{\nu}$ represents noise in the data. If the problem is linear, the resolution matrix $\mathbf{R}$ is given by the expression[15]

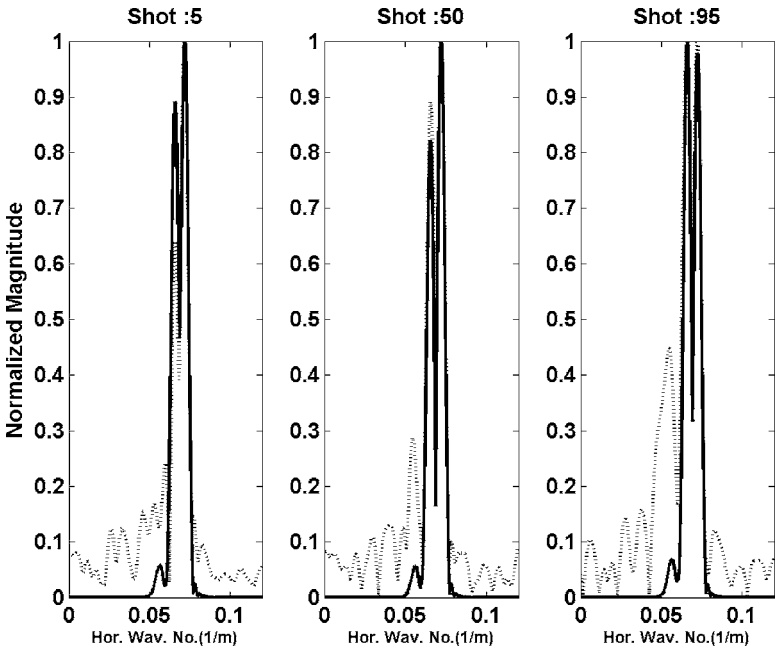


FIG. 6. The wave number spectrum obtained from field data (dotted lines) is compared with the spectrum obtained from the predicted field (full lines). The spectra from the three shots at 20 Hz are shown.

TABLE II. Correlation values between the measured field and the predicted field obtained from modal inverse models.

Frequency (Hz)	Shot 5	Shot 50	Shot 95
18	0.83	0.66	0.70
20	0.86	0.90	0.87
25	0.83	0.84	0.73
30	0.80	0.55	0.62

$$\mathbf{R} = (\mathbf{G}^{\mathbf{T}}\mathbf{C}_{\boldsymbol{\nu}}^{-1}\mathbf{G} + \mathbf{C}_{\mathbf{m}}^{-1})^{-1}(\mathbf{G}^{\mathbf{T}}\mathbf{C}_{\boldsymbol{\nu}}^{-1}\mathbf{G}). \tag{18}$$

In Eq. (18), $\mathbf{C}_{\boldsymbol{\nu}}$ and $\mathbf{C}_{\mathbf{m}}$ are the data and model covariance matrices, respectively. When the nonlinear inverse problem is solved iteratively as a sequence of linear inverse problems, the matrix $\mathbf{G}$ in Eq. (18) is replaced by a matrix that uses the final solution to determine the elements of the resolution matrix $\mathbf{R}$.

The data errors were assumed to be uncorrelated, yielding a diagonal matrix for the data covariance matrix. The error ε in the estimation of the modal wave numbers is related to the aperture of the horizontal array and the signal to noise ratio (SNR), and is given by the expression[1] $\varepsilon = 2\pi/L\sqrt{\mathrm{SNR}}$, where SNR is the signal to noise ratio and L is the length of the horizontal array over which the acoustic field is measured. The data errors were obtained using a SNR based on the average value of the signal to noise of the modal peaks, as determined from the wave number spectrum. The model covariance matrix was also assumed to be diagonal with the standard deviation of the error in the compressional wave speed of the sediment layers estimated as 20 m/s, and that of the sound speed in the water column assumed to be 5 m/s. The estimate of model variance for the sediment layers is based on the compressional wave speed variability for silty sediments.[16] A much smaller variability has been assumed for the sound speed in the water column.

The resolution matrix determines how well the estimated parameters are resolved. If they are well resolved, then the resolution matrix will be a diagonal matrix. Its deviation from the ideal is determined by the resolution length which is given by the expression

$$rl(i) = \frac{\sum_{j=1}^{M} R_{ij}^2 dz}{R_{ii}^2}, \tag{19}$$

where $rl(i)$ is the resolution length of the ith layer, dz is the layer thickness, and R_{ij} is the element in the ith row and jth column of the $\mathbf{R}$ matrix. The resolution length is a measure of the interval over which the estimate is averaged. It is seen from Eq. (19) that if $\mathbf{R}$ is a diagonal matrix, the resolution length of each layer is equal to the layer thickness.

The left panels of Fig. 7 show the resolution length of each layer for one of the three shots. The plot shows the resolution length for six layers. In our inverse, we obtained the wave speed for the water column and three sediment layers. Layers 5 and 6 represent the half-space. It is seen that for the top four layers the resolution lengths are approximately 47 m. The subsequent resolution length for the half-space is much higher. It is to be noted that only lower-order modes were used in the inversion; hence, the poor resolution at greater depths.

2. Variance

For the linear problem of Eq. (17), the variance of the solution can be determined from the expression[16]

$$\hat{\mathbf{C}}_{\mathbf{m}} = (\mathbf{G}^{\mathbf{T}}\mathbf{C}_{\boldsymbol{\nu}}^{-1}\mathbf{G} + \mathbf{C}_{\mathbf{m}}^{-1})^{-1}, \tag{20}$$

where $\hat{\mathbf{C}}_{\mathbf{m}}$ is the *a posteriori* covariance matrix of the model parameters. The diagonal of this matrix shows the variance of the model parameters. One would expect the variance of model parameters to reduce after inversion. If the variance

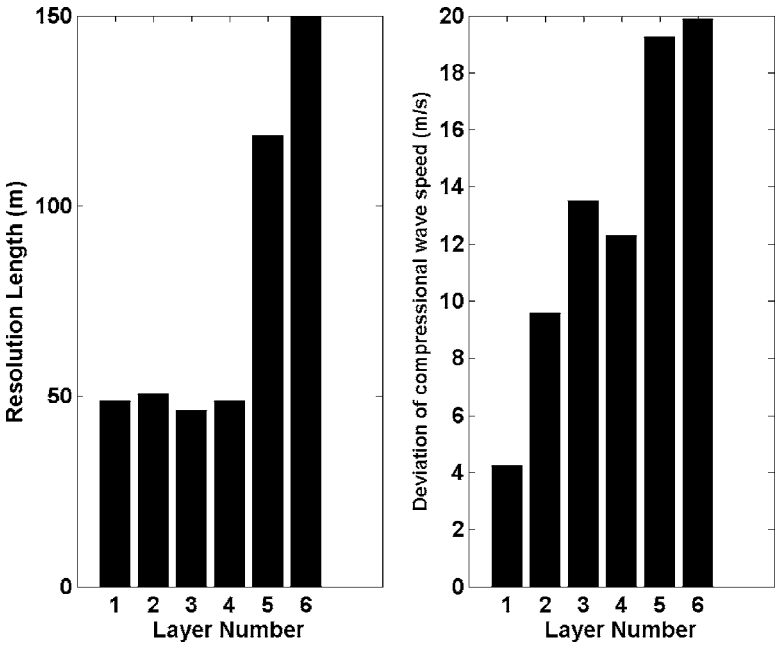


FIG. 7. The panels on the left show the resolution lengths at different layers for one of the three shots. Layer 1 is the water column. Layers 2 to 4 correspond to three sediment layers. Layers 5 and 6 represent the sediment in the half-space. The water column is approximately 37 m thick, while the sediment layers are modeled as 35 m thick. The right panel shows the deviation of the compressional wave speed estimate in each layer.

does not change after inversion, it implies that the inverse has little impact on these model parameters.

The *posteriori* model standard deviation determined from Eq. (20) is shown in the right panels of Fig. 7. It is seen that the *posteriori* value is less than the prior value in Layers 1 to 4. For Layers 5 and 6, there is very little change in the standard deviation. This indicates that the inverse did not have any effect on the deeper layers, a conclusion that is supported by the resolution estimates as well.

V. CONCLUSIONS

The data used in this analysis were broadband signals acquired on a horizontal array during a geophysical survey. Such data are routinely collected by oil companies, and an important thrust of this research is to determine whether such data can be used to determine near-surface geoacoustic properties of the sediment. This paper looks at estimating the sediment acoustic properties from estimates of modal eigenvalues obtained from the broadband data. Sources of errors in the estimation of the modal eigenvalues were considered. Since the water column properties were not known, joint inversion of the water column and sediment acoustic properties was carried out.

The source was an air gun with most of its energy in the frequency band of 15–50Hz. The receiver array consisted of 96 hydrophones with an interelement spacing of 25 m. The data were transformed into the frequency domain, and thus the pressure field, as a function of range at a given frequency, was extracted from the data. We have used these data to extract the horizontal wave numbers of the propagating modes, and attempted to obtain a water column and sediment model using these wave numbers.

The pressure field data obtained from the horizontal array have a spatial sampling of 25 m. At frequencies higher than 30 Hz, the spatial sampling rate is below the Nyquist rate required to avoid aliasing when transforming the pressure field into the wave number spectrum. We demonstrated with synthetic data that interpolation of the data at the required sampling rate does not introduce errors in the estimate of modal wave numbers. We also investigated the effect of errors in the assumed range and receiver depth. Our simulations indicated that the effect of these errors on modal wave number estimates of the dominant low-order modes is negligible. The addition of random Gaussian noise to the simulated broadband data results in considerable changes in both the pressure field obtained from it and in the corresponding wave number spectrum. But the dominant modes seem to be unaffected by noise. The weaker higher-order modes are affected, and spurious modes appear at wave numbers higher than the water wave number. This results in a highly oscillatory pressure-field structure. It is to be noted that the pressure fields obtained from the field experiment also exhibited a similar structure. It was therefore decided to use the wave numbers of dominant modes (Modes 1 and 2) only for the parameter estimation. The wave numbers of dominant modes at 18, 20, 25, and 30 Hz were used, and the joint water column-sediment bottom model was determined from Shot 5, Shot 50, and Shot 95 data.

The estimates of water column sound speed and water column thickness are very close to the values obtained through matched field processing. The estimates of sediment compressional wave speed up, to a sediment depth of 105 m, are within two standard deviations of the mean values calculated using Gibb's sampling approach[2] in the case of models from Shot 50 and Shot 95 data. In the case of Shot 5, the agreement is limited to sediment depth less than 70 m. The penetration of these modes is restricted to the top layers; hence, the poor resolution for the deeper layers.

The comparison between the fields predicted by the models and the measured fields indicates good correlation at the lower frequencies. At 30 Hz, the correlation is poor because only the wave number of Mode 1 at this frequency was incorporated in the data.

The resolution estimates indicate that the top layers are fairly well resolved, and this conclusion is supported by the variance estimates.

Errors in the data impose restrictions on the modal wave numbers that can be used as data in the linear inverse procedure. This results in poor resolution of the deeper layers. However, the estimates of the shallower layers are comparable to the values obtained by matched field processing. This suggests that the modal inverse method can be used to determine the initial model and the search range for a global optimization procedure, thus providing improved efficiency in the execution of matched field processing.

ACKNOWLEDGMENTS

The work was supported by the Naval Research Board, Ministry of Defense, Government of India.

[1]S. D. Rajan, J. F. Lynch, and G. V. Frisk, "Perturbative inverse methods for obtaining bottom geoacoustic parameters in shallow water," J. Acoust. Soc. Am. **82**, 998–1017 (1987).

[2]P. V. Nagesh, S. D. Rajan, and G. V. Anand, "Estimation of sediment acoustic properties from horizontal array data," *Conf. on Underwater Acoustic Measurements: Technologies and Results*, Heraklion, Crete, Greece, (2005).

[3]J. F. Lynch, S. D. Rajan, and G. V. Frisk, "A comparison of broadband and narrowband inversions for bottom geoacoustic properties at a site near Corpus Christi, Texas," J. Acoust. Soc. Am. **89**, 648–651 (1991).

[4]G. V. Frisk, J. F. Lynch, and S. D. Rajan, "Determination of wave speed profiles using modal inverse techniques in a range-dependent environment in Nantucket Sound," J. Acoust. Soc. Am. **86**, 1928–1939 (1989).

[5]G. V. Frisk and J. F. Lynch, "Shallow water characterization using Hankel transform," J. Acoust. Soc. Am. **76**, 205–211 (1980).

[6]H. Schmidt, *Oases version 2.2, User Guide and Reference Manual* (Massachusetts Institute of Technology, Cambridge, MA, 1999).

[7]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean acoustics* (Springer, New York, 2000).

[8]K. E. Hawker, "A normal mode theory of acoustic Doppler effects in the ocean waveguides," J. Acoust. Soc. Am. **56**, 58–64 (1974).

[9]H. Schmidt and W. A. Kuperman, "Spectral and modal representations of the Doppler-shifted field in ocean waveguides," J. Acoust. Soc. Am., **96**, 386–395 (1994).

[10]K. M. Becker, S. D. Rajan, and G. V. Frisk, "Results from geoacoustic inversion techniques workshop using modal inverse method," IEEE J. Ocean. Eng. **28**, 331–341 (2003).

[11]S. D. Rajan and S. Bhatta, "Evaluation of high-resolution methods for estimating modal modal wave numbers in a shallow water environment," J. Acoust. Soc. Am. **93**, 378–389 (1993).

[12]S. D. Rajan, "Simultaneous reconstruction of compressional wave speed and density profiles from modal eigenvalues," J. Comput. Acoust. **6**, 257–268 (1998).
[13]S. D. Rajan, G. V. Frisk, and J. F. Lynch, "On the determination of modal attenuation coefficients and compressional wave attenuation profiles in a range-dependent environment in Nantucket sound," IEEE J. Ocean. Eng. **17**, 118–126 (1992).
[14]R. T. Kessel, "The variation of modal wave numbers with geoacoustic parameters of a layered media," J. Acoust. Soc. Am. **102**, 2690–2696 (1997).
[15]S. D. Rajan, "Waveform inversion to obtain the geoacoustic properties of sediments in deep water," J. Acoust. Soc. Am. **91**, 3228–3241 (1992).
[16]E. L. Hamilton, "Geoacoustic modeling of the seafloor," J. Acoust. Soc. Am. **68**, 1313–1340 (1980).

Haro Strait geometry (sloping bottom)

A. Tolstoy
ATolstoy Sciences, 1538 Hampton Hill Circle, McLean, Virginia 22101

(Received 2 February 2006; revised 11 May 2006; accepted 12 May 2006)

This paper is a followup to a previous JASA paper which assumed a flat bottom. New results include (i) estimations assuming a *sloping* bottom where, as before, the *geometric parameters* of various examined Haro St. scenarios are estimated based on the time domain signals *alone*. Additionally, this new paper considers (ii) *four* parts of the time domain signal (rather than the previous 2): The time differences for each phone for (1) surface reflected minus direct arrivals; (2) bottom reflected minus direct arrivals; *plus* (3) surface then bottom reflected minus direct arrivals; *plus* (4) bottom then surface reflected minus direct arrivals. Finally, (iii) 12 to 16 phones (rather than the previous eight) may now be considered as well (up to 16 phones are available for some source-array paths). The determined geometric parameters remain: source range, source depth, individual phone ranges, individual phone depths, and now include two water depths (at the source and array) per scenario. We have retained constraints to again improve realism, e.g., the array shape is restricted, sources can only be within certain ranges and depths, the top array phone depth and water depths can only be within certain regimes. Finally, as before, some of the new optimal positions are "confirmed" by simulating the signals for the predicted geometries by means of a pulse (PE) (RAMGEO) propagation code [M. D. Collins, "Generalization of the split-step Pade solution," J. Acoust. Soc. Am. **96**, 382–385 (1994)] and then comparing those simulated signals with the observed data. Unfortunately, based on geometric concerns alone, even with these new considerations, non-uniqueness of any solution geometry is still *not* guaranteed.
[DOI: 10.1121/1.2214174]

PACS number(s): 43.30.Pc [EJS] Pages: 1335–1346

I. INTRODUCTION

The Haro Strait experiment was conducted south of Vancouver Island in June 1996 and was designed to examine broadband (200–800 Hz) signals from numerous distributed sources on three (nearly) vertical arrays in a range-dependent shallow water region. The data have been studied by a number of researchers and are considered to be extremely difficult to invert for geoacoustic properties as a result of strong local currents and complicated topography. One of the major difficulties resulting from the strong currents is more uncertainty in the geometric parameters of each data set, i.e., the ranges and depths of each source, water depths at the sources and arrays, and the phone locations of each array at any chosen time. The motivation here is to "preprocess" the data in order to estimate geometric parameters prior to a later estimation (still to be performed) of extensive environmental parameters. The test itself is more fully described in Tolstoy,[1] Corré and Chapman,[2] Jaschke,[3] Jaschke and Chapman,[4] and Pignot and Chapman[5] with helpful figures also shown in those papers. We will again analyze only selected shots (14 and 15) to the (NW) array. Arrival time data from these data are used to produce improved estimates of the source and NW array phone locations as well as the endpoint water depths.

The structure of this paper is as follows: Section II provides a brief presentation of the new geometric relations (assuming a sloping bottom) for the first five arrivals (four reflections) used in the time domain inversions for source range and depth. Section III shows the effects of the additional information on uniqueness for the geometric parameters including additional receiver phones. Section IV presents "optimized"[6] results for the selected paths. Section V discusses a new optimization of the array geometry. Section VI discusses arrival matching and sensitivities in detail, including pulse simulation results via RAMGEO comparing them with the data for the selected paths. Section VII summarizes the work and presents conclusions.

II. GEOMETRIC RELATIONS

There is an obvious and simple relationship between the time (T_d) of a direct arrival and a time (T_S) for the first surface reflected arrival for a source in a *constant* sound-speed ocean when modeled by ray acoustics. For this arrival (at angle ϕ) bottom slope (given by γ) does *not* matter, and the result is shown in Fig. 1 [and in Fig. 2(a) of Ref. 1]. In particular, we see that (as before) for c_0 as the (assumed) constant sound speed of the ocean[7] estimated times are given by

$$\hat{T}_d = d/c_0,$$

$$\hat{T}_S = (d_1 + d_2)/c_0,$$

where rge is the range from the source to the phone, zsou is the source depth, and zph is the depth of the phone so that

$$d^2 = \mathrm{rge}^2 + (\mathrm{zsou} - \mathrm{zph})^2,$$

$$(d_1 + d_2)^2 = \mathrm{rge}^2 + (\mathrm{zsou} + \mathrm{zph})^2.$$

Thus, we can easily compute an estimated time difference

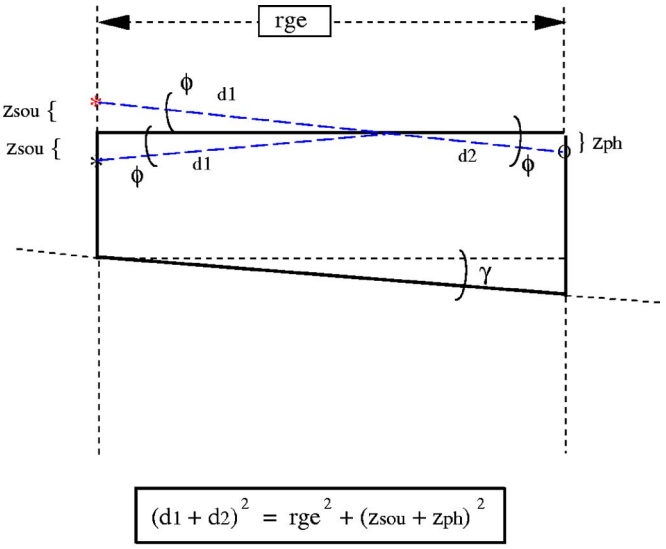


FIG. 1. (Color online) Geometric relationship for the first surface reflected arrival distance (d_1+d_2).

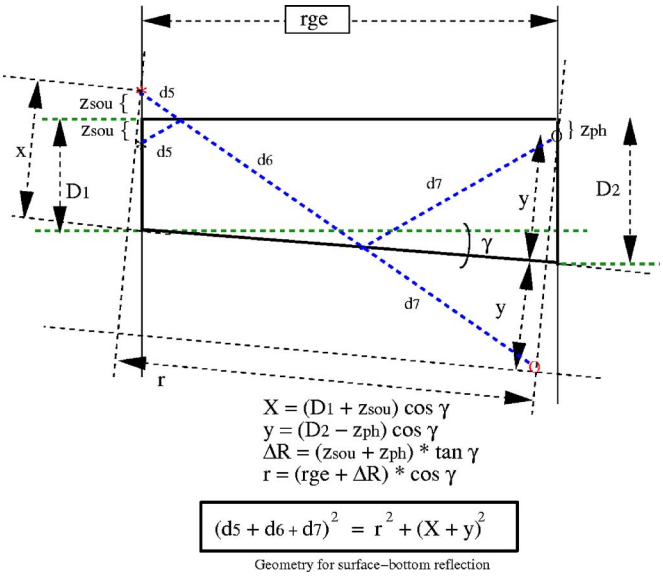


FIG. 3. (Color online) Geometric relationship for the first surface-bottom reflected arrival at distance ($d_5+d_6+d_7$) assuming a sloping bottom at angle γ.

$$\hat{\Delta}_S = \hat{T}_S - \hat{T}_d$$

for a given source and phone position. This difference can be compared to the measured data difference ($\Delta_S=T_S-T_d$) to optimize the estimated phone and source positions, i.e., we can minimize $|\hat{\Delta}_S-\Delta_S|$ as a function of zsou, zph, and rge. See also Michalopoulou and Ma.[8]

Next, we can compute the estimated time ($\hat{T}_B$) of a first bottom reflected arrival (making angle ψ with the bottom as seen in Fig. 2) for a *sloping* bottom, then compute the time difference $\hat{\Delta}_B=\hat{T}_B-\hat{T}_d$ for a variety of bottom depths D_1, D_2 (or equivalently D_1 and slopes γ), and finally compare the estimated time differences with that observed in the data ($\Delta_B=T_B-T_d$) to find the values of zsou, zph, rge, D_1, D_2 which minimize $|\hat{\Delta}_B-\Delta_B|$. For $D_2 \geq D_1$ we have $\gamma = \tan^{-1}(D_2-D_1)/\text{rge}$. In general we have

$$\hat{T}_B = (d_3 + d_4)/c_0,$$

$$(d_3 + d_4)^2 = r^2 + (x + y)^2,$$

$$x = (D_1 - z_{\text{sou}}) * \cos\ \gamma,$$

$$y = (D_2 - z_{\text{ph}}) * \cos\ \gamma,$$

$$\Delta r = (z_{\text{sou}} - z_{\text{ph}}) * \tan\ \gamma,$$

$$r = (\text{rge} - \Delta r) * \cos\ \gamma,$$

$$\hat{\Delta}_B = \hat{T}_B - \hat{T}_d.$$

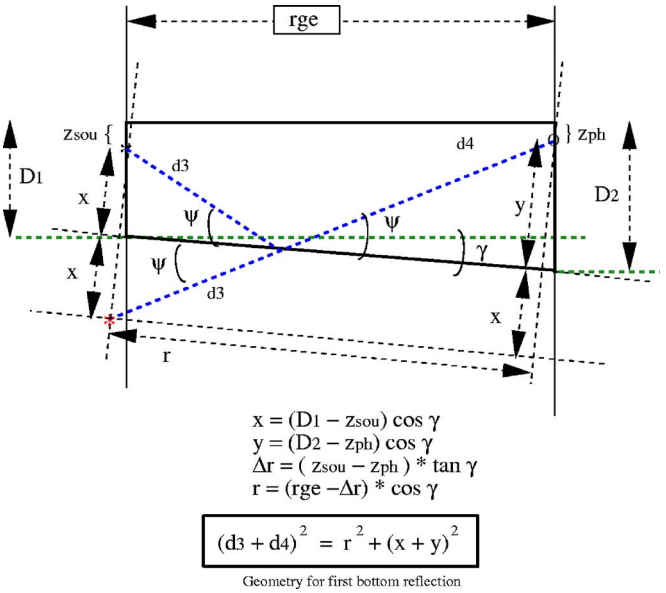


FIG. 2. (Color online) Geometric relationship for the first bottom reflected arrival at distance (d_3+d_4) assuming a sloping bottom at angle γ. We note that for $D_2 \geq D_1$ we have $\gamma=\tan^{-1}(D_2-D_1)/\text{rge}$.

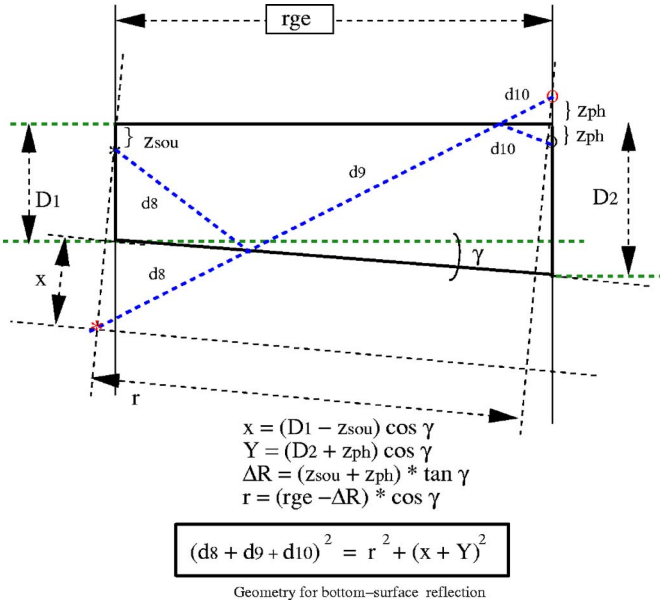


FIG. 4. (Color online) Geometric relationship for the first bottom-surface reflected arrival at distance ($d_8+d_9+d_{10}$) assuming a sloping bottom at angle γ.

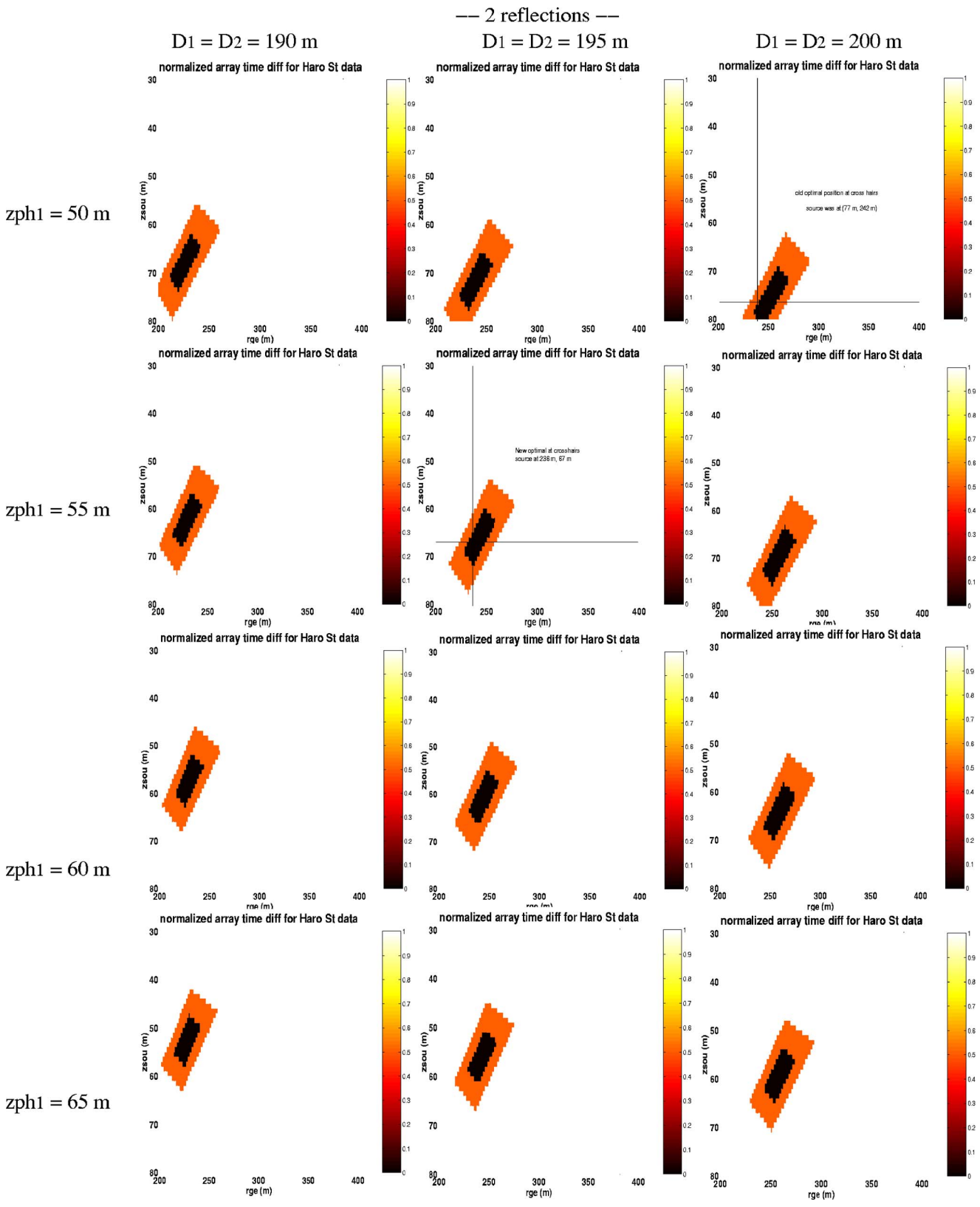


FIG. 5. (Color online) Plot showing the "excellent" (black) source positions assuming three flat bottoms (D=190,195,200 m) and two reflections (one from the surface, one from the bottom) for four candidate depths for the top (first) phone. These phone depths as shown are 50, 55, 60, and 65 m. These plots were generated by comparing two *data* time differences for nw014 with two *estimated* time differences based on the geometry. We see that an infinite number of fits are possible. We note that as the water depth becomes deeper, the likely depth of the phone becomes deeper and farther from the source. All axes on subfigures are identical: x axis is rge from 200 to 400 m in 50 m increments, y axis is zsou from 30 to 80 m (downward) in 10 m increments.

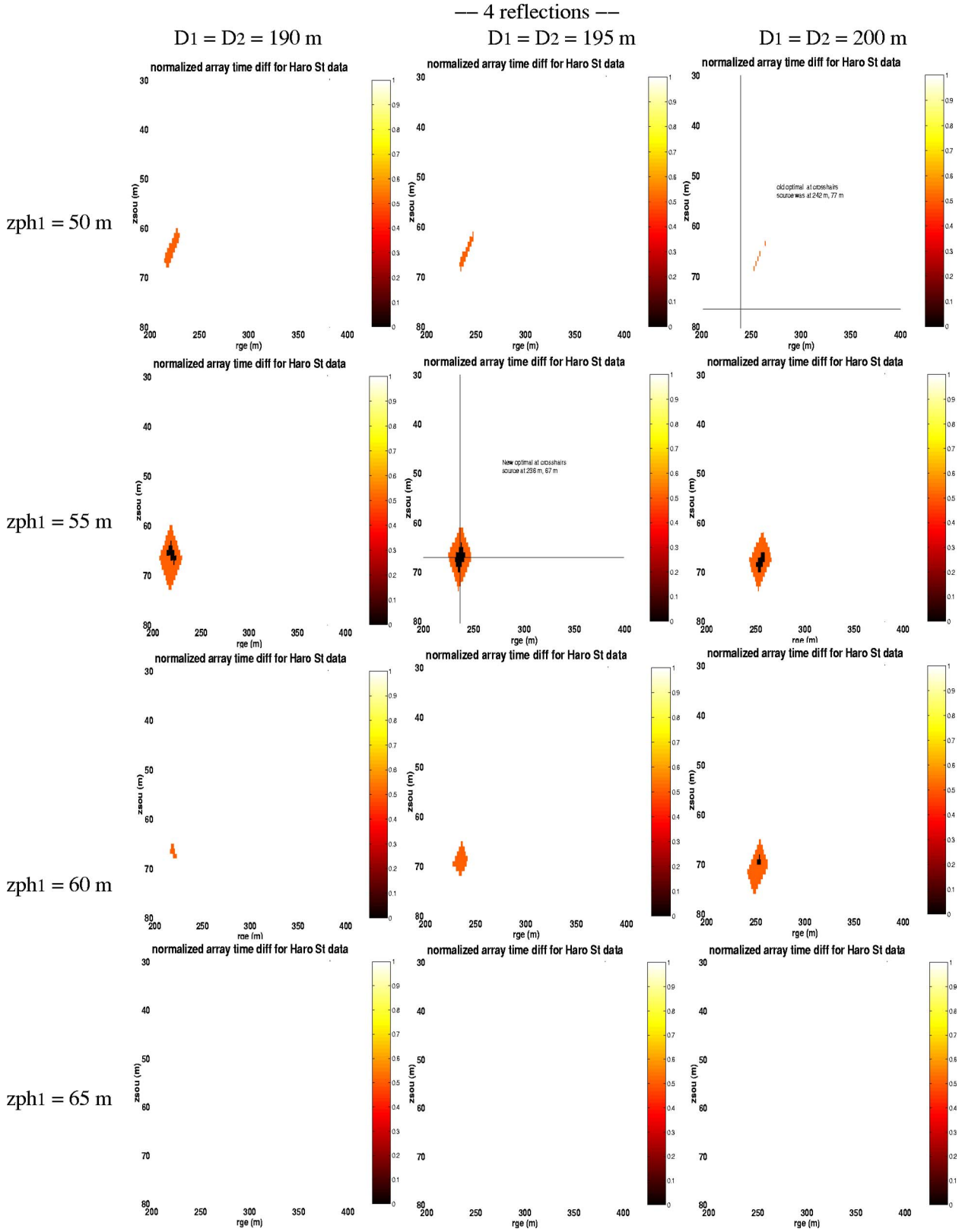


FIG. 6. (Color online) This plot is similar to the preceding figure (nw014) but now allows for four reflections: one from the surface, one from the bottom, one from the surface bottom, one from the bottom surface. We see that the number of "excellent" (black) fits to the data has been significantly reduced. In particular, very shallow arrays, i.e., top phone above 55 m, do not fit these data (nw014) nor do very deep arrays, i.e., top phone below 60 m.

For the third difference, we can compute the estimated time ($\hat{T}_{\mathrm{SB}}$) of a first surface-bottom reflected arrival for a sloping bottom (as seen in Fig. 3), then compute the time difference $\hat{\Delta}_{\mathrm{SB}}=\hat{T}_{\mathrm{SB}}-\hat{T}_d$ for a variety of bottom depths D_1, D_2, and finally compare the estimated time differences with that observed in the data ($\Delta_{\mathrm{SB}}=T_{\mathrm{SB}}-T_d$) to find the values of zsou, zph, rge, D_1, D_2 which minimize $|\hat{\Delta}_{\mathrm{SB}}-\Delta_{\mathrm{SB}}|$. We have

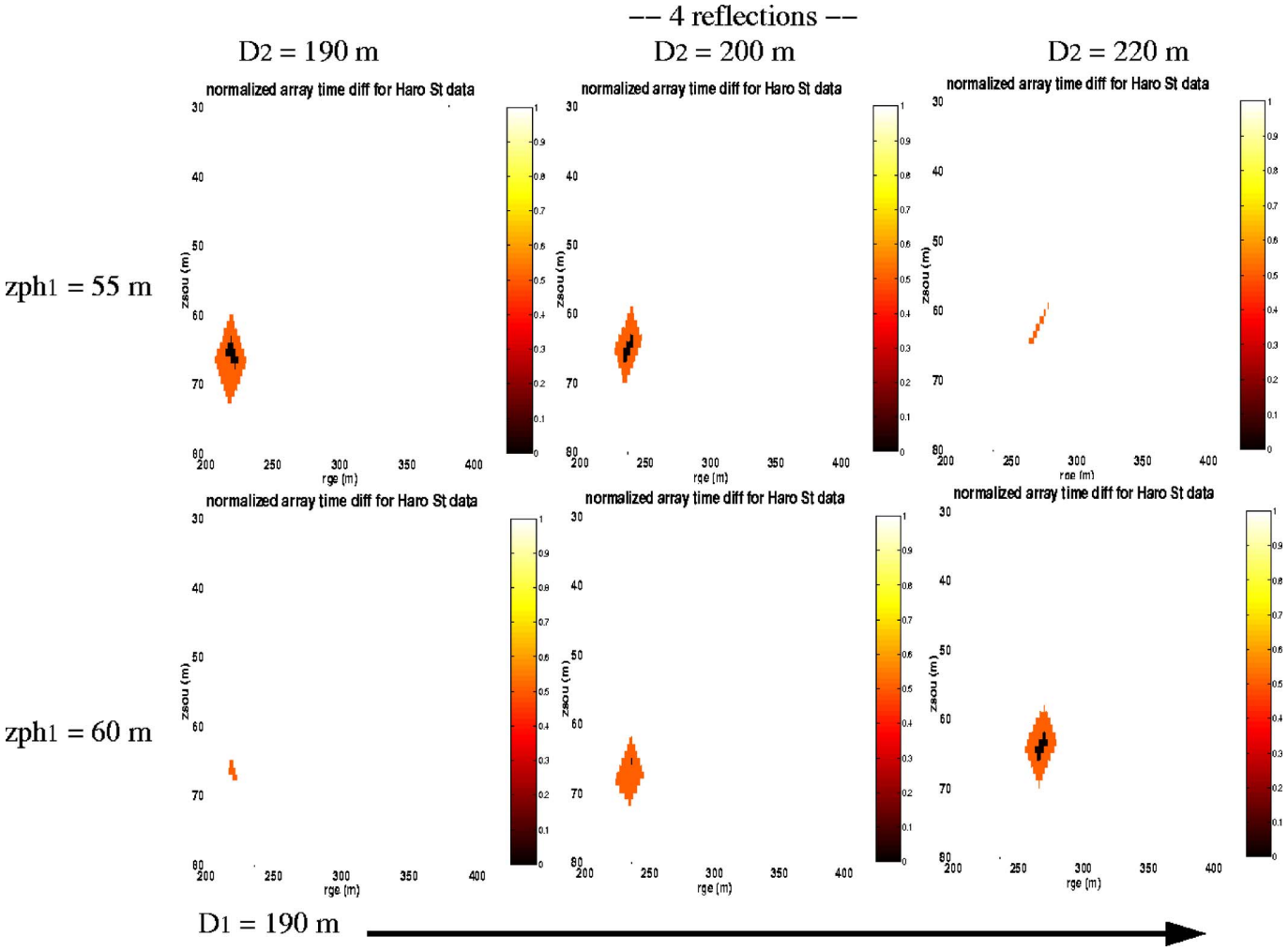


FIG. 7. (Color online) This plot is similar to the preceding one but now allows for a sloping bottom. In particular, $D_1=190$ while $D_2=190,200,220$ m. We see that there is some change from the previous figure but not a lot.

$$\hat{T}_{\mathrm{SB}} = (d_5 + d_6 + d_7)/c_0,$$
$$(d_5 + d_6 + d_7)^2 = r^2 + (X + y)^2,$$
$$X = (D_1 + z_{\mathrm{sou}}) * \cos\,\gamma,$$
$$y = (D_2 - z_{\mathrm{ph}}) * \cos\,\gamma,$$
$$\Delta R = (z_{\mathrm{sou}} + z_{\mathrm{ph}}) * \tan\,\gamma,$$
$$r = (\mathrm{rge} - \Delta R) * \cos\,\gamma,$$
$$\hat{\Delta}_{\mathrm{SB}} = \hat{T}_{\mathrm{SB}} - \hat{T}_d.$$

For the fourth difference, we can compute the estimated time ($\hat{T}_{\mathrm{BS}}$) of a first bottom-surface reflected arrival for a sloping bottom (as seen in Fig. 4), then compute the time difference $\hat{\Delta}_{\mathrm{BS}}=\hat{T}_{\mathrm{BS}}-\hat{T}_d$ for a variety of bottom depths D_1, D_2, and finally compare the estimated time differences with that observed in the data ($\Delta_{\mathrm{BS}}=T_{\mathrm{BS}}-T_d$) to find the values of zsou, zph, rge, D_1, D_2 which minimize $|\hat{\Delta}_{\mathrm{BS}}-\Delta_{\mathrm{BS}}|$. We have

$$\hat{T}_{\mathrm{BS}} = (d_8 + d_9 + d_{10})/c_0,$$
$$(d_8 + d_9 + d_{10})^2 = r^2 + (x + Y)^2,$$
$$x = (D_1 - z_{\mathrm{sou}}) * \cos\,\gamma,$$
$$Y = (D_2 + z_{\mathrm{ph}}) * \cos\,\gamma,$$
$$\Delta R = (z_{\mathrm{sou}} + z_{\mathrm{ph}}) * \tan\,\gamma,$$
$$r = (\mathrm{rge} - \Delta R) * \cos\,\gamma,$$
$$\hat{\Delta}_{\mathrm{BS}} = \hat{T}_{\mathrm{BS}} - \hat{T}_d.$$

Next, let us examine how a sloping bottom, more surface and bottom reflections, and additional phones help with regard to the nonuniqueness problem.

III. NONUNIQUENESS

A very important addition to the calculations seems to be the higher-order reflections, i.e., the surface-bottom and bottom-surface reflections. First, let us revisit the flat bottom case for nw014.

Before beginning, the axes of the figure parts of Figs. 5–10 have very small labels. They are all the same: The x axis runs from rge of 200 m (left) to 400 m in 50 m increments while the y axis runs from zsou of 30 m (top) to 80 m (bottom) in 10 m increments. Since Figs. 5 and 6 contain 12 figures parts, the axes are difficult to enlarge. The computa-

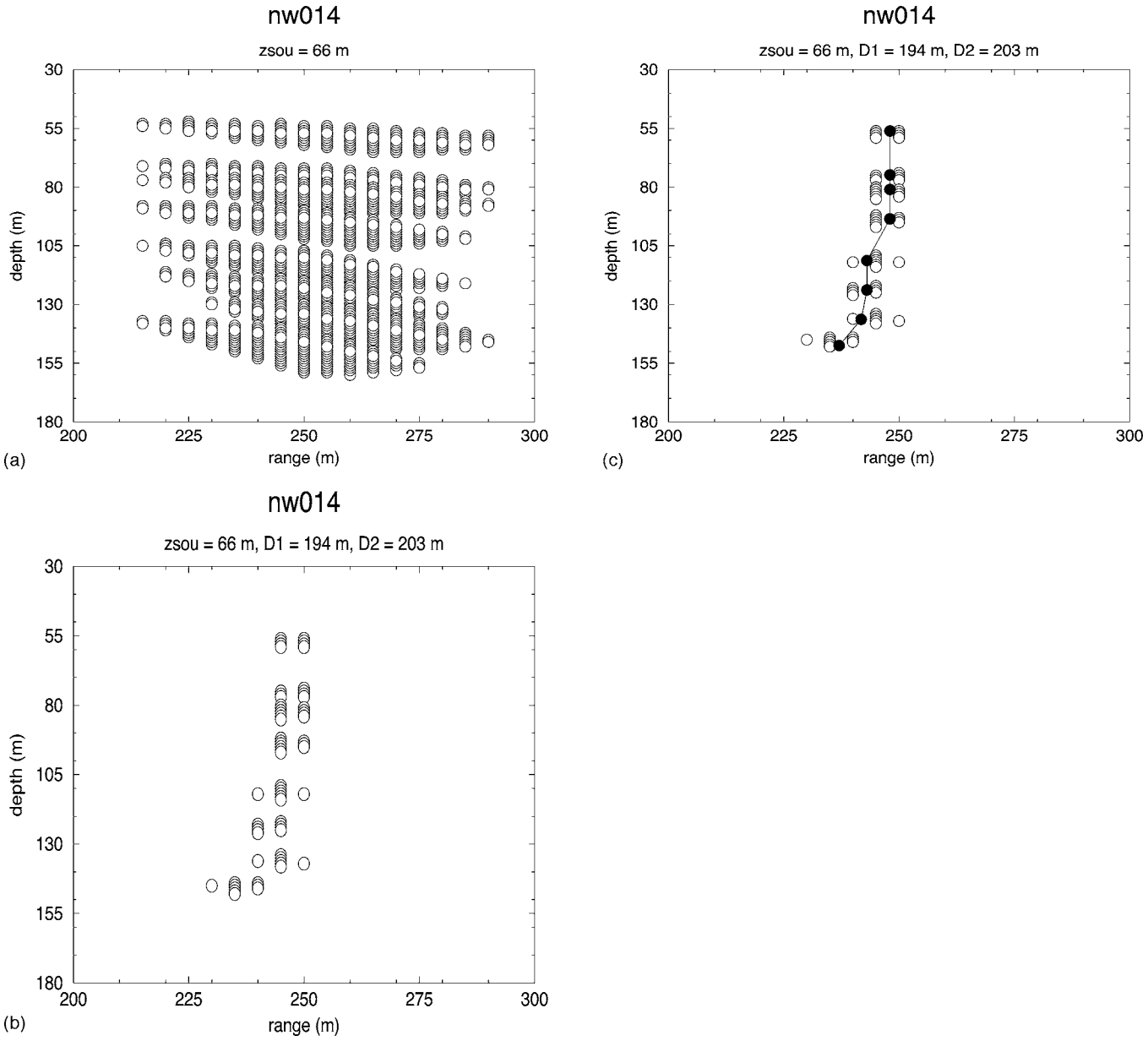


FIG. 8. (Color online) Plots showing the candidate ranges and depths of selected phones (1,4,5,7,9,11,13,15) for NW014. For each circle the phone location results in time differences for all four boundary reflections versus the data to be less than 0.002 s. In (a) we see all possible locations for phones given that the source depth must be 66 m. We note obvious clustering of the phones. In (b) we constrain the water depths so that $D_1=194$ m, $D_2=203$ m. In (c) we impose the black filled-in circles shown the array of Table I.

tions are simply the individual differences (black for each less than 0.002 s, gray (orange online) for some values larger than 0.002 s but each still less than 0.004 s) between the ray arrival times (computed by the formulas of Sec. II) and the data arrival times (computed by examining the data) for the indicated number of boundary reflections, e.g., 2 or 4.

In Fig. 5 we show the top phone depths(zph_1) and the flat bottom ($D=D_1=D_2$) depths which allow for excellent agreement with the time domain data for two reflections only. In particular, the black regions of the figure show excellent agreement (barely discernible differences) between the data and geometric arrivals. That is, the black regions correspond to source ranges and depths for which the indicated top phone depth and water depth D produce geometric arrivals which agree to within 0.002 s for time differences in *both* the surface reflected minus direct arrivals and in the bottom reflected minus direct arrivals. The white regions show disagreements larger than 0.004 s (in at least one arrival), while the gray regions show disagreements larger than 0.002 s but less than 0.004 s (discernible but slight differences with the data). There are clearly many top phone depths, water depths, source ranges, and source depths which show excellent agreement with the data (black regions). In the original paper we concluded that the first phone for nw014 was at depth 49.75 m, the average water depth was 199 m, the source at range 242 m, depth 77 m. This position is approximately indicated by the crosshairs in the upper right of Fig. 5.

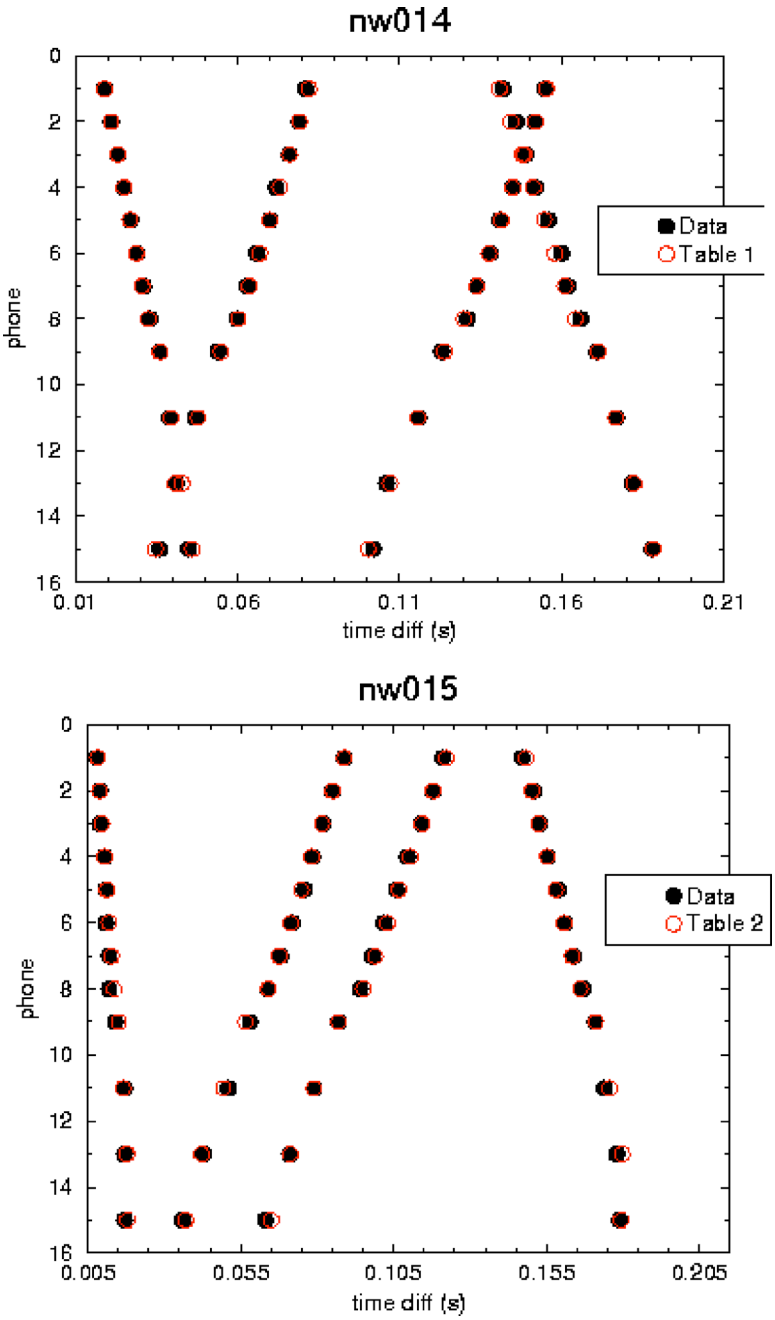


FIG. 9. (Color online) The upper plot shows the nw014 *data* versus the simulated (predicted) arrivals using the parameters of Table I. The lower plot shows the nw015 *data* versus the simulated (predicted) arrivals using the parameters of Table II.

In Fig. 6 we allow for four reflections, i.e., we have added the surface-bottom and bottom-surface reflections (equations above). The allowable configurations are now a subset of those of Fig. 5, and we have clearly improved the nonuniqueness of the possible solution space, i.e., reduced the sizes of the black regions. The new optimal position for the shown configurations is indicated by the crosshairs of the middle of Fig. 6, second from the top with the source at 236 m, 67 m. The overall optimal position for a flat bottom occurs at $\mathrm{zph}_1=56.50$ m, $D=200$ m, $z_{\mathrm{sou}}=68$ m, rge $=253$ m. Using four reflections we now conclude that:

- The use of four reflections significantly improves the localization of the source and array (compared to the use of only two reflections).
- The top phone for nw014 cannot be shallower than 50 m.
- The top phone for nw014 cannot be deeper than 65 m.
- We have *new* optimal array and source configurations using four reflections. In particular, the nw014 array seems to be deeper (top phone at 55 versus 49.75 m) and the source shallower (67 versus 77 m) than before.

Next, let us consider the effect of a sloping versus flat bottom. In Fig. 7 we see variations like those of Fig. 6 (using four reflections) but with the sloping bottom. We see that increasing D_2 from 190 to 220 m (moving to the right of the figure) shifts the source location as well as changes the fit to the data. In particular, increasing the water depth at the array (D_2) results in shallower source depths and larger source ranges to fit the data. Moreover, the deeper top phone, e.g., at 60 m, results in an excellent fit for $D_1=190$ m, $D_2=220$ m (rge$=255$ m, $z_{\mathrm{sou}}=63$ m), whereas for the flat bottom

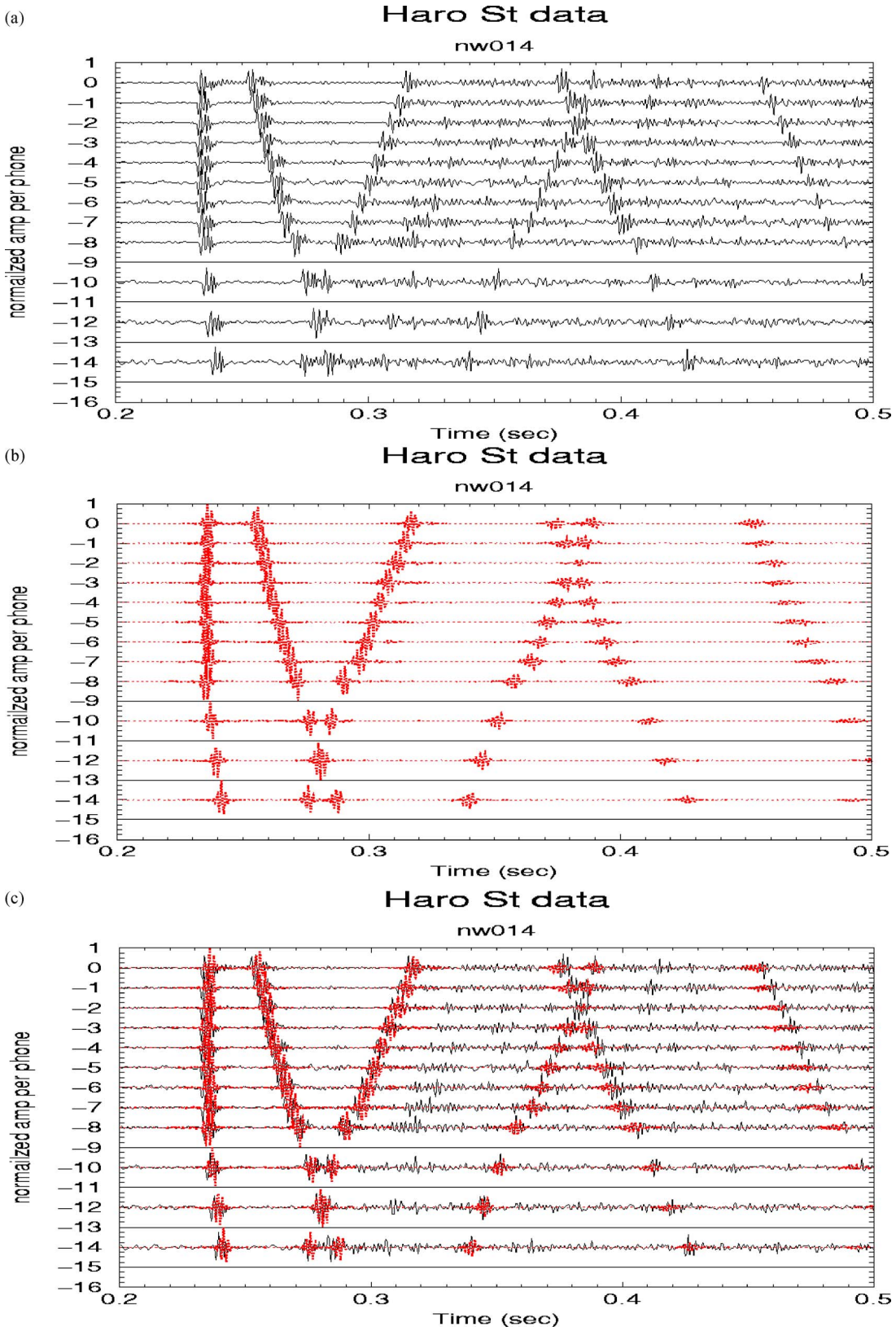


FIG. 10. (Color online) The top (a) shows the data for nw014. The middle (b) shows the simulated data using the parameters of Table I as input to RAMGEO. The bottom (c) shows the overlay of the two time domain arrivals at all the phones. We note that the final (fifth) reflection (sixth arrival) for the simulated data (b) appears to come in a bit early. This third-order reflection was not considered in the calculations here but may be in future work.

D=190 m, there was no excellent (black) fit for the top phone at 60 m to the data.

We can conclude that while the nonflat bottom does not have as pronounced an effect on improving resolution as the addition of more reflections, it does result in shifts for the optimal phone depth and source location. Where the optimal top phone depth was approximately 55 m for the flat bottom D=200 m, it becomes more like 60 m for a bottom depth of

190 m at the source, 220 m at the array. Moreover, for this phone the source has gone from (253, 68) to (265, 64). Thus, the nonflat nature of the bottom can significantly influence the geometry of the signal arrivals.

Finally, let us consider the effect of more phones, i.e., arrays. First, we optimize over all (realistic) arrays, source locations, and sloping bottoms using four reflections and all available phones (for nw014 phones 10, 12, 14, and 16 were flooded and not usable). We find that one excellent fit to the data is possible for a vertical array with top phone at 61.25 m and the source at 258 m range, 64 m depth. Consider the uniqueness of the source location using only the top eight of those phones, and then using only the top phone. We note that using only the top phone resulted in excellent fits to a number of source locations; using the next eight phones improved the possible uniqueness of the source location, while using all the available data (12 phones) resulted in a (nearly) unique best fit to the data (note: we have eliminated the third and fourth reflections for the second and third phones since they were ambiguous and overlapped in the data). Clearly and not surprisingly, more phones significantly improve the resolution of the source localization.

Unfortunately, one excellent fit can be misleading. We find that other geometric configurations can also provide excellent, high resolution fits to the data. We can also have a nonvertical array (slight tilt for the bottom three phones) with top phone at 54.0 m depth and source at 230 m range, 64 m depth, $D_1=190$ m, $D_2=195$ m which *also* provides an excellent fit. Thus, we still do not have final uniqueness even using a great deal of the time domain data.

IV. HARO STRAIT DATA—FURTHER RESULTS

While this method has not resulted in unique solutions, it has significantly *reduced* the geometric possibilities. For nw014 the source is likely to be located in the range 230–260 m, at depths 64–68 m, with D_1 (the water depth at the shot) in $[190,196]$ m, D_2 (the water depth at the array) in $[200,206]$ m. Moreover, for this data the array is likely to be at least as deep as 53.5 m but not deeper than 62.5 m. If we constrain the water depths as per the experimental expectations,[9] i.e., $D_1 \approx 193$ m, $D_2 \approx 202.5$ m, and we look for the shallowest array fitting the data, then we arrive at the geometry of Table I.

Unfortunately, this array seems to be predicted to be deeper than expected (zph_1 was expected around 30 m). Thus, we decided to next examine another shot to that NW array, i.e., shot 15. The shallowest top phone depth found for nw015 was 54 m. This is a very similar minimum depth to that found for nw014. If we restrict the depth D_2 to again be in $[200,206]$ m, then we arrive at the values of Table II which give excellent fits to the data. Thus, the array *at that time* is likely to have had a top phone at $[56,60]$ m. We do note that the source for shot 15 at 32 m depth was shallow (and resulted in weaker reflections from the bottom).

TABLE I. Table shows one optimal source depth (zsou), phone locations (zph, rge), and water depths D_1, D_2 given Haro Strait data for a selected path (nw014). This shows the shallowest array with restrictions on the water depths (within experimental expectations).

zsou=66.0 m in [30,80]	$D_1=194.0$ m, $D_2=203.0$ m in [190,220]
Phone depth (m)	Phone range (m)
$zph_1=56.00$ in [45,85]	$rge_1=248.00$ in [200,400]
$zph_2=62.25$	$rge_2=248.00$
$zph_3=68.50$	$rge_3=248.00$
$zph_4=74.75$	$rge_4=248.00$
$zph_5=81.00$	$rge_5=248.00$
$zph_6=87.25$	$rge_6=248.00$
$zph_7=93.50$	$rge_7=248.00$
$zph_8=99.75$	$rge_8=248.00$
$zph_9=111.25$	$rge_9=243.00$
zph_{10}	
$zph_{11}=123.75$	$rge_{11}=243.00$
zph_{12}	
$zph_{13}=136.25$	$rge_{13}=241.75$
zph_{14}	
$zph_{15}=147.50$	$rge_{15}=237.00$
zph_{16}	

We conclude that the top phone depth for the NW array for the cluster of shots including numbers 14 and 15 is likely to truly be around 56–60 m.

We note that the arrays of Tables I and II suggest *opposite* tilts for the same array with data sets only slightly separated in time. This is a concern. One can interpret this discrepancy to mean either that a diversity of shapes can match the data (we can certainly expect this) or that there is more to the puzzle, e.g., the bottom is *NOT* a simple slope. Efforts to find a single array or one with similar tilt which can fit *both* data sets—if the bottom must be a simple slope—have *not*

TABLE II. Table shows one optimal source depth (zsou), phone locations (zph, rge), and water depths D_1, D_2 given Haro Strait data for a selected path (nw015). This shows the shallowest array given restrictions on the water depths. Note the shallow source depth of 32 m.

zsou=32.0 m in [30,80]	$D_1=196.0$ m, $D_2=204.0$ m in [190,220]
Phone depth (m)	Phone range (m)
$zph_1=58.00$ in [30,70]	$rge_1=296.00$ in [200,400]
$zph_2=63.75$	$rge_2=299.00$
$zph_3=69.50$	$rge_3=301.00$
$zph_4=75.75$	$rge_4=301.00$
$zph_5=82.00$	$rge_5=301.00$
$zph_6=88.25$	$rge_6=301.00$
$zph_7=94.50$	$rge_7=301.00$
$zph_8=100.75$	$rge_8=301.00$
$zph_9=113.25$	$rge_9=301.00$
zph_{10}	
$zph_{11}=125.50$	$rge_{11}=301.00$
zph_{12}	
$zph_{13}=137.00$	$rge_{13}=301.00$
zph_{14}	
$zph_{15}=144.00$	$rge_{15}=311.00$
zph_{16}	

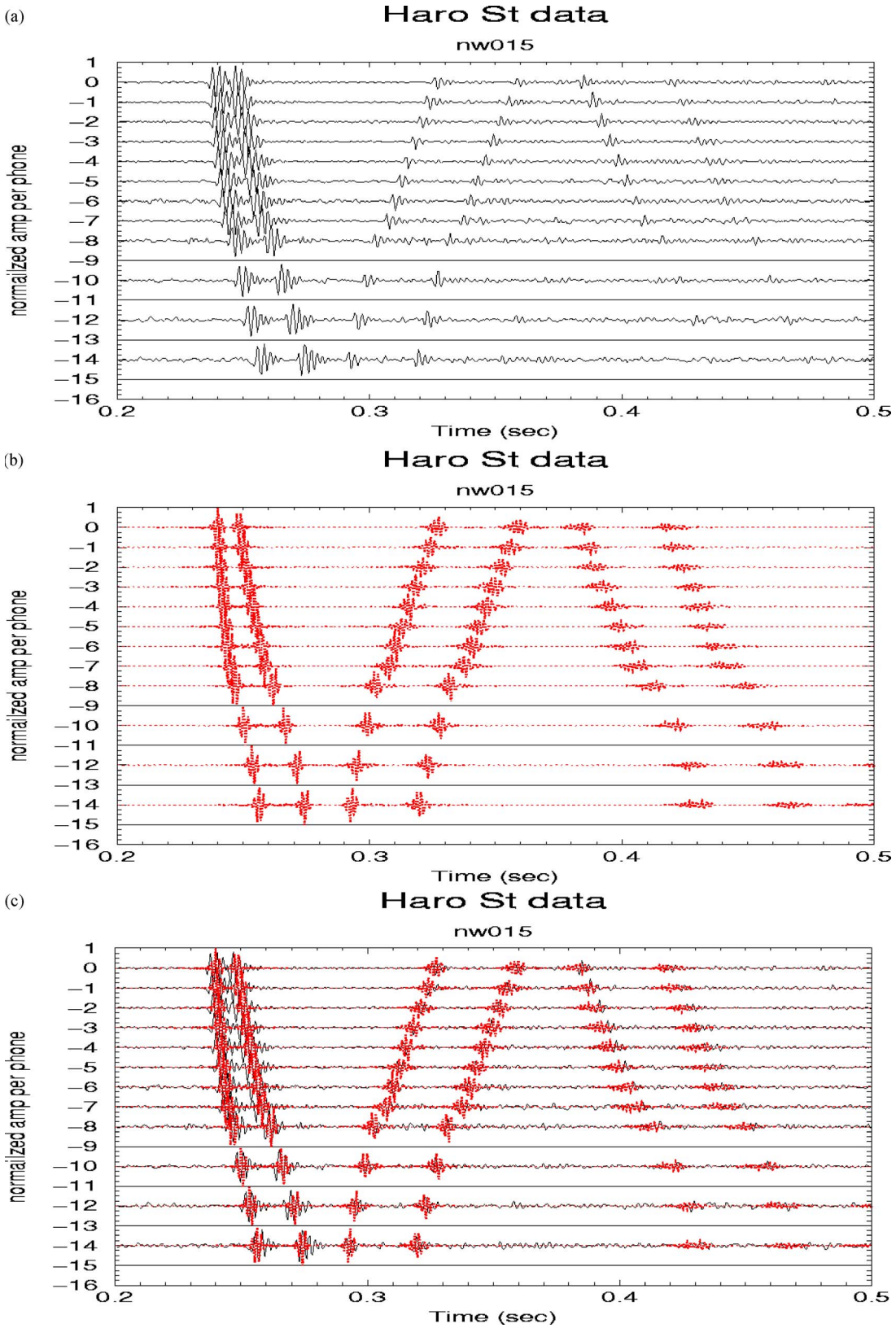


FIG. 11. (Color online) The top (a) shows the data for nw015. The middle (b) shows the simulated data using the parameters of Table II as input to RAMGEO. The bottom (c) shows the overlay of the two time domain arrivals at all the phones. For this calculation the final (fifth) reflection (sixth arrival) for the simulated data (b) appears to come in nearly on time.

been successful. Thus, the second option may be necessary to consider. Chapman[9] has pointed out that shot 15 was very low power, and as such it may not carry much weight with regard to inversion confidence.

One important question remaining is, How were the arrays computed? That is, did they include *all* possible candidates so as not to have missed possibilities including shallower arrays for nw014, nw015?

V. ARRAY GEOMETRY OPTIMIZATION

The array phones were found by an exhaustive search which guaranteed that all possible good-fit arrays could be found. Thus, the calculations of array configurations are believed to contain *all* "excellent" fits to the data. We note that this computation is now different from that of the earlier paper and is discussed in detail next.

We first computed all depths and ranges $(\mathrm{zrec}_1, \mathrm{rge}_1)$ for which the top (first) phone allowed for a fit (all reflections) within 0.002 s to the data for *some* source depth and range $(z_{\mathrm{sou}}, r_{\mathrm{sou}})$ and some D_1, D_2. These possibilities were put into file1. We continued in the same fashion for each remaining phone: phone2, ..., phone15 (excluding 10, 12, 14, and 16 for the NW array). Thus, for the NW array we had 12 files for a given data set each of which allowed for "perfect" fits to the time domain data for each phone.

In Fig. 8(a) we see a representative sampling of the selected phones given some source depth, e.g., zsou=66 m and all considered D_1, D_2. That is, each circle on the plot indicates a range and depth for which the corresponding phone has geometric parameters resulting in less than 0.002 s differences with each of the four data reflections for that phone. In Fig. 8(b) we allow only $D_1=194$ m, $D_2=203$ m. Thus, each circle represents a candidate phone location. This figure shows candidate locations for phones 1, 4, 5, 7, 9, 11, 13, and 15. We note that the candidate phone locations cluster (corresponding to the various files/phones). The next step is to select a single phone from each cluster such that the distances between each are appropriate, and there is no unacceptable snaking of phones.

Thus, eight of a possible 12 data sets[10] were examined together (not in sequence like the first paper but rather in nested loops) to see what phones allowed for appropriate distances $\Delta_{(i,i+1)}$ between neighbors $i, i+1$ where $[\Delta_{i,i+1} = \sqrt{(\mathrm{zrec}_i - \mathrm{zrec}_{i+1})^2 + (\mathrm{rge}_i - \mathrm{rge}_{i+1})^2}]$. For example, $\Delta_{i,i+1} = 6.25$ m for $i=1, \ldots, 7, \Delta_{i,i+1}=12.5$ m for $i=8$, 9, 11, 13. For each array we examined fits for nearly all the odd phones (eight phones from 1–15). We did not use phone 3 of shot 14 since there was more uncertainly about the fourth and fifth arrival times for these data; for nw014 we used phone 4 instead.

In addition, each acceptable phone combination in the set had to have the same water depths at the source, i.e., the same D_1, and the same source depths z_{sou}. Moreover, the water depths D_2 at all of the neighboring array phones had to agree to within 2 m.

Finally, the arrays had to "tilt" only in one direction or the other, i.e., they could not snake in random directions.

In this fashion we could be sure that all possible "reasonable" arrays fitting the data could be found. There were thousands. Thus, we eventually considered only the "shallowest" and "deepest" plus constraints on D_1, D_2 as available. The quotation marks indicate that the solutions were within the search spaces which allowed for variations (errors) of source range of 5 m, source depth of 2 m, water depths of 1 m, and phone depths of 1 m. We, thus, arrived at Tables I and II. The array of Table I is also shown by the black filled circles of Fig. 9. It should be mentioned that examining a figure similar to Fig. 9 but for NW015 suggests that a single array may be found for *both* nw014 and nw015—if bottom depths are not constrained by simple D_1, D_2 endpoints.

VI. ARRIVAL MATCHING (SENSITIVITIES AND DATA)

We can also examine the sensitivity effects on the time differences in the various arrivals of varying source range (rge), depth (z_{sou}), water depth at the source (D_1), water depth at the array (D_2), and array depth (depth of the top phone: zph_1). We examine simulated data only, assuming the general parameters

$$\mathrm{rge} = 250 \text{ m}, \quad z_{\mathrm{sou}} = 70 \text{ m}, \quad D_1 = 185 \text{ m},$$

$$D_2 = 205 \text{ m}, \quad \mathrm{zph}_1 = 45 \text{ m}.$$

These values are close to those found for nw014. The simulated array is a vertical array.

We note a number of behaviors:

- The source range (rge) most strongly influences the third and fourth arrivals.
- Moving the source farther away compresses the arrivals in time.
- z_{sou} changes produce relatively small effects on all arrivals.
- D_1 influences the third arrival most.
- D_2 influences the fourth arrival most.
- Moving the array *up* (reducing zph_1) moves the crossing point for the *X* between the third and fourth arrivals *down*. That is, the crossing point of that *X* occurs at the phone just below the source depth z_{sou}.

These behaviors also hold for longer ranges where the nominal range is approximately 500 m.

We conclude that for the nw014 and nw015 *data* (the filled in circles of Fig. 9) the NW array must be deep at this time (strong, changing currents can change the geometries over time). Using the simulated data (the open circles of Fig. 9) with parameters of Table I for nw014 and with parameters of Table II for nw015 we see the expected excellent agreement with the *measured* data.

Finally, in Figs. 10 and 11 we see simulations of the nw014 and nw015 path parameters, respectively, via RAMGEO (Refs. 11 and 1) for the values of the tables versus the data. We see excellent agreement (observed differences may be the result of geoacoustic parameters which remain to be estimated by other methods). We do note that for those two paths there appears to be an additional, later, reflection which may be worth considering for improved convergence to the optimal values. For the nw014 data this additional feature seems to arrive a bit early in the simulations, suggesting that slight changes to the table values may improve the complete fit to the data.

VII. CONCLUSIONS

We conclude that the use of additional information in the time domain signals improves the overall fit to the data, particularly the fit to the later arrivals corresponding to second-order reflections. This work now allows for a sloping rather

than a flat bottom, more phones, and most importantly two additional boundary interactions. This work also considers a new search for the phone configurations which should account for *all* possible shapes likely for the array. Finally, the results are examined for sensitivities to various geometric values and are successfully compared via RAMGEO to the data. We do note that nonuniqueness is still an issue even in the presence of ever more geometric information.

ACKNOWLEDGMENTS

The author would like to thank ONR for support through Grant No. N00014-06-M-0017 and for direction on the Haro St. data analysis; N. R. Chapman for his continued time and efforts to supply and discuss the data from this unique but difficult test data; and E. Michalopoulou for her very helpful discussions and thoughts.

[1]A. Tolstoy, "Haro Strait geometry (flat bottom)," J. Acoust. Soc. Am. **119**(3), 1388–1395 (2006).
[2]V. Corré and N. R. Chapman, "Vertical-slice matched field tomography," Acoustica **87**, 637–646 (2001).
[3]L. Jaschke, "Geophysical inversion by the freeze bath method with an application to geoacoustic ocean bottom parameter estimation," Ph.D. thesis, U. Victoria, 1997.
[4]L. Jaschke and N. R. Chapman, "Matched field inversion of broadband data using the freeze bath method," J. Acoust. Soc. Am. **106**(4), 1838–1851 (1999).
[5]P. Pignot and N. R. Chapman, "Tomographic inversion in a range-dependent shallow water environment," J. Acoust. Soc. Am. **110**(3), 1338–1348 (2001).
[6]"Optimal" results are taken to mean those for which time domain agreement is within 0.002 s of the data.
[7]c_0 is believed to be essentially constant as a result of a very active mixing environment. Sensitivity to this parameter was investigated in any case and found to be extremely small for the data considered here. See Jaschke,[3] Jaschke and Chapman,[4] and Michalopoulou and Ma.[8].
[8]Z.-H. Michalopoulou and X. Ma, "Source localization in the Haro Strait using arrival time estimation and linearization" (unpublished).
[9]N. R. Chapman (private communication).
[10]We use only a subset of phones because the CPU time becomes enormous as more phones are considered. We do use the full span of the array by selecting a well-separated subset of eight of the 12 phones (1,4,5,7,9,11,13,15) rather than just the first eight of the 12 phones.
[11]M. D. Collins, "Generalization of the split-step Pade solution," J. Acoust. Soc. Am. **96**, 382–385 (1994).

Inverting acoustic communication signals for the sound speed profile

Elin Svensson[a)]
MWL, Department of Aeronautical and Vehicle Engineering, Royal Institute of Technology, SE-100 44 Stockholm, Sweden

(Received 25 January 2006; revised 29 June 2006; accepted 29 June 2006)

Given data from a shallow-water acoustic communication experiment, the acoustic environmental parameters influencing the sound propagation conditions are estimated, particularly the sound speed profile. This inverse problem is solved using a differential evolution algorithm with an objective function measuring the mismatch between observed and modeled impulse response estimates. The model wave field is computed by a hybrid ray-trace/plane-wave method, and its accuracy is assessed in a related range-independent case using solutions obtained with a full-field transform integral method. The environmental parameters obtained by inversion reduce the mismatch between the modeled and the observed impulse response estimates, compared to the mismatch using a sound speed profile measured two days before the experiment. The resulting environmental model is used to investigate alternative source-receiver configurations and to predict the optimal transmitter depth. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2234851]

PACS number(s): 43.30.Pc, 43.30.Cq [AIT] Pages: 1347–1355

I. INTRODUCTION

Acoustic waves in water have low attenuation, which makes them useful for transferring information over long distances. The challenge is the refractive medium and, especially in shallow water, the many boundary interactions. This causes long impulse responses and hence intersymbol interference (ISI) at data communication, which complicates the decoding and interpretation of the received signal.

In order to achieve fast and reliable acoustic communication in shallow water, it is necessary to study different kinds of modulation and equalization schemes as well as the influence of the physical conditions on the transmission of the communication signals. Numerical simulation of acoustic transmission is much cheaper than performing actual experiments and can be used both to analyze the results of an experiment[1] and to decide which experiments to perform.[2]

Realistic simulation of an experiment requires a lot of data. It is necessary to know the bathymetry, the seabed properties, the sound speed profile, the position of source and receiver, and perhaps also how some of these parameters vary with time. Sometimes some of this information is missing and to be able to analyze the measured data, the missing parameters need to be determined. For instance, it is not always possible to measure the sound speed profile exactly at the time the experiments are performed and this can be a major source of error, limiting the agreement between simulation and measurements.

This paper analyzes a communication experiment in 40 m deep water with a source-receiver distance of 3.9 km from a physical point of view. In particular, given acoustic data from the experiment along with the sound speed profile measured two days earlier, the actual sound speed profile during the experiment is estimated. The transmitted data was a quadrature amplitude modulated signal with a carrier frequency of 10 kHz.

This inverse problem is similar to that occurring in *ocean acoustic tomography*, see Ref. 3 for an overview. The inversion approach here is to match estimates of the impulse response obtained from experimental data at a single receiver to model predictions.

The paper is composed as follows. Section II describes the studied communication experiment. In Secs. III and IV the forward propagation model and the optimization method are presented. Section V describes the choice of environmental parameters for the simulation and finally the results are presented and discussed in Secs. VI and VII.

II. EXPERIMENTS

The data were recorded at an underwater acoustics data communication experiment in May 2003 in the Baltic Sea.[4] The source was positioned at a depth of 30 m and a short receiving array with eight elements about 10 cm apart was located at an approximate depth of 36 m. The distance between source and receiver was 3.9 km.

In the right panel of Fig. 1 is the bottom profile, with the source at a range of 0 km and the receiver at a range of 3.9 km. The average depth is around 40 m, with a ridge in the middle and valleys close to the source and the receiver.

Two different sound speed profiles are shown in the left panel of Fig. 1, profile A in black and profile B in gray. They were both measured two days before the communication experiment, but at different times and positions. Both profiles have a well-defined thermocline, however with slightly different depths and slopes. There is hardly any sound speed minimum below the thermocline, but profile A has a local sound speed minimum above the thermocline.

[a)]Electronic mail: elins@kth.se

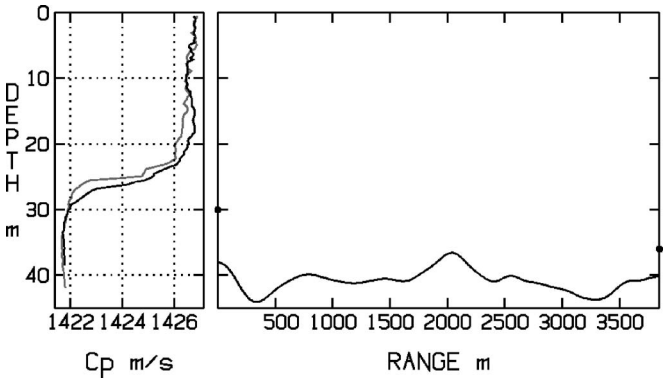


FIG. 1. Source and receiver geometry and measured sound speed profiles for the experiments. Profile A in black and profile B in gray.

The transmitted message was a quadrature amplitude modulated (QAM) signal, where two pseudorandom binary sequences (PRBS), each of a length of 2047, were used to modulate the in-phase and quadrature part, respectively.[5]

In complex baseband this signal can be written

$$s_{\mathrm{bb}}(t) = p_1(t) + ip_2(t), \tag{1}$$

where $p_1(t)$ and $p_2(t)$ are two PRBS signals, which take the value 1 or -1 with the symbol rate of 7350 Hz. The real bandpass version then becomes

$$\begin{aligned} s_{\mathrm{bp}}(t) &= \mathrm{Re}\{(p_1(t) + ip_2(t))e^{i\omega_c t}\} \\ &= p_1(t)\cos\,\omega_c t - p_2(t)\sin\,\omega_c t, \end{aligned} \tag{2}$$

where ω_c is 2π times the carrier frequency f_c, which in this case was 10 000 Hz.

The received signal is the convolution of the transmitted signal and the measured impulse response. In the frequency domain this can be written

$$S_r(f) = H(f)S_{\mathrm{bp}}(f) + N(f) = T(f)C(f)R(f)S_{\mathrm{bp}}(f) + N(f). \tag{3}$$

The total transfer function $H(f)$ consists of three factors, the transmitter transfer function $T(f)$, the channel transfer function $C(f)$, and the receiver transfer function $R(f)$. $N(f)$ is additive noise from the channel.

$T(f)$ and $R(f)$ are unknown characteristics of the transmitter and the receiver. However, assuming they are constant within the frequency range of the signal, it follows that the total transfer function $H(f)$ is proportional to the channel transfer function $C(f)$.

III. FORWARD PROPAGATION MODEL

The forward propagation problem was solved by the hybrid ray-trace/plane-wave method XRAY.[6] This method uses ray tracing in the water column and models the bottom interaction by plane-wave reflection coefficients. The surface is treated as a horizontal plane, with the reflection coefficient $R=-1$. The method is similar to that presented by Hovem and Knobles.[7]

XRAY can handle layered seabeds of fluid or solid materials with range-dependent geometry and material parameters. The layer interfaces are required to be smooth functions and are represented by splines. The material parameters within each layer (water column and seabed layers) are similarly smoothed and represented by variance reducing B-spline expansions.[8] If the sound speed is given in just a few points, these points will be used to construct a smooth function in the entire water column.

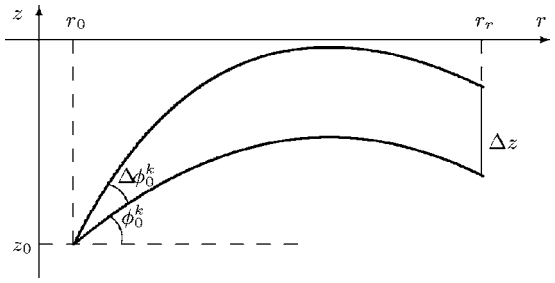


FIG. 2. Ray no. k and the candidate to ray $k+1$.

A ray trajectory $(r(s), z(s))$ where s denotes the arc length, is in the range-independent case $c=c(z)$ a solution to the ordinary differential equation (ODE) system[9]

$$\begin{aligned} \frac{\partial r}{\partial s} &= \cos\,\phi, \\ \frac{\partial z}{\partial s} &= \sin\,\phi, \\ \frac{\partial \phi}{\partial s} &= -\frac{\cos\,\phi}{c}c'(z), \end{aligned} \tag{4}$$

with the initial conditions $r=r_0$, $z=z_0$, $\phi=\phi_0$ at $s=0$. Here, ϕ is the grazing angle, ϕ_0 the launch angle, and (r_0, z_0) the position of the source. The horizontal coordinate is r and the vertical coordinate z with positive direction upward as shown in Fig. 2. In the following, let $\partial/\partial s$ denote differentiation at constant ϕ_0 and $\partial/\partial\phi_0$ at constant s.

The intensity along the ray is inversely proportional to the cross section ∂S of an infinitesimal ray tube,[9] which varies with the launch angle ϕ_0 according to

$$\frac{\partial S}{\partial \phi_0} = -\frac{\partial r}{\partial \phi_0}\sin\,\phi + \frac{\partial z}{\partial \phi_0}\cos\,\phi. \tag{5}$$

To compute this, XRAY solves the ODE system (4) augmented by

$$\begin{aligned} \frac{\partial}{\partial s}\frac{\partial r}{\partial \phi_0} &= -\frac{\partial \phi}{\partial \phi_0}\sin\,\phi, \\ \frac{\partial}{\partial s}\frac{\partial z}{\partial \phi_0} &= \frac{\partial \phi}{\partial \phi_0}\cos\,\phi, \\ \frac{\partial}{\partial s}\frac{\partial \phi}{\partial \phi_0} &= \frac{\sin\,\phi}{c}c'(z)\frac{\partial \phi}{\partial \phi_0} - \left(c''(z) - \frac{[c'(z)]^2}{c}\right)\frac{\cos\,\phi}{c}\frac{\partial z}{\partial \phi_0}. \end{aligned} \tag{6}$$

In addition, the equation

$$\frac{\partial \tau}{\partial s} = \frac{1}{c} \tag{7}$$

for the travel time τ along the ray is included.

Let $z_{\text{end}} = z_{\text{end}}(\phi_0)$ be the value of z when $r = r_r$, the range of the receiver. For efficiency and robustness in the present investigation, the search technique for eigenrays used in Ref. 6 had to be improved by: (i) using an adaptive search method based on the derivative $z'_{\text{end}}(\phi_0)$; (ii) introducing a criterion for discarding rays with low energy content. The derivative $z'_{\text{end}}(\phi_0)$ can be calculated from the derivatives in the ODE system by

$$z'_{\text{end}}(\phi_0) = \frac{\partial z}{\partial \phi_0} - \frac{\partial r}{\partial \phi_0} \tan \phi_{\text{end}}. \tag{8}$$

The adaptive method is a shooting technique, where a selected launch angle interval is searched by marching. Starting from launch angle ϕ_0^k the new launch angle is defined as

$$\phi_0^{k+1} = \phi_0^k + \Delta\phi_0^k. \tag{9}$$

The angle step size $\Delta\phi_0^k$ is chosen to satisfy the following inequalities (see Fig. 2):

$$\Delta\phi_0^k \leqslant \Delta\phi_{0\max}, \tag{10}$$

$$\Delta\phi_0^k \leqslant |\Delta z_{\max} / z'_{\text{end}}(\phi_0^k)|, \tag{11}$$

$$\Delta\phi_0^k \leqslant |\Delta z_{\max} / z'_{\text{end}}(\phi_0^{k+1})|. \tag{12}$$

First $\Delta\phi_0^k$ is chosen as large as possible with respect to Eqs. (10) and (11). If this $\Delta\phi_0^k$ is too large according to Eq. (12), the step size is halved until all inequalities are satisfied. A large derivative $z'_{\text{end}}(\phi_0^k)$ requires small steps, whereas a small derivative allows for larger steps. Thus, the maximum step size $\Delta\phi_{0\max}$ is introduced to avoid missing a region with a rapidly changing derivative.

This adaptive method reduces the risk of overlooking eigenrays compared to using a constant step size. The difficult case where the method might fail is when $z'_{\text{end}}(\phi_0)$ is zero or $z_{\text{end}}(\phi_0)$ is discontinuous while $z_{\text{end}}(\phi_0)$ is close to the receiver depth z_r.

The resulting set of rays is used to find launch angle intervals enclosing single eigenrays. Then a final search for the eigenray in each interval is done by a derivative-free nonlinear equation solver.[10]

The method for discarding intervals with low energy content is introduced to limit the calculation time. The intensity along each ray is monitored and the ray is discarded if it falls below a preselected threshold times the intensity at spherical spreading.

For spherical spreading the cross section $\partial S^{\text{sphere}}$ of an infinitesimal ray tube is given by

$$\frac{\partial S^{\text{sphere}}}{\partial \phi_0} = \sqrt{r^2 + (z - z_0)^2} \approx r. \tag{13}$$

Thus, by Eq. (5), the condition for a ray to survive can be written

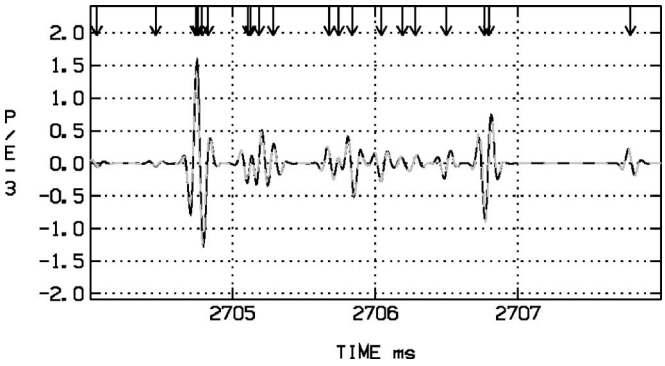


FIG. 3. Comparison of the hybrid method XRAY in black and the full-field method XFEM in dashed gray. The arrows indicate the arrival times of the different propagation paths as calculated by XRAY.

$$\left| -\frac{\partial r}{\partial \phi_0} \sin \phi + \frac{\partial z}{\partial \phi_0} \cos \phi \right| < \text{threshold} \cdot r. \tag{14}$$

The value of the threshold must be chosen so that the excluded eigenrays do not have a major influence on the result.

A. Validation of the model

The accuracy of the hybrid method was assessed in a range-independent case, using an approximation of profile B (see Fig. 1) with a slightly absorbing bottom at a depth of 40 m. Both source and receiver were placed at a depth of 30 m. The results of a simulation with the hybrid method were compared to those obtained by the full-field transform integral method (XFEM).[11] XFEM is a finite element method, with an option to use exact local solutions to the governing differential equation as basis functions, assuming the medium parameters to have a piecewise simple form.

In Ref. 11 homogeneous elements were used giving exponentials as solutions. In the case studied here homogeneous elements are not economical, due to the fast sound speed changes at the thermocline, which can be seen in Fig. 1. Instead elements with linear squared slowness $1/c^2$ are used. The depth-separated frequency-domain wave equation can then be transformed into the Airy differential equation and thus the solutions can be described using Airy functions.

For both XRAY and XFEM the transmitted signal was a 10 kHz Ricker pulse, which was chosen to represent the transmission of an impulse. It resembles the autocorrelation of the bandpass form of the experimental signal shown in Fig. 4. The comparison of the two methods is illustrated in Fig. 3. The black XRAY curve and the dashed gray XFEM curve coincide, showing that XRAY has found all propagation paths and that the frequency is high enough to use ray-tracing. The arrival times of the signal from the different propagation paths are marked by arrows in the figure.

IV. OPTIMIZATION

A. Differential evolution

The *differential evolution algorithm* (DE) is a global search technique for the minimum of an *objective function*.[12,13] Each set of parameter values is called a *model*, and a collection of models is a *population*. The population

evolves through a sequence of generations in order to reach the global optimum, which makes DE conceptually similar to *genetic algorithms*.[14] The use of an entire population of models makes it possible to search the parameter space efficiently and limit the risk of getting trapped at a local minimum.

The first generation is randomly distributed over the parameter space and for each model m_i the objective function value f_i is calculated. The new generation is obtained as follows. For each model m_i:

- Randomly select three other models of the present generation, $m_{p(i)}$, $m_{q(i)}$, and $m_{r(i)}$.
- Define $v_i = m_{p(i)} + F \cdot (m_{q(i)} - m_{r(i)})$.
- Create a *trial model* u_i by choosing each parameter in u_i from v_i with probability CR and from m_i with probability $1-CR$.
- Calculate the objective function value $f(u_i)$. For the next generation m_i is kept unless $f(u_i) < f(m_i)$, in which case m_i is replaced by u_i.

F (weighting factor) and CR (crossover factor) are constants, here set to 0.5 and 0.9, respectively.

B. Objective function

When the sound field is structured into a small number of dominating propagation paths, a possible approach used successfully in ocean acoustic tomography is to match ray arrival times with the peaks of the observed impulse response. Here however, the ray arrivals are not separable and the approach is to find an objective function that measures the discrepancy between modeled and observed impulse response estimates.

For modeled data, the impulse response is estimated by cross correlating the received signal with the transmitted one, yielding

$$\tilde{h}^{\mathrm{sim}}(t) = s_{\mathrm{bp}}(t) \star s_r^{\mathrm{sim}}(t) = h^{\mathrm{sim}}(t) * [s_{\mathrm{bp}}(t) \star s_{\mathrm{bp}}(t)], \quad (15)$$

where $\star$ denotes cross correlation and $*$ convolution.

This should be compared with the experimental impulse response $\tilde{h}(t)$ given by

$$\tilde{h}(y) = s_{\mathrm{bp}}(t) \star s_r(t) = h(t) * [s_{\mathrm{bp}}(t) \star s_{\mathrm{bp}}(t)] + s_{\mathrm{bp}}(t) \star n(t), \quad (16)$$

where $n(t)$ is noise. These estimates are convolutions of the impulse response and the auto-correlation function of the transmitted signal, and they are normalized to have the maximum absolute value 1.

The PRBS signals resemble noise and have an autocorrelation with a single peak, which makes it easy to interpret the impulse response visually. When the bandpass signal is studied, this property is lost and the autocorrelation becomes a series of peaks, as can be seen in Fig. 4.

An estimate of the complex baseband impulse response is given by multiplying $\tilde{h}(t)$ by $e^{-i\omega_c t}$ and applying a low-pass filter.

A natural measure of the similarity of two real functions x and y is the normalized cross correlation

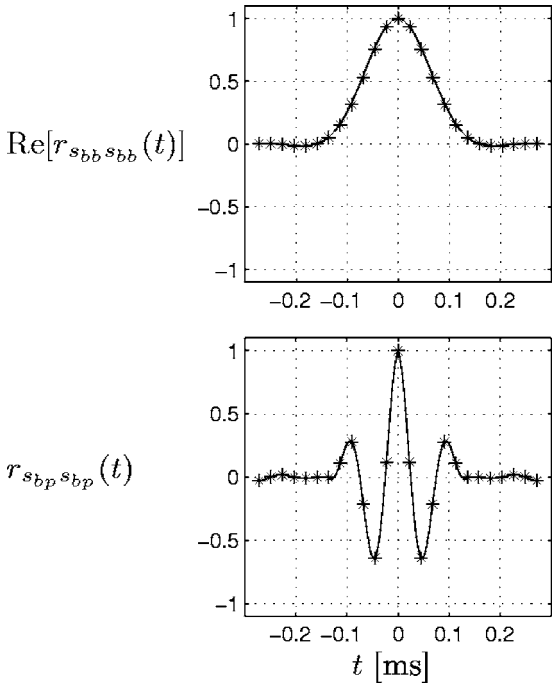


FIG. 4. Real part of the autocorrelation of the complex baseband signal s_{bb} in the upper panel. The imaginary part is essentially zero. Autocorrelation of the bandpass signal s_{bp} in the lower panel.

$$r_{xy}(t) = \frac{x \star y}{\|x\| \, \|y\|}, \quad (17)$$

where $r_{xy}(0)$ can be written

$$r_{xy}(0) = \frac{\langle x, y \rangle}{\|x\| \, \|y\|}, \quad (18)$$

using the notation $\langle x, y \rangle = \int_{-\infty}^{\infty} x(t) y(t) dt$ and $\|x\| = \sqrt{\langle x, x \rangle}$.

However, the measure should focus on the similarity in the interval where $\tilde{h}(t)$ is large and be little affected by differences outside this interval, where the noise content is high. Let this main interval be the minimum interval outside where the signal amplitude is below 0.2. Let T be the length of the interval and let $t=0$ at the beginning of the interval.

Define

$$\lfloor x \rfloor = \begin{cases} x & \text{if } |x| > 0.2, \\ 0 & \text{if } |x| \leqslant 0.2. \end{cases} \quad (19)$$

The target signal is then defined as

$$\tilde{h}^{\mathrm{target}} = \begin{cases} \tilde{h}(t) & \text{if } 0 < t < T, \\ \lfloor \tilde{h}(t) \rfloor = 0 & \text{otherwise}, \end{cases} \quad (20)$$

see Fig. 5.

Now define, for all real τ

$$(\tilde{h}^{\mathrm{sim}})^{\tau} = \begin{cases} \tilde{h}^{\mathrm{sim}}(t+\tau) & \text{if } 0 < t < T, \\ \lfloor \tilde{h}^{\mathrm{sim}}(t+\tau) \rfloor & \text{otherwise}. \end{cases} \quad (21)$$

A measure that focuses on the main interval is now given by

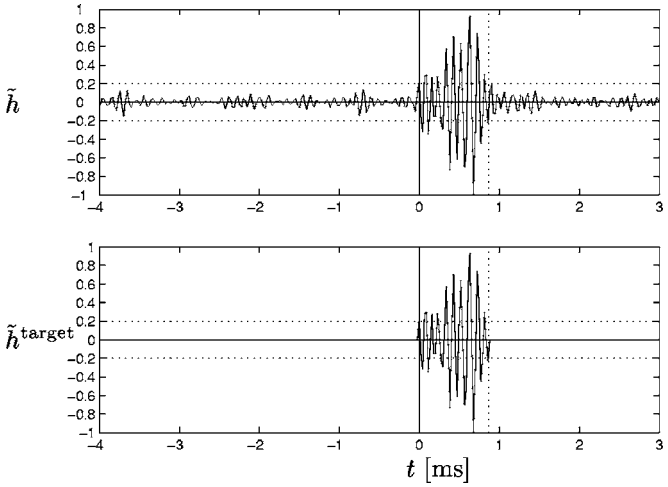


FIG. 5. Construction of the target signal $\tilde{h}^{\text{target}}$.

$$r(\tau) = \frac{\langle (\tilde{h}^{\text{sim}})^{\tau}, \tilde{h}^{\text{target}} \rangle}{\|(\tilde{h}^{\text{sim}})^{\tau}\| \, \|\tilde{h}^{\text{target}}\|}. \tag{22}$$

Like the cross correlation, this function reaches the maximum 1 when the match is perfect, and the objective function f is defined by

$$f = 1 - \max_{\tau} \frac{\langle (\tilde{h}^{\text{sim}})^{\tau}, \tilde{h}^{\text{target}} \rangle}{\|(\tilde{h}^{\text{sim}})^{\tau}\| \, \|\tilde{h}^{\text{target}}\|}. \tag{23}$$

The maximization over τ serves to time align the target and the modeled data, since no absolute time reference for the target is available.

V. MODEL PARAMETERS

The seabed was modeled as an infinitely thick fluid layer, with parameters characteristic for clay.[9,15] The density was set to 1424 kg/m^3, the sound speed to 1480 m/s, and the attenuation to 0.06 dB/λ.

The sound speed profile is constructed by the ray tracing program XRAY using a given set of depth-speed couples as described in Sec. III. Figure 1 shows the two original sound speed profiles measured two days before the communication experiment. To describe these profiles properly and efficiently, eight points were chosen: the surface, the bottom, four points for the thermocline, and two points for the minimum above the thermocline seen in profile A. The point representing the bottom was placed at a depth of 40 m. In Fig. 6 these chosen points are marked for each of the two original profiles. The resulting approximative sound speed profile is also shown together with the original one.

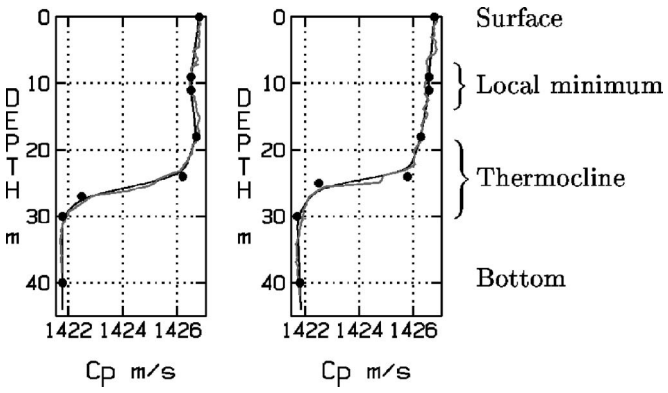


FIG. 6. Approximation of profile A in the left panel and profile B in the right panel with the eight points marked.

TABLE I. The search parameter space described by the eight points.

Point index i	z_i	c_i
1	$z_1=0$	$c_1=1426.8$
2	$8<z_2<10$	$1426.4<c_2<1426.9$
3	$z_3=z_2+2$	$c_3=c_2$
4	$z_4=18$	$1426.3<c_4<1426.8$
5	$23<z_5<25$	$c_5=c_4-0.5$
6	$z_5+1<z_6<z_5+4$	$c_6=1422.5$
7	$z_7=30$	$1421.7<c_7<1421.8$
8	$z_8=40$	$c_8=1421.84$

Among the 16 sound speed profile parameters, six were allowed to vary and the other ones were fixed. The search parameter space is defined in Table I and was chosen to cover sound speed profiles similar to profiles A and B.

The profile characteristics that change in this parameter space are the depth and slope of the thermocline and the properties of the possible shallow sound speed minimum. The characteristics of the thermocline have a strong influence on the ray pattern and the impulse response and are therefore the most important to determine. The parameter space is kept narrow to enable a thorough search. Both the measured sound speed profiles can be described using these parameters as shown in Fig. 6. The entire search parameter space is illustrated in Fig. 7.

The ray tracing was done for launch angles between $-10°$ and $10°$. Rays outside this interval do not contribute to the dominating part of the impulse response, but form a weak tail spread over later time instants.

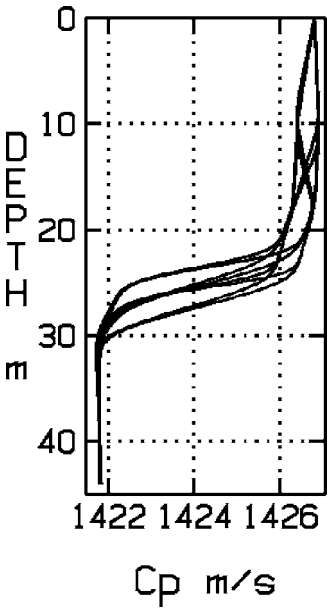


FIG. 7. The resulting sound speed profiles when the extreme values of the parameters in Table I are chosen in all possible combinations.

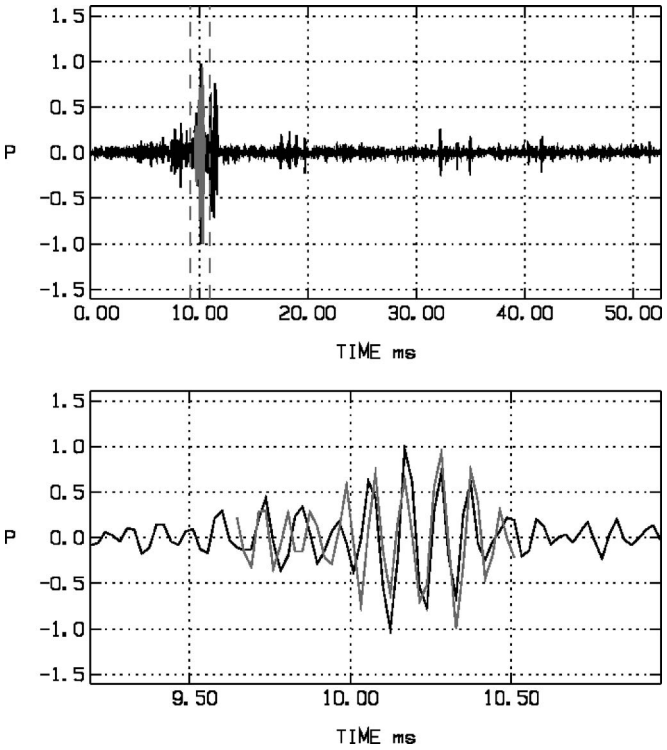


FIG. 8. Profile A—estimated bandpass impulse response (black), target signal at optimal time (gray).

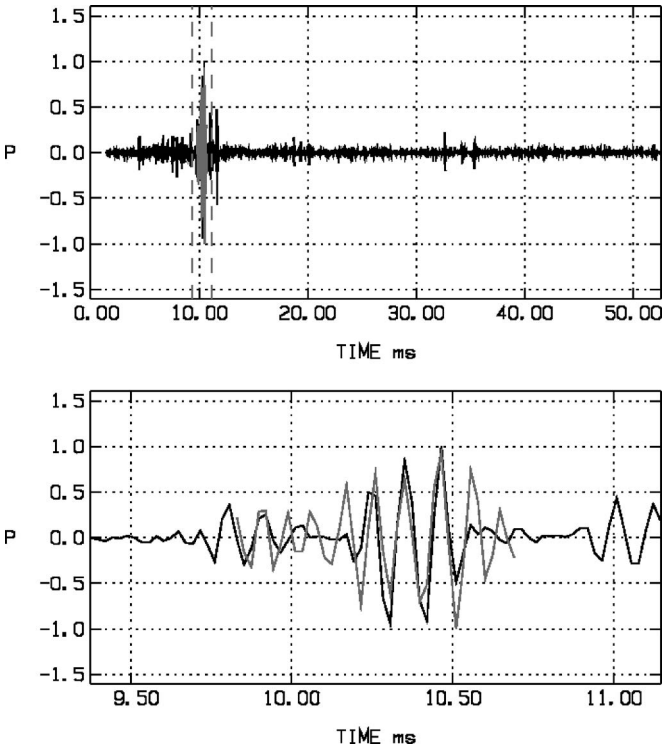


FIG. 10. Profile B—estimated bandpass impulse response (black), target signal at optimal time (gray).

VI. RESULTS

We begin by looking at the estimated impulse responses of profiles A and B that were used to construct the search parameter space described in Table I. The impulse responses are shown in Figs. 8–11. All figures have a time axis with zero at the left end to increase the readability.

Profile A has an impulse response that is too long compared to the target as can be seen in the upper panel of Fig. 8. The gray signal is the target impulse response placed at the time displacement yielding the lowest objective function value, and the dashed vertical lines at approximately 10 ms in the left panel mark a region centered at the target signal. This region is shown in the lower panel and here it is clear that the model prediction has its maximum at the wrong time. At the beginning of the interval, the simulated impulse response has the wrong sign.

These differences can also be illustrated by looking at the magnitude of the complex baseband impulse response, shown in Fig. 9. Here it is easy to see that the total width of the two main peaks is too small.

The impulse response of profile B (Fig. 10) is also slightly too long, but not as much as for profile A. The closeup in the lower panel indicates that the energy in the target region is contained in two time spans, with a gap between them. In the end of the region there is also a lack of energy. Outside the target region one of the narrow peaks from the left panel appears. The baseband response in Fig. 11 confirms this.

To find a better fit between simulated and experimental data the differential evolution algorithm was used to search the six-dimensional parameter space of sound speed profiles defined in Table I. The model with the lowest objective function value was chosen as a starting point for a final local

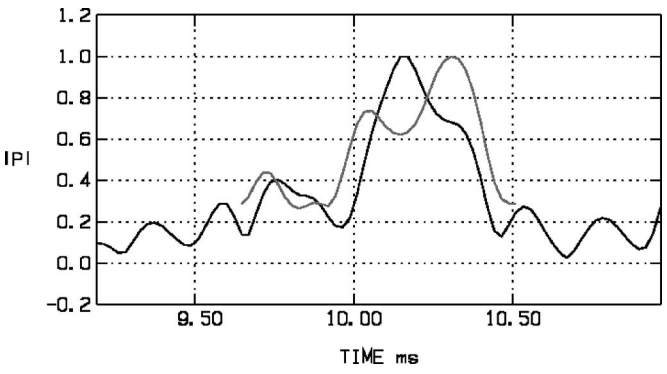


FIG. 9. Profile A—magnitude of the estimated baseband impulse response (black), target signal at optimal time (gray).

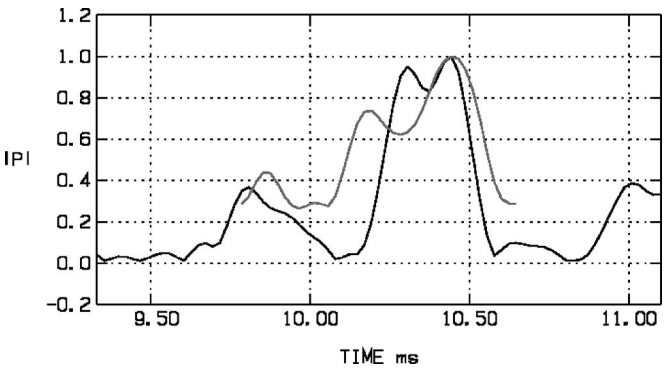


FIG. 11. Profile B—magnitude of the estimated baseband impulse response (black), target signal at optimal time (gray).

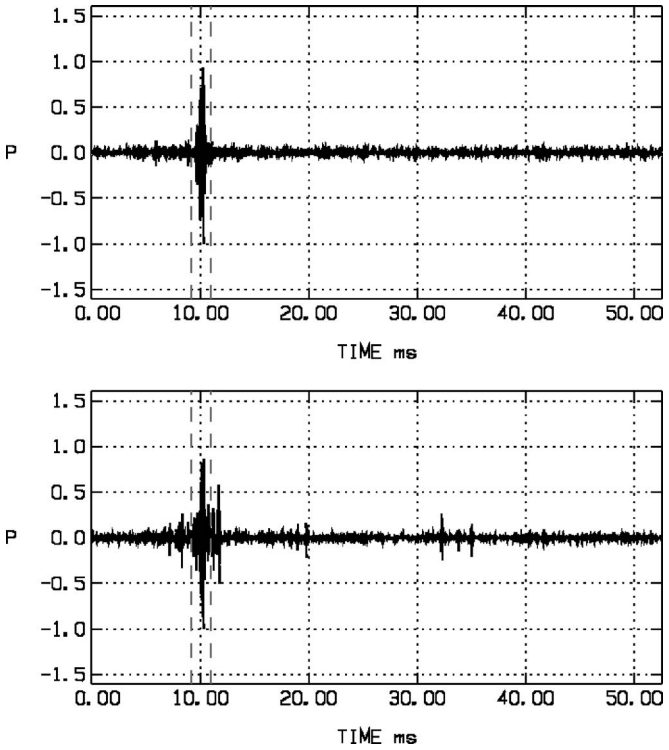


FIG. 12. The observed impulse response at optimal time (upper panel) and the modeled impulse response using the best fit environmental model (lower panel).

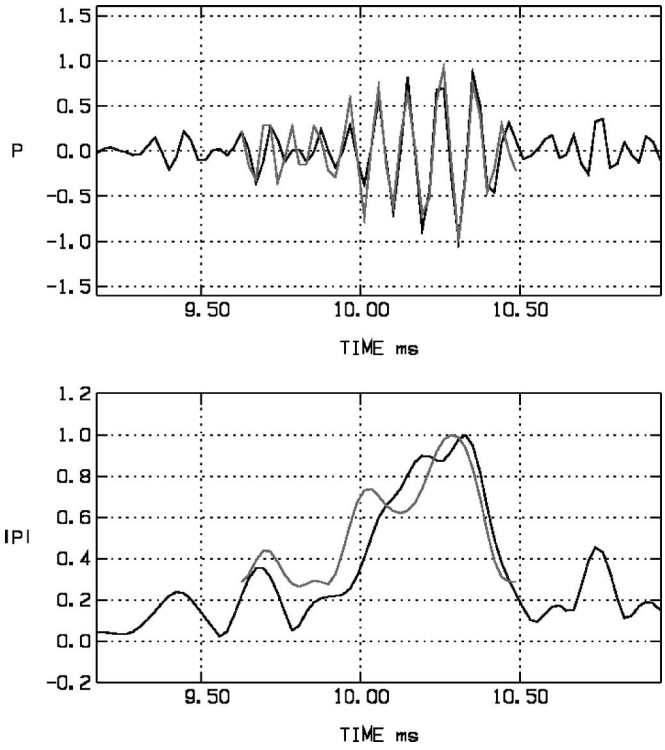


FIG. 13. The best fit profile—estimated bandpass impulse response (upper panel) and magnitude of the estimated baseband impulse response (lower panel). The gray curves show the target signal at the optimal time.

minimization yielding the best fit model, whose estimated impulse response is shown in the lower panel of Fig. 12, with the observed impulse response at the optimal time in the upper panel for comparison.

The dominating part of the impulse response is still somewhat too long, but the fit in the target region has improved significantly, as can be seen in the upper panel of Fig. 13. Many different acoustic propagation paths have their arrival time in this region and contribute to the impulse response.

The differences between simulation and target are easier to see in baseband (Fig. 13, lower panel). Among the three main peaks of the target impulse response, the first and the third one are well modeled by the simulation.

The best fit environmental model is shown in Fig. 14 in black together with profiles A and B in gray. Profile A has the thermocline at about the same depth, but above it the best fit model almost follows a constant slope, whereas profile A has a local sound speed minimum.

A. Different source-receiver configurations

The best fit environmental model was used to investigate different source-receiver configurations. To illustrate the variation of the impulse response with these parameters, simulations were performed where the source depth varied from 4 to 36 m and the receiver depth from 4 to 38 m, with calculations for every second meter.

An example is shown in the left panel of Fig. 15 where the source is kept at a depth of 30 m and the receiver depth is changed. When the receiver is placed below the thermocline, the impulse response is more focused and the influence of late arrivals decreases.

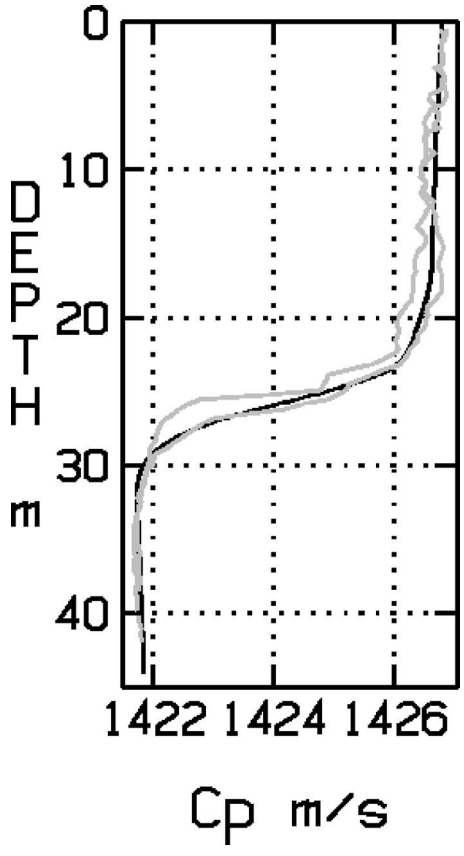


FIG. 14. The best fit environmental model in black, profiles A and B in gray.

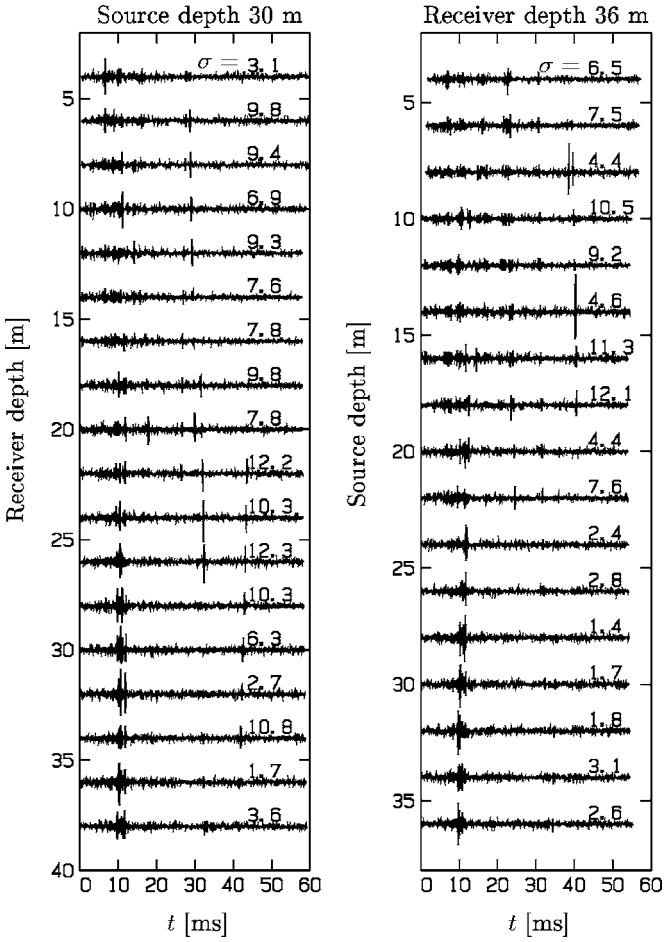


FIG. 15. Impulse responses at varying source and receiver depths: source at depth 30 m with varying receiver depth (left panel) and receiver at depth 36 m with varying source depth (right panel).

Instead, letting the depth of the source vary and keeping the receiver at a depth of 36 m yield the same type of behavior in the right panel of Fig. 15. Here the impulse responses when the source is close to the surface are both more spread in time, and contain less energy.

The different source-receiver configurations offer different conditions for communication. Transmission through a single dominating propagation path is expected to be the easiest situation, and the longer the impulse response is, the more complicated it is to decode the transmitted message. To find the optimal transmitter depth for this environmental model a probability density function $p(t)$ based on the impulse response is considered for which mean arrival time and standard deviation are calculated. To emphasize the importance of high peaks, the impulse response is raised to the power of four, and $p(t)$ is defined by

$$p(t) = \frac{[\tilde{h}^{\mathrm{sim}}(t)]^4}{\int_I [\tilde{h}^{\mathrm{sim}}(\tau)]^4 d\tau}, \tag{24}$$

where I is an interval of width 110 ms centered at $t_{\mathrm{cen}} = \int_{-\infty}^{\infty} t(\tilde{h}^{\mathrm{sim}})^4 dt \Big/ \int_{-\infty}^{\infty} [\tilde{h}^{\mathrm{sim}}(\tau)]^4 d\tau$. The considered interval will thus include all parts of the impulse responses shown in Fig. 15.

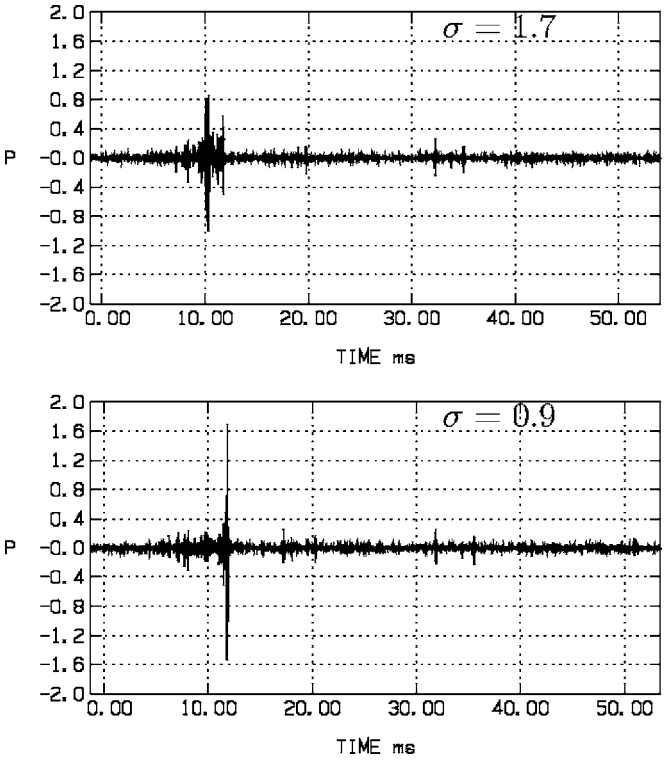


FIG. 16. The best fit impulse response (from optimization), source depth 30 m (upper panel) and the impulse response when source depth is changed to 25 m (lower panel).

The mean arrival time $\bar{t}$ and the standard deviation σ are now defined by

$$\bar{t} = \int_I t p(t) dt, \tag{25}$$

$$\sigma^2 = \int_I (t - \bar{t})^2 p(t) dt. \tag{26}$$

The standard deviation σ is a measure of how focused the impulse response is, and this is used to compare the conditions for data communication at the different transmitter depths. In the experiment the transmitter was placed at a depth of 30 m. The left panel of Fig. 15 shows that deeper transmitter positions are more promising, and calculations of $\sigma(z_0)$ with one meter spacing between transmitter depths z_0 showed that the optimal transmitter depth (the depth with minimal standard deviation) is 25 m. This combination of source and receiver position was also the best when 450 different source-receiver configurations spread over the water column were compared.

The change of source depth from 30 to 25 m is shown in Fig. 16. The same amplitude scaling as in the previous figures is used. Positioning the transmitter at a depth of 25 m gives a stronger and more focused impulse response.

VII. SUMMARY AND CONCLUDING REMARKS

An estimate of the sound speed profile at an underwater data communication experiment was computed, based on two sound speed profiles measured two days before the experiment. The inversion was done by matching modeled im-

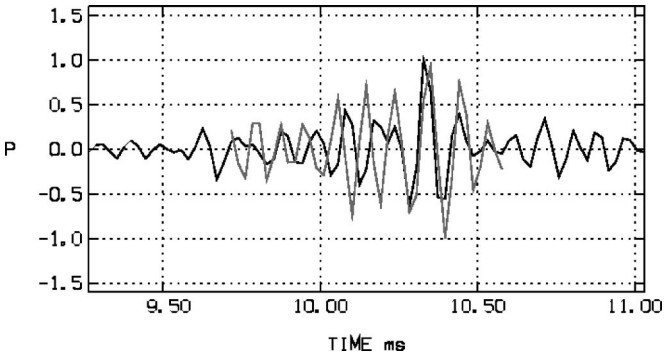


FIG. 17. The best fit profile—estimated bandpass impulse response for rigid bottom (black), target signal at optimal time (gray).

pulse response estimates to observed ones using a differential evolution algorithm combined with a local search technique. The hybrid ray-trace/plane-wave method XRAY was used as a forward model at the inversion.

Rays in the searched launch angle interval lose just a small part of their intensity when they are reflected. The incidence angle at reflection is in most cases subcritical, resulting in almost total reflection (critical grazing angle: 16°). However, the introduced phase shift at the reflection influences the resulting impulse response significantly, as can be seen in Fig. 17, where the best fit sound speed profile has been combined with a rigid bottom. When the phase shifts at the reflections are removed, the agreement with the target signal from Fig. 13 is lost.

The mismatch between observed and modeled impulse response estimates was significantly reduced using the best fit model, compared to the mismatch using profile A or B.

The best fit environmental model was used to investigate the impulse response using different source-receiver configurations. The impulse response was found to be more focused when source and receiver were positioned at or below the thermocline, as was the case in the experiment. The optimal transmitter and receiver depths were predicted to be 25 and 36 m, respectively, compared to 30 and 36 m used at the experiment.

ACKNOWLEDGMENTS

This work was supported by the Swedish National Defence College.

[1]M. Siderius, M. Porter, and P. Hursky, "Measurements and modeling comparisons of underwater communications performance at three shallow-water sites," Proc. Oceans 2003, Vol. **4** (IEEE, Piscataway, NJ, USA, 2003), pp. 1773–1779.

[2]C. Bjerrum-Niese, L. Bjørnø, M. Pinto, and B. Quellec, "A simulation tool for high data-rate acoustic communication in a shallow-water, time-varying channel," J. Oceanic Eng. **21**, 143–149 (1996).

[3]M. Taroudakis, "Ocean acoustic tomography," in Lecture Notes, ECUA 2002, (Gdańisk University of Technology, Gdańsk, Poland 2002), pp. 77–95.

[4]B. Nilsson, N. Olofsson, E. Sangfeldt, and T. Öberg, "Acoustic underwater communication methods for high data rate and for low signal to noise ratios," Technical Report FOI-R–1070–SE, (FOI, Stockholm, Sweden 2003).

[5]J. Proakis and M. Salehi, *Communication Systems Engineering* (Prentice-Hall, Englewood Ciffs, NJ, 2002).

[6]E. Svensson, I. Karasalo, and J.-P. Hermand, "Hybrid raytrace modelling of an underwater communication channel," Proc. ECUA 2004 (TNO Physics and Electronics Laboratory, The Hague, The Netherlands, 2004), pp. 1211–1216.

[7]J. Hovem and D. Knobles, "A range dependent propagation model based on a combination of ray theory and plane wave reflection coefficients," Proc. of the Tenth Int. Cong. Sound Vib. (ITAV, 2003), pp. 2593–2600.

[8]C. de Boor, *A Practical Guide to Splines*, Applied Mathematical Sciences, Vol. **27** (Springer-Verlag, Berlin, 1978).

[9]F. Jensen, W. Kuperman, M. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP Press, Melville, NY, 1994).

[10]N. Anderson and A. Björk, "A new high order method of regula fasi type for computing a root of an equation," BIT, Nord. Tidskr. Inf.behandl. **13**, 253–264 (1973).

[11]I. Karasalo, "Exact finite elements for wave propagation in range-independent fluid-solid media," J. Sound Vib. **172**, 671–688 (1994).

[12]H. Gothäll and R. Westin, "Evaluation of four global optimisation techniques (ASSA, DE, NA, Tabu Search) as applied to anechoic coating design and inverse problem uncertainty estimation," Technical Report FOI-R–1593–SE (FOI, Stockholm, Sweden, 2005).

[13]R. Storn and K. Price, "Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces," J. Global Optim. **11**, 341–359 (1997).

[14]M. Mitchell, *An Introduction to Genetic Algorithms* (MIT Press, Cambridge, MA, 1996).

[15]E. L. Hamilton, "Geoacoustic modeling of the sea floor," J. Acoust. Soc. Am. **68**, 1313–1340 (1980).

Feeding behavior of wild dugongs monitored by a passive acoustical method

Chika Tsutsumi, Kotaro Ichikawa, and Nobuaki Arai
Graduate School of Informatics, Kyoto University, 606-8501 Kyoto, Japan

Tomonari Akamatsu[a)]
National Research Institute of Fisheries Engineering, Fisheries Research Agency, 314-0408, Kamisu, Ibaraki, Japan

Tomio Shinke
System Intech Co., Ltd. R&D Center, 424-8610 Shizuoka, Japan

Takeshi Hara
Japan Fisheries Resource Conservation Association, 104-0054 Tokyo, Japan

Kanjana Adulyanukosol
Phuket Marine Biological Center, 83000 Phuket, Thailand

(Received 24 September 2005; revised 13 May 2006; accepted 12 June 2006)

Little is known about feeding behavior of wild dugongs (*Dugong dugon*) because direct measurements of feeding events in the water were scarcely feasible. In this study, the authors achieved the first successful feeding sound monitoring in a seagrass area using a full-band underwater recording system (called automatic underwater sound monitoring system for dugong: AUSOMS-D). In total, 175 feeding sounds were identified in 205 h of recording. Feeding sounds were only detected at night, implying diurnal differences in the feeding behavior of the studied dugong population. Differences in periodicity of feeding sounds suggested that two or more individuals were in the acoustically observable area. Furthermore, a feeding position monitored by two AUSOMS-Ds was used to calculate source levels of dugong feeding sounds. Assuming spherical propagation, source levels were measured between 70.6 and 79.0 dB rms *re* 1 μPa/$\sqrt{\text{Hz}}$. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2221529]

PACS number(s): 43.30.Sf, 43.66.Gf [WWA] Pages: 1356–1360

I. INTRODUCTION

Dugongs are endangered herbivorous marine mammals inhabiting tropical and subtropical shallow waters that include the Indian and Pacific Oceans. They feed almost exclusively on seagrasses of the families Potamogetonaceae and Hydrocharitaceae (Heinsohn and Birch, 1972). The assessment of feeding ground utilization by dugongs is essential for their conservation. We must understand when and where dugong use specific feeding ground, and daily foraging duration for effective measures of conservation acts.

Present methodologies cannot provide comprehensive monitoring of dugong feeding behavior. Little is known about feeding behavior of wild dugongs because direct measurements of feeding events in the water have not been feasible, especially at nighttime. Today, acoustic monitoring methods are widely used to observe marine mammals. Sonar signals (Akamatsu *et al.*, 2001; Wang *et al.*, 2005) and other sounds (reviewed by Desharnais and Hay, 2004) produced by cetaceans have been acoustically detected, allowing positioning and even tracking of migrating individuals. Dugong vocalizations have also been used in behavioral studies (Anderson and Barclay, 1995; Ichikawa *et al.*, 2004; Shinke *et al.*, 2004). The full band automatic underwater sound monitoring system for dugongs (AUSOMS-D) recording system has revealed the vocalizing behavior of dugongs (Ichikawa *et al.*, 2006). In the present study, we used AUSOMS-D to record the feeding sounds of dugongs in seagrass areas.

II. MATERIALS AND METHODS

For the initial assessment of dugong feeding sounds, we recorded the underwater feeding sounds of a captive male at Toba Aquarium, Toba City, Mie Prefecture, Japan. Each day, the dugong fed on whole eelgrass (*Zostera marina*) anchored in a net plate. A hydrophone (ST 1020; Oki Electric Industry Co., Tokyo, Japan) was suspended in the tank at a depth of 1 m. During feeding, sounds were recorded by a digital audio tape recorder (TCD-D8; Sony Corp., Tokyo, Japan). The mouth movement of the feeding dugong was videotaped simultaneously. In the off-line analysis, mouth movements and recorded sounds were compared to ensure that the sounds corresponded to the feeding events.

The AUSOMS-D had a stereo full band recording system developed by System Intech Co., Ltd. of Tokyo, Japan. The sampling frequency was 44.1 kHz and the dynamic range was 74–140 dB *re* 1 μPa with a 16-bit resolution. The system was housed in a water-resistant case. The devise also

[a)]Corresponding author, Tomonari Akamatsu. Electronic mail: akamatsu@affrc.go.jp

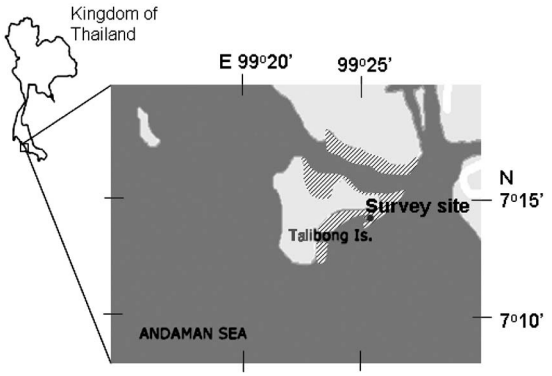


FIG. 1. Map of the survey site off Talibong Island, Trang Province, Thailand. Shaded area is the seagass grounds in tidal flats.

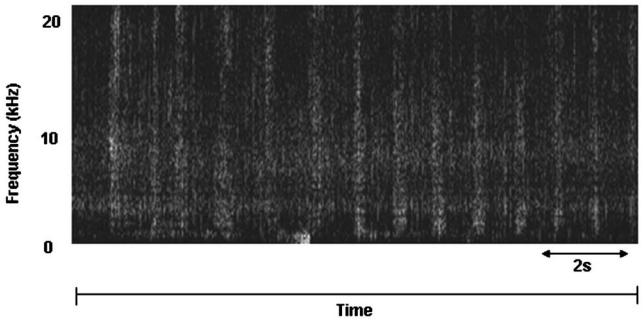


FIG. 2. Sonagram of a typical feeding event. A feeding event consists of a sequence of a feeding sounds. Each feeding sound corresponds to single mastication action of a dugong.

had an 80-GB hard disk and could be operated up to 124 h. The sound source direction can be calculated from the time arrival difference of sound to dual-channel hydrophones of AUSOMS-D. The sound source direction is calculated using trigonometric function assuming underwater sound speed as 1542 m/s (measured *in situ*, personal communication of K. Ichikawa). Further details of the AUSOMS-D have been provided by Ichikawa *et al.* (2006).

From 21 February to 4 March, 2005, we conducted a passive acoustic survey in seagrass grounds off Talibong Island, Trang Province, Thailand (Fig. 1). We deployed single AUSOMS-D at the survey site. A time depth recorder (UME-DT; Little Leonardo, Tokyo, Japan) was deployed beside the AUSOMS-D to measure the tidal level during observations.

From 27 October to 1 November, 2005, we conducted a similar survey in the same seagrass grounds in March. We deployed two AUSOMS-Ds at the survey site to identify the sound source position using two bearing angle from AUSOMS-Ds. The distance between these AUSOMS-Ds was 10 m. The interhydrophone intervals of two AUSOMS-Ds were 2.87 and 2.94 m. The left hydrophones of these AUSOMS-Ds were placed on the north side approximately perpendicular to the water's edge during lowest tide.

We also mapped dugong feeding trails by visual observation and measured the length and width of each trail at the beginning and the end of the acoustical observation. Comparing between two maps, newly appeared feeding trails in the focal area could be identified.

The first author of this paper listened the sound files recorded by the AUSOMS-D. No other person performed audial identification of feeding sounds. The spectral characteristics of the sounds were analyzed using COOL EDIT PRO software (Syntrillium Software Corp., Phoenix, AZ, USA). Once a feeding sound was detected either audially or by visually inspecting the sonagram, the absolute time and duration were noted in a spreadsheet. The data collected when the AUSOMS-D was exposed to the air during the lowest tide were excluded. Direction of dugong feeding sound was calculated by cross correlation between recordings by two hydrophones. The position of each sound was plotted by the combination of short base line of AUSOMS-D. Then the trajectories of feeding events were obtained. Visually observed feeding trails were matched with the acoustically identified trajectory.

III. RESULTS

In total, 237.5 h of data were recorded at the study site in February and March 2005. Only data collected when the AUSOMS-D was submerged were used for analyses to exclude wave noises at low tide. After excluding those data, 205 h of effective recordings remained for analyses. The several to tens of feeding sounds were observed successively in one feeding event (Fig. 2). Separations of more than 8 s between feeding sounds were defined as cutoffs between different feeding events. In total, 175 feeding events were identified (Table I).

Four feeding events were recognized both acoustically and visually (Fig. 3). The acoustical characteristics of these feeding sounds such as peak frequencies, noise levels, and source sound pressure levels are summarized in Table II. The power spectrum is indicated in Fig. 4. The peak frequency of sounds in the four feeding events ranged from 1430 to 3232 Hz. The source level of four feeding sound events ranged from 70.6 to 79.0 dB rms *re* 1 μPa-m/$\sqrt{\text{Hz}}$. The spherical sound propagation was assumed for the source level calculation because the distance to the animal was within 6 m and the water depth of the present area was a few meters, depending on the tide level. The acoustical characteristics of feeding sounds in captivity could not be estimated because pump noise from the aquarium filtering device masked the lower frequency component of the feeding sounds.

The occurrence frequencies of sound intervals during feeding were compared between wild and captive recordings (Fig. 5). The most frequent interval of feeding sounds was 0.9 s (S.D.=0.4) in the wild and 0.8 s in captivity. As clearly seen in Fig. 5, the feeding sounds had regular interval and a smaller standard deviation in both environments.

Irregular sound intervals were occasionally observed in the wild. Figure 6(a) shows the time interval sequence of feeding sounds. This figure was divided into two separate periodical interval sequences [Figs. 6(b) and 6(c)]. The average intervals of each sequence were 0.8 and 0.9 s, respectively.

The minimum depth of the observed site when the feeding sound was observed was 0.86 m in February and March 2005. All feeding sounds were recorded at night for the study sites (Table I). In contrast, no feeding sounds could be detected during the daytime, even at high tide.

TABLE I. Recording sessions and number of feeding events observed in two different types of seagrass- (*Halophila ovalis* and *Enhalus acoroides*) dominated sites.

Date (2005) Recording duration	Dominant species	No. of feeding events
21–21 Feb. 9:50–14:40 (4.8 h)	*Halophila ovalis*	0
21–22 Feb. 17:00–2:20 (9.3 h)	*Halophila ovalis*	94
22–22 Feb. 7:00–14:40 (7.6 h)	*Halophila ovalis*	0
22–23 Feb. 17:00–3:10 (10.2 h)	*Enhalus acoroides*	15
23–23 Feb. 7:00–16:10 (9.2 h)	*Enhalus acoroides*	0
23–24 Feb. 18:30–3:00 (8.5 h)	*Enhalus acoroides*	17
24–24 Feb. 7:10–16:10 (9 h)	*Enhalus acoroides*	0
24–25 Feb. 18:40–4:00 (9.3 h)	*Enhalus acoroides*	31
25–25 Feb. 7:40–16:30 (8.8 h)	*Enhalus acoroides*	0
25–26 Feb. 19:20–4:40 (9.3 h)	*Enhalus acoroides*	11
27–27 Feb. 9:40–18:30 (8.8 h)	*Enhalus acoroides*	0
27–28 Feb. 18:30–6:30 (12 h)	*Enhalus acoroides*	0
28 Feb.–3 Mar. Daytime (6:30–18:30) (48 h)	*Enhalus acoroides*	0
28 Feb.–4 Mar. Nighttime (18:30–6:30) (48 h)	*Enhalus acoroides*	7
4–4 Mar. 6:30–8:30 (2 h)	*Enhalus acoroides*	0

IV. DISCUSSION

Sounds extracted as feeding events in the present field study were considered to be the feeding sounds of wild dugongs. The acoustically monitored positions of the sound source matched with the visually confirmed feeding trails (Fig. 3). The sound occurrence intervals during feeding agree well between captive and wild data (Fig. 5). Actually, the human listener found it hard to discriminate feeding sounds of dugongs between in captivity and in the wild.

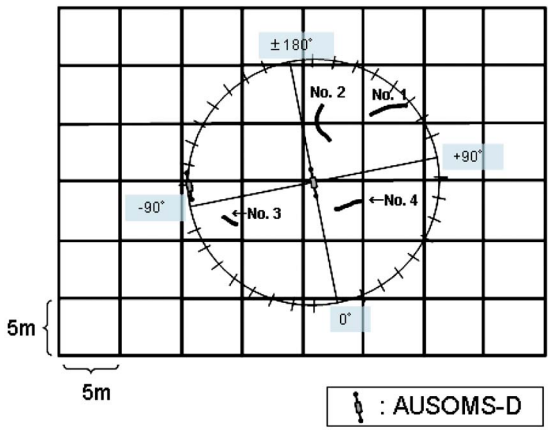


FIG. 3. The feeding trails appeared identified both visually and acoustically during the observation period.

TABLE II. (a) The peak frequency (Freq), sound pressure level (SPL), noise level (NL) and S/N ratio of three local maximums of power spectrum (fp1 to 3) in four feeding events (dB *re* 1 μPa/$\sqrt{\text{Hz}}$). See also Fig 4. (b) The source sound pressure level of four feeding sound events (dB *re* 1 μPa/$\sqrt{\text{Hz}}$).

(a)

Event no.		fp1	fp2	fp3
1	Freq (Hz)	1430.0	2290.0	3050.0
	SPL (dB)	58.8	58.1	59.6
	NL (dB)	51.2	51.2	54.0
	S/N (dB)	7.6	6.9	5.6
2	Freq (Hz)	1659.0	2754.0	···
	SPL (dB)	61.1	62.4	···
	NL (dB)	51.4	54.0	···
	S/N (dB)	9.7	8.4	···
3	Freq (Hz)	1445.0	1862.0	3232.0
	SPL (dB)	73.0	71.1	65.2
	NL (dB)	53.6	53.3	54.0
	S/N (dB)	19.4	17.8	11.2
4	Freq (Hz)	1788.0	2344.0	3162.0
	SPL (dB)	67.6	67.2	65.3
	NL (dB)	50.0	51.8	54.7
	S/N (dB)	17.6	15.4	10.6

(b)

Event no.	fp1	fp2	fp3	distance (m)
1	74.4	73.7	75.2	6
2	70.6	71.9	unclear	3
3	79.0	77.1	71.2	2
4	73.6	73.2	71.3	2

As depicted in Fig. 3, acoustically monitored feeding trajectory and visually confirmed feeding trails could be matched to each other. The reliable observable distance of clearly monitored feeding sound was less than 10 m. According to the signal-to-noise ratio of 7.6 to 19.4 dB (Table II) of these feeding events, theoretical observable range would be approximately 10 m. Feeding sounds of dugong were very faint, but the periodical structure of the sound makes dis-

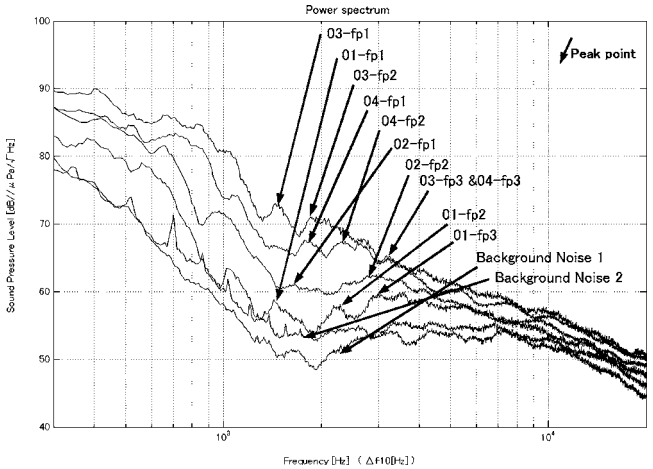


FIG. 4. The power spectrum of four feeding sessions. Three local peaks were indicated by arrows. The unclear third peak of second feeding event was not indicated.

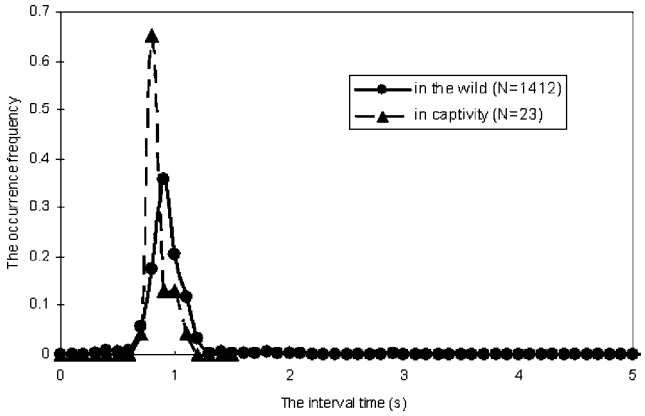


FIG. 5. The occurrence frequency of feeding sounds in the wild and in captivity versus the interval time.

crimination possible from background noise. Although the observable range is limited, the advantage of the acoustical monitoring is nocturnal operation. In the present study, all of the feeding events were observed at nighttime.

Irregular sound intervals strongly suggested that two or more dugongs fed in the observation area at the same time (Fig. 6). The sound interval sequence was divided into two periodic sequences. Each interval of these sequences (0.8 and 0.9 s) was close to the most frequent interval for the feeding sound (0.9 and 0.8 s in the wild and in captivity, respectively; Fig. 5). Thus, the minimum number of feeding dugongs should also be observable acoustically.

We could measure the minimum water depth in which dugongs could enter. This information has been difficult to observe by commonly used visual observation. The minimum depth when feeding sounds were recorded in the present study was 0.86 m. The body height is 0.39±0.067 m, when the body length of a dugong is 2.2±0.3 m (measured by Adulyanukosol, personal communication). A dugong could therefore enter a seagrass bed as shallow as twice its body height.

By far, the use of seagrass areas by dugongs has been measured by trail observations (Heinsohn *et al.*, 1977; Anderson and Birtles, 1978; De Iongh *et al.*, 1995, Preen, 1995; De Iongh *et al.*, 1997; Nakaoka and Aioi, 1999; Mukai *et al.*, 2000; Nakaoka *et al.*, 2002; Mukai and Watanabe, 2003). We can now use additional acoustical cues to monitor underwater feeding behavior of dugongs. The feeding sounds provide not only information on the attendance of feeding dugongs but also on duration of a feeding session and the position of dugongs at a feeding ground. Even more, the passive acoustical system enables observation of the minimum number of simultaneous feeders. All of these mentioned above are the key factors for the better understanding of dugong habitat use and thus facilitate their conservation.

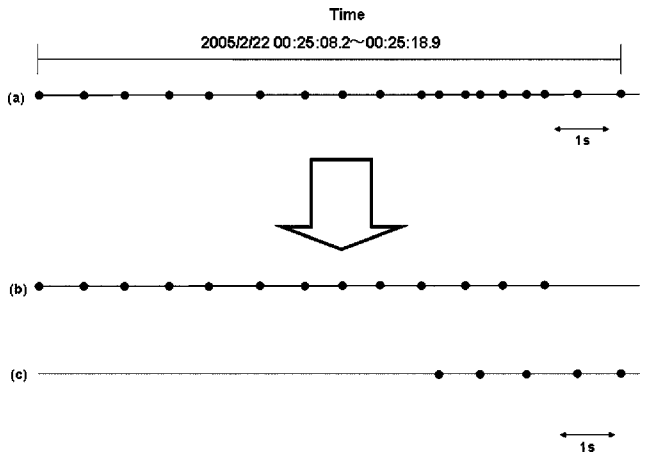


FIG. 6. Diagram (a) indicates the irregular intervals of mastication sounds. (a) was divided into two separate periodical interval sequences (b) and (c). The solid circle indicates the time at feeding sound detected.

ACKNOWLEDGMENTS

We thank all members of the National Research Council in Thailand, Phuket Marine Biological Center, Chaomai National Park (National Park, Wildlife and Plant Department), System Intech Co. Ltd., Metocean Environment Inc., and Okinawa Environmental Research Co., Ltd. who helped support these field experiments. We also thank Dr. Kongkiat Kittiwattanawong, Anuwat Prasittipornkun, Pantaruk Buakaew, and Sumain Panpum of the Phuket Marine Biological Center, Hideharu Tamori and Kazuo Takeuchi of the Japan Fisheries Resource Conservation Association, and Akihiko Matsuda of Fisheries Rsearch Agency, Japan for their assistance in conducting the acoustical survey in Thailand. We thank Dr. Hiroshi Mukai of Hokkaido University for helpful comments and assistance in the seagrass survey. We further thank the staff and students in the laboratories of Biosphere Informatics, Graduate School of Informatics, Kyoto University. This study was partly supported by the Cabinet Office, Grants-in-Aid for Scientific Research (Grant Nos. 16255011 and 16658081), Program for Promotion of Basic Research Activities for Innovative Biosciences, and Information Research Center for Development of Knowledge Society Infrastructure, the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Akamatsu, T., Wang, D., Wang, K., and Wei, Z. (**2001**). "Comparison between visual and passive acoustic detection of finless porpoises in the Yangtze River, China," J. Acoust. Soc. Am. **109**, 1723–1727.

Anderson, P. K., and Barclay, R. M. R. (**1995**). "Acoustic signals of solitary dugongs: physical characteristics and behavioral correlates," J. Mammal. **76**, 1226–1237.

Anderson, P. K., and Birtles, A. (**1978**). "Behaviour and ecology of the dugong, *Dugong dugon* (Sirenia): observations in Shoalwater and Cleveland Bays, Queensland," Aust. Wildl. Res. **5**, 1–23.

De Iongh, H. H., Bierhuizen, B., and van Orden, B. (**1997**). "Observations on the behaviour of the dugong (*Dugong dugon* Müller, 1776) from waters of the Lease Islands, eastern Indonesia," Contr. Zool. **67**, 71–77.

De Iongh, H. H., Wenno, B. J., and Meelis, E. (**1995**). "Seagrass distribution and seasonal biomass changes in relation to dugong grazing in the Moluccas, East Indonesia," Aquat. Bot. **50**, 1–19.

Desharnais, F., and Hay, A. E. (**2004**). "Overview of the 2003 workshop on detection and localization of marine mammals using passive acoustics," Can. Acoust. **32**, 9–16.

Heinsohn, G. E., and Birch, W. R. (**1972**). "Foods and feeding habitats of the dugong, Dugong dugon (Erxleben), in northern Queensland, Australia," Mammalia **36**, 414–422.

Heinsohn, G. E., Wake, J., Marsh, H., and Spain, A. V. (**1977**). "The dugong (*Dugong dugon* (Müller)) in the seagrass system," Aquaculture **12**, 235–248.

Ichikawa, K., Arai, N., Akamatsu, T., Shinke, T., Hara, T., and Adulyanukosol, K. (**2004**). "Acoustical analyses on the calls of dugong," Proceedings of the 4th SEASTAR 2000 workshop, pp. 72–76.

Ichikawa, K., Tsutsumi, C., Arai, N., Akamatsu, T., Shinke, T., Hara, T., and Adulyanukosol, K. (**2006**). "Dugong (*Dugong dugon*) vocalization patterns recorded by automatic underwater sound monitoring systems," J. Acoust. Soc. Am. **119**, 3726–3733.

Mukai, H., and Watanabe, K. (**2003**). "Habitat preference for foraging by dugongs within *Halophila ovalis* beds in Southern Thailand," Conservation of Tropical Seagrass Beds with special reference to their role on function of coastal ecosystem, edited by I. Koike, Ocean Research Institute, University of Tokyo, pp. 149–174.

Mukai, H., Aioi, K., Lewmanomont, K., Matsumasa, M., Nakaoka, M., Nojima, S., Supanwanid, C., Suzuki, T., and Toyohara, T. (**2000**). "Dugong grazing on *Halophila* beds in Haad Chao Mai National Park, Thailand," Biol. Mar. Medit. **7**, 268–270.

Nakaoka, M., and Aioi, K. (**1999**). "Growth of seagrass *Halophila ovalis* at dugong trails compared to existing within-patch variation in a Thailand intertidal flat," Mar. Ecol.: Prog. Ser. **184**, 97–103.

Nakaoka, M., Mukai, H., and Chunhabundit, S. (**2002**). "Impacts of dugong foraging on benthic animal communities in a Thailand seagrass bed," Ecol. Res. **17**, 625–638.

Preen, A. (**1995**). "Impacts of dugong foraging on seagrass habitats: observational and experimental evidence for cultivation grazing," Mar. Ecol.: Prog. Ser. **124**, 201–213.

Shinke, T., Shimizu, H., Ichikawa, K., Arai, N., Matsuda, A., and Akamatsu, T. (**2004**). "Development of automatic underwater sound monitoring system version 1," Proceedings of The 2004FY Annual Meeting of the Marine Acoustics Society of Japan, pp. 33–36., (in Japanese).

Wang, K., Wang, D., Akamatsu, T., Li, S., and Xiao, J. (**2005**). "A passive acoustic monitoring method applied to observation and group size estimation of finless porpoises," J. Acoust. Soc. Am. **118**, 1180–1185.

Estimated detection distance of a baiji's (Chinese river dolphin, *Lipotes vexillifer*) whistles using a passive acoustic survey method

Kexiong Wang and Ding Wang
Institute of Hydrobiology, The Chinese Academy of Sciences, Wuhan 430072, People's Republic of China

Tomonari Akamatsu,[a] Kaoru Fujita, and Rika Shiraki
National Research Institute of Fisheries Engineering, Fisheries Research Agency, Hasaki, Kamisu, Ibaraki 314-0408, Japan

(Received 20 September 2005; revised 12 May 2006; accepted 12 June 2006)

Source levels and phonation intervals of whistles produced by a free-ranging baiji (Chinese river dolphin) were measured in the seminatural reserve of Shishou in Hubei, China. A total of 43 whistles were recorded over 12 recording sessions. The mean dominant frequency (the frequency at the highest energy) was 5.7 kHz (s.d.=0.67). The calculated source level was 143.2 dB rms re 1 μPa (s.d.=5.8). Most phonation intervals were shorter than 460 s, and the average interval was 205 s (s.d.=254). Theoretical detection range of baiji's whistle was 6600 m at the present study site, but it could reduce a couple of hundred meters in practical noisy situation in the Yangtze River. Sporadic phonation (205 s interval on average) with relatively faint signal of baiji was considered to be difficult to be detected by a towing hydrophone system. Stationed monitoring or slow speed towing of hydrophones along the river current is recommended. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2221416]

PACS number(s): 43.30.Sf, 43.66.Gf, 43.80.Ka [WWA] Pages: 1361–1365

I. INTRODUCTION

The freshwater baiji (*Lipotes vexillifer*) once thrived in their only habitat, the Yangtze River, which runs though central China. But fewer than 100 dolphins are thought to be left in the river, which has become a busy, polluted highway (Dalton, 2006). The baiji will become extinct if effective measures are not taken to halt its declining population. The Chinese government's conservation plan involves capturing baiji in the Yangtze River and subsequently releasing them into a temporary refuge such as an oxbow lake (the seminatural reserve) located in Shishou, Hubei (Wei *et al.*, 2002), or into a rearing pool at the Institute of Hydrobiology, the Chinese Academy of Sciences in Wuhan. Conservation of baiji requires successful location and capture in the Yangtze River. However, the extremely small population size and the turbidity of the river's water quality prevent efficient capture.

Over the last few decades, researchers have applied passive acoustic survey methods when observing phonating aquatic mammals (e.g., Miller and Tyack, 1998; Janik *et al.*, 2000; Akamatsu *et al.*, 2001; Lammers and Au, 2003; Desharnais and Hay, 2004; Wang *et al.*, 2005). Passive acoustic surveys can be effective for identifying species in wild and semi-wild habitats. Passive acoustic surveys have several advantages over visual surveys: they can be operated automatically and limit observer bias, since an equivalent detection threshold can be maintained over long time periods; also, acoustic surveys can be conducted continuously. The uniqueness of acoustic signals permits nocturnal observations without illumination. This methodology does have a limitation, in that the source level and the number of phonations occurring over a unit of time can both affect detection distance and probability.

Baiji produce narrow-band tonal vocalizations called whistles; dominant whistle frequencies and duration are 5.8 kHz (±0.85) and 907 ms (±384), respectively (Wang *et al.*, 1999). While acoustic characteristics of the baiji's whistle have already been identified, practical application of passive acoustic methodology to this species requires additional information about source levels and phonation intervals. Use of these parameters during a passive acoustic survey would allow calculation of baiji detection distance and provide optimal acoustic survey methodology. This study used recording conducted simultaneously with Wang *et al.* (1999) using different type of recording equipment

II. MATERIALS AND METHODS

A female baiji was captured in the main stream of the Yangtze River on 19 December 1995, and was subsequently introduced into the seminatural reserve. At the time of capture, she weighed 150 kg and measured 229 cm in body length (Wang *et al.*, 1999). Recordings were made from 28 to 31 January 1996, approximately one month after capture.

Three recording systems were employed for the recordings. The primary recording system consisted of a hydrophone (B&K 8103, Denmark; sensitivity, −211 re 1 V/μPa), a charge amplifier (B&K 2635, Denmark), and a data recorder (Sony PCHB 244, Japan). The secondary recording system was a backup system consisting of a hydrophone (Oki ST1020, Japan; sensitivity, −175.1 dB re 1 V/μPa at

[a]Electronic mail: akamatsu@affrc.go.jp

1 kHz), a sound level meter (Oki SW1020), and the audio track of the 8-mm video recorder (Sony CCD-TR11). The third recording system shared the hydrophone and the sound level meter of the secondary recording system, but had a different recorder (PMD-430; Marantz, Japan). The present study mainly used the primary system and Wang *et al.* (1999) used the third system. The double or triple use of sound recording systems ensured secure observation of the free-ranging baiji, a unique opportunity for study. All three systems had flat (within 1 dB) frequency responses up to 16 kHz.

The hydrophone was suspended at 2 m in depth from a floating stationary boat with its engine shut off during all observation; this helped to eliminate noise contamination and prevented disturbing the animal's phonating behavior. During the recording, the animal kept similar position swimming circular for a couple of minutes to several tens of minutes. While the boat was stopped, underwater sound was recorded continuously. When the animal moved away from the boat and no further sound could be detected, the recording session was terminated and we moved to a new position for the next recording session.

Data recorded by the primary recording system were used to measure sound pressure levels. Due to the 8-mm video recorder's automatic gain control, data recorded by the secondary system were not used to measure the sound pressure levels. The internal calibration signal (1 kHz pure tone signal with 2 V peak to peak amplitude) of the data recorder (PCHB244) was used as a reference level to compare the voltage output of the charge amplifier (B&K 2635) while sound pressure levels were measured. The output voltage of the charge amplifier was converted to the sound pressure level by using the hydrophone sensitivity −211 re 1 V/μPa.

Two experienced visual observers measured the distance to the animal during each of its respirations. The distance estimation was relatively easier in the present area, using the bank and trees on both sides of the lake as reference points. The observers could know the present position and distance to either of the banks by a handy GPS and map of the oxbow. The width of the channel was 1 km. Therefore the nearest bank was within 500 m from the observation boat. The estimated distance to the animal was 100–250 m, which was fairly comparable to the distance to the bank. Additionally, there was only one baiji in the focal area. The animal swam circular and surfaced nearly the same position. This circumstance allowed cross checking of the distance to the animal by both of the visual observers.

Vocalizations recorded within 1 min of the beginning of the observers' announcement of the distance were used to estimate sound pressure level to prevent error of distance estimation due to the movement of the animal. The source level was calculated by the overall sound pressure level within the frequency band of whistles and the propagation loss, assuming spherical sound propagation from the animal to the hydrophone. This calculation did not take into account absorption loss, because the absorption coefficient at 5.8 kHz, the previously measured baiji whistle frequency (Wang *et al.*, 1999), was negligible (Urick, 1983).

The band sound pressure levels (hereafter, "band levels") of baiji whistles were calculated using COOL EDIT PRO (Syntrillium Software Co., USA). An analysis bandwidth of 43 Hz corresponded to the 1024 point analysis window (23 ms) with a 44.1 kHz sampling frequency. The present band level is close to the overall sound pressure level of the whistle within each 23 ms bin, because 43 Hz band includes frequency band of the baiji's whistle and excludes noise outside of the band. Frequency and amplitude of whistles of baiji are usually modulated. Bandwidth and spectrum level changes even within a whistle. Therefore, maximum sound pressure level at specific point in a whistle is difficult to be measured by a long analysis window. Short FFT size with 43 Hz band is a practical choice to analyze the maximum sound pressure level in a whistle.

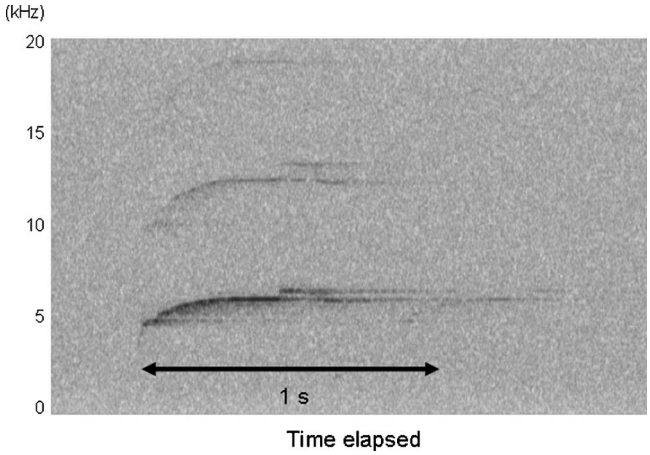


FIG. 1. Typical sonagram of a baiji whistle: a narrow-band tone burst with relatively upswept characteristics (see Table I for details).

Band levels were measured at the time of each whistle's greatest intensity. The present study measured the beginning, end, minimum, and maximum frequencies independently, as well as the duration of each whistle for comparison with previous data sets measured by Wang *et al.* (1999). The −6 dB frequency bandwidth of whistles was also determined. A longer sampling duration 743 ms of 32 768 points at 44.1 kHz sampling frequency was employed for the bandwidth analysis to improve frequency resolution.

The repetition rate of phonation is essential to properly evaluate the performance of passive acoustic monitoring. During each recording session, phonation intervals were measured between successive whistles using data recorded with primary (less sensitive but with fixed gain) and secondary measures (sensitive but with floating gain).

III. RESULTS

A total of 43 baiji whistles were identified during 12 recording sessions, adding 6 whistles to the analysis by Wang *et al.* (1999). Figure 1 shows a typical whistle sonagram; the mean dominant frequency (the frequency at the highest energy) was 5.7 kHz (s.d.=0.67). Baiji whistles had a narrow-band tone burst lasting approximately 1 s, usually with an upswept frequency modulation. Initially, frequency modulations were relatively steep, and then stabilized to around 5–6 kHz.

TABLE I. Comparisons between acoustic whistle characteristics taken in a previous study (Wang *et al.*, 1999) and in the present study using the same free-ranging baiji subject. Based on the fundamental sonagram contours, the beginning, end, minimum, and maximum frequencies and the vocalization duration were measured according to the method described in Wang *et al.* (1999). Minimum, maximum, average, and standard deviation are indicated for each parameter.

	Present	Wang (1999)
Beginning frequency (kHz)		
Min	3.8	3.8
Max	6.2	6.0
Average	5.0	5.0
s.d.	0.47	0.59
End frequency (kHz)		
Min	4.1	4.5
Max	6.8	6.7
Average	5.8	5.8
s.d.	0.63	0.56
Minimum frequency (kHz)		
Min	3.8	3.8
Max	6.1	6.0
Average	4.9	5.0
s.d.	0.48	0.60
Maximum frequency (kHz)		
Min	4.6	4.5
Max	6.8	6.8
Average	6.1	5.9
s.d.	0.60	0.62
Duration (s)		
Min	0.5	0.4
Max	1.6	1.8
Average	1.0	1.0
s.d.	0.23	0.36

The average values for beginning, end, minimum, and maximum frequencies and duration time were 5.0, 5.8, 4.9, 6.1 kHz, and 1.0 s, respectively. Table I summarizes all other parameters; our results corroborated the measurements taken by Wang *et al.* (1999). Averaged 6 dB bandwidth of 14 whistles was 9.51 Hz (s.d.=8.0).

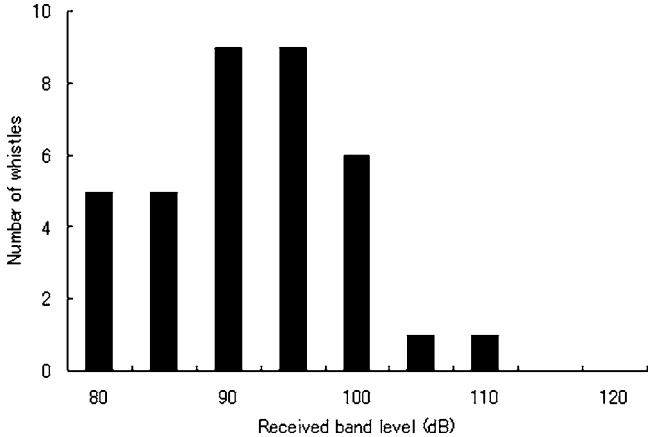


FIG. 2. Received band levels (dB rms re 1 μPa) within the analyzed bandwidth of 43 Hz. In most cases, the received band level was less than 105 dB. Each bin of 105–110 and 110–115 dB bin has only one example, respectively.

TABLE II. Acoustic characteristics of six measurement of baiji's whistle. RL: received band level dB (bandwidth=43 Hz)), SL: source level dB rms re 1 μPa within the 43 Hz bandwidth.

YY/MM/DD	HH:MM:SS	Duration (s)	Distance (m)	RL	SL
1996/1/29	9:59:21	1.02	100	101.5	141.5
1996/1/29	9:59:24	1.18	100	110.8	150.8
1996/1/30	15:45:30	1.05	200	101.0	147.0
1996/1/30	16:20:14	1.22	250	89.5	137.5
1996/1/30	16:33:01	1.27	150	92.4	135.9
1996/1/30	16:35:18	1.21	150	102.8	146.3

The seminatural reserve provided ideal conditions for measuring underwater sound during this experiment; it had little wind and the water surface was completely flat. These conditions allowed wakes to be observed in the water even after the animal had submerged. Received band levels within 43 Hz analysis band ranged from 72.3 to 110.8 dB rms re 1 μPa (Fig. 2), and the average band level was 93.4 dB rms re 1 μPa (s.d.=8.27). Six whistles were associated with distance estimated by a visual observer (Table II); assuming a spherical sound propagation, the source level was calculated at 143.2 dB rms re 1 μPa (s.d.=5.8). At this study site, the background noise level was 66.8 dB re 1 μPa within the 43 Hz band at around 5.7 Hz.

The baiji whistled several times during each recording session. Figure 3 shows the distribution of phonation intervals. Most phonation intervals were shorter than 460 s; the average phonation interval, including the two longer intervals, was 205 s (s.d.=254). One of the 12 recording sessions contained only 1 whistle and was excluded from the phonation interval analysis.

IV. DISCUSSION

The data set presented here corroborates the findings of Wang *et al.* (1999). The upswept tonal nature of baiji whistles observed in this study also supports previous research using baiji vocalizations (Jing *et al.*, 1981; Xiao and Jing, 1989; Wang *et al.*, 1995). However, none of these previous studies presented baiji whistle source levels or phonating intervals. All whistle recordings collected in this study have been digitized and securely stored in laboratories both at the Institute of Hydrobiology, the Chinese Academy of

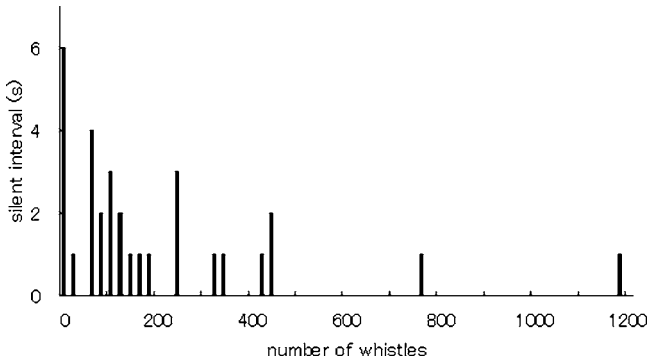


FIG. 3. Phonation interval of a baiji's successive whistle; the average interval was 205 s. The baiji's phonation was sporadic.

Sciences, China, and at the National Research Institute of Fisheries Engineering, Japan. Even though the sample size was limited (one subject and 43 whistles), this is a unique database of free-ranging baiji whistles.

The 43 Hz bandwidth employed in this study was the practical choice to calculate the sound pressure levels. The present band level is close to the overall sound pressure level of the baiji's whistle at the maximum amplitude. The 43 Hz band covered the whistle frequency band (9.51 Hz) and excluded noise components outside the band. Received band level was 23–44 dB higher than the background noise level in the band (Table II). Therefore, most of the energy measured in the band was contributed by the whistle. Additionally, short analysis window is suitable for fast online computing aboard a survey vessel to compare the band level and detection threshold level, which should be determined by the noise level of each section of the river.

Using the estimated source level of 143.2 dB and a background noise level of 66.8 dB re 1 μPa, the signal-to-noise ratio at 1 m away from the animal is 76.4 dB, equivalent to transmission loss over 6600 m assuming spherical propagation. This is the theoretical maximum detection range of a baiji's whistle in an extremely calm environment of the present study site. In a practical situation, such as ship traffic and ambient noise caused by waves, noise spectrum levels will raise to approximately 80 and 100 dB re 1 μPa^2/Hz, respectively, even at 5 kHz (Wenz, 1962; Richardson *et al.*, 1995). This can be converted to 96.3–116.3 dB in 43 Hz band. In these conservative circumstances, the signal-to-noise ratio would be 26.92–46.9 dB. Therefore, in practice, the baiji's whistle detection range could be estimated between 22 and 220 m depending on background noise levels.

We have limited information of underwater behavior of the focal animal. Bahl *et al.* (2003) calculated the swimming depth of the present animal as 0 m (surface) to 15.3 m at maximum. They used broadband recording data by the primary recording system and calculated time-arrival difference of high frequency clicks and water surface echo. Lloyd mirror effect occurs $r < 4\pi dH/L$ (Etter, 1996), where $r = 100-250$ m (direct distance between the hydrophone and the animal), $d = 2$ m (depth of the hydrophone), $H = 0-15.3$ m (depth of sound source), $L = 0.26$ m (wavelength at 5.7 kHz). The right side of the inequality ranges from 0 to 1478 m. This calculation indicates the mirror effect potentially occurred in the present data. In case of constructive interference, the detection range would be double (+6 dB error). When the interference is destructive, the detection range could drop to nearly zero.

The phonation interval averaged 205 s; baiji phonation was sporadic (Fig. 3), unlike the ceaseless phonation of finless porpoise sonar signals (Akamatsu *et al.*, 2005). A monitoring vessel should be operated slowly, otherwise it fails to receive the signal during silence. Additionally, all operating vessels produce continuous noise, which could mask a baiji whistle's frequency band. Towing a hydrophone from a vessel also creates water flow noise. Therefore, allowing a vessel to drift along a river's stream current is a better method for acoustically surveying baiji.

A passive acoustic survey from a stationary system seems to be more effective than a recording system on a moving vessel; noise levels should be much lower than in a towing system. Under circumstances of low noise levels, a hydrophone array system can identify a sound source direction or position, e.g., a stand-alone stereo recording system operated over 120 h (Ichikawa *et al.*, 2006) is a good candidate for monitoring whistles over a prolonged period in a baiji hotspot. Monitoring baiji proximity using a stationary recording system could help to assess an animal's habitat use.

It should be noted, however, that there is heavy ship traffic (9.4 ships/h in the Yangtze River's middle reach and 101.8 ships/h in its lower reach; Chen *et al.*, 1997). Masking due to ship noise potentially reduces the detection range of any stationary monitoring system.

ACKNOWLEDGMENTS

X. Zhang, Q. Zhao, Z. Wei, X. Wang, B. Yu (IHCAS), Y. Naito (NIPR), Institute of Hydrobiology, the Chinese Academy of Sciences, Field Station of Tian-e-zhou Baiji Nature Reserve, National Institute of Polar Research of Japan, greatly supported our experiments. This research was supported by Bio-oriented Technology Research Advancement Institution (BRAIN) Promotion of Basic Research Activities for Innovative Biosciences, the Chinese Academy of Sciences (CAS) (No. KSCX2-SW-118), the Institute of Hydrobiology, CAS (220103).

Akamatsu, T., Wang, D., Wang, K., and Naito, Y. (**2005**). "Biosonar behavior of free-ranging porpoises," Proc. R. Soc. London, Ser. B **272**, 797–801.

Akamatsu, T., Wang, D., Wang, K., and Wei, Z. (**2001**). "Comparison between visual and passive acoustic detection of finless porpoises in the Yangtze River, China," J. Acoust. Soc. Am. **109**, 1723–1727.

Bahl, R., Yano, M., Ura, T., Akamatsu, T., Wang, D., and Wang, K. (**2003**). "Single-hydrophone method for reconstructing dynamic behavior of endangered Chinese river dolphin," Proc. Oceans **2003**, 473–477.

Chen, P., Liu, R., Wang, D., and Zhang, X. (**1997**). *Biology, Rearing and Conservation of Baiji* (Science, Beijing, pp. 1–252.

Dalton, R. (**2006**). "Last hope for river dolphins," Nature (London) **440**, 1096–1097.

Desharnais, F., and Hay, A. E. (**2004**). "Overview of the 2003 Workshop on detection and localization of marine mammals using passive acoustics," Can. Acoust. **32**, 9–16.

Etter, P. C. (**1996**), *Underwater Acoustic Modeling*, 2nd ed. (E & FN Spon), pp. 57–58.

Ichikawa, K., Tsutsumi, C., Arai, N., Akamatsu, T., Shinke, T., Hara, T., and Adulyanukosol, K. (**2006**). "Dugong (*Dugong dugon*) vocalization patterns recorded by automatic underwater sound monitoring systems," J. Acoust. Soc. Am. **119**, 3726–3733.

Janik, V. M., Van Parijs, S. M., and Thompson, P. M. (**2000**). "A two-dimensional acoustic localization system for marine mammals," Marine Mammal Sci. **16**, 437–447.

Jing, X., Xiao, Y., and Jing, Y. (**1981**). "Acoustic signals and acoustic behavior of the Chinese rive dolphin *Lipotes vexillifer*," Sci. Sin. **24**, 407–415.

Lammers, M. O., and Au, W. W. L. (**2003**). "Directionality in the whistles of Hawaiian spinner dolphins (*Stenella longirostris*): A signal feature to cue direction of movement?," Marine Mammal Sci. **19**, 249–264.

Miller, P. J. O., and Tyack, P. L. (**1998**). "A small towed beamforming array to identify vocalizing resident killer whales (*Orcinus orca*) concurrent with focal behavioral observation," Deep-Sea Res., Part II **45**, 1389–1405.

Richardson, W. J., Greene, C. R., Jr., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego), p. 576.

Urick, R. J. (**1983**). *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill,

New York), pp. 1–423.
Wang, D., Wang, K., Akamatsu, T., and Fujita, K. (**1999**). "Study on whistles of the Chinese river dolphin or baiji (*Lipotes vexillifer*)," Oceanologia et Limnologia Sinica **30**, 349–354.
Wang, D., Wenxiang, L., and Zhifan, W. (**1995**). "A preliminary study of the acoustic behavior of the baiji, *Lipotes vexillifer*," in *Biology and Conservation of the River Dolphins*, edited by W. R. Perrin *et al.* (IUCN Special Survival Commission, Gland, Switzerland), pp. 137–140.
Wang, K., Wang, D., Akamatsu, T., Li, S., and Xiao, J. (**2005**). "A passive acoustical monitoring method applied to observation and group size estimation of finless porpoises," J. Acoust. Soc. Am. **118**, 1180–1185.
Wei, Z., Wang, D., Kuang, X., Wang, K., Wang, X., Xiao, J., Zhao, Q., and Zhang, X. (**2002**). "Observations on behavior and ecology of the Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientalis*) group at Tian-e-Zhou Oxbow of the Yangtze River," The Raffles Bulletin of Zoology, Supplement **10**, 97–103.
Wenz, G. M. (**1962**). "Acoustic ambient noise in the ocean: Spectra and sources," J. Acoust. Soc. Am. **34**, 1936–1956.
Xiao, Y., and Jing, R. (**1989**). "Underwater acoustic signals of the baiji, Lipotes vexillifer," in *Biology and Conservation of the River Dolphins*, Occasional paper, edited by W. R. Perrin *et al.*, IUCN special survival commission 3, Gland, Switzerland. pp. 137–140.

An international key comparison of free-field hydrophone calibrations in the frequency range 1 to 500 kHz

Stephen P. Robinson, Peter M. Harris, Justin Ablitt, Gary Hayman, and Alex Thompson
National Physical Laboratory, Teddington, Middlesex, TW11 0LW, United Kingdom

A. Lee van Buren, Joseph F. Zalesak, and Robert M. Drake
USRD (Code 151), Naval Undersea Warfare Center, 1176 Howell Street, Newport, Rhode Island 02841-1708

Alexander E. Isaev and Alexander M. Enyakov
VNIIFTRI, Mendeleevo, Moscow Region 141570, Russia

Christopher Purcell
Defence Research and Development Canada, 9 Grove Street, Dartmouth, Nova Scotia B2Y 3Z7, Canada

Zhu Houqing, Wang Yuebing, and Zhang Yue
Acoustics Laboratory, National Institute of Metrology, 18 Bei San Huan Dong Lu, Beijing 100013, China

Pierre Botha
CSIR—National Metrology Laboratory, P.O. Box 395, 0001 Pretoria, South Africa

Dieter Krüger
Wehrtechnische Dienstelle für Schiffe und Marinewaffen (WTD71), Berliner Strasse 115, 24340 Eckernförde, Germany

(Received 19 September 2005; revised 24 May 2006; accepted 24 June 2006)

A description is given of the results of a Key Comparison of primary free-field standards for underwater acoustics at frequencies from 1 to 500 kHz. This is the first such Key Comparison exercise in the field of underwater acoustic calibration and measurement. Laboratories from UK, Germany, USA, Russia, China, Canada, and South Africa participated by calibrating three reference hydrophones, with project coordination provided by the National Physical Laboratory, UK. The agreement between the results obtained from the comparison was generally encouraging, with the calibration values reported by the laboratories agreeing within quoted uncertainties over the majority of the frequency range, and the results generally lying within a ±0.5-dB band for frequencies up to 300 kHz. A discussion is given of the general sources of uncertainties in the calibrations, in particular those which are thought to have contributed to the differences in the results between laboratories. The results of the participants have been used to estimate the equivalence of national measurement standards within this field. [DOI: 10.1121/1.2228790]

PACS number(s): 43.30.Xm, 43.58.Vb, 43.30.Yj [KGF] Pages: 1366–1373

I. INTRODUCTION

International trade requires a secure technical foundation. To avoid disagreement between the measurements made in separate countries, there must be demonstrable equivalence between the measurement standards maintained by the individual countries. One method of determining the equivalence is to undertake an international comparison of measurements and calibrations. These exercises often indicate where improvements are required and the results may represent progress toward a more accurate realization of SI units.

In 1999, a major development in this field was the ratification of the Mutual Recognition Arrangement (MRA), which was prepared by the Comité International des Poids et Mesures (CIPM).[1] All member states of the Metre Convention are signatories to the MRA, which has the objective of providing for the mutual recognition of calibration certificates. This in turn requires an evaluation of the degree of equivalence of national measurement standards maintained by individual national metrology institutes, and the mechanism for determining this equivalence is through international comparisons. Such comparisons may be at a world level, in which case they are known as Key Comparisons and are organized under the auspices of the relevant consultative committee of the CIPM. The participants in a Key Comparison will typically include at least one representative from each regional metrology organization (RMO). Comparisons may also be organized within a region and fall within the auspices of the relevant RMO, such as EUROMET in Europe, or SIM in the Americas. These regional comparisons may then be linked to the Key Comparison via the common participant(s).

With the formation in 1999 of the Consultative Committee on Acoustics, Ultrasound, and Vibration (CCAUV) of the CIPM, a number of comparisons were commenced to fulfill the requirements of the MRA. For acoustics in air, there has been considerable activity to compare the calibrations of

standard microphones. This work has been undertaken using Key Comparisons and regional comparisons, and this has already been extensively reported in the literature.[2–6]

For water-borne acoustics, the infrastructure for metrology is perhaps less mature than for air acoustics. However, at megahertz frequencies a Key Comparison of hydrophone comparisons has recently been completed,[7] this activity following on from less formal comparisons undertaken some years earlier.[8,9] In the field of underwater acoustics at frequencies less than 1 MHz, there have been few comparisons of hydrophone calibrations organized in the past. The first significant attempt was made by Trott in 1968, who organized a pioneering worldwide comparison.[10] Although other regional comparisons have been undertaken in the meantime,[11–14] until now there has been no attempt to repeat the exercise on a worldwide scale.

This paper describes the first Key Comparison for the primary free-field standards for sound in water at frequencies between 1 and 500 kHz (designated comparison identifier: CCAUV.W-K1).[15] The exercise involved participants in seven countries each calibrating the same three hydrophones. The goal of the comparison was to compare the measurement standards of the participating nations, and express their equivalence in terms of the difference between their results and the uncertainties on those differences. To achieve this, the results of the participants have been used to calculate a reference value termed the Key Comparison reference value (KCRV) at each acoustic frequency. The degree of equivalence of national measurement standards has then been calculated from the differences of the participants' results from the KCRV.

II. THE COMPARISON

A. Organization of the comparison

The comparison had seven participating countries, each represented either by the respective National Metrology Institute (NMI), or by an organization officially designated as representing the country for this exercise. For Canada, Defence Research and Development Canada (DRDC) was nominated by the Institute for National Measurement Standards of the National Research Council; for Germany, Wehrtechnische Dienstelle für Schiffe und Marinewaffen (WTD) was nominated by the Physikalisch-Technische Budesanstalt; and for the USA, the Underwater Sound Reference Division (USRD) of the Naval Undersea Warfare Center (NUWC) was nominated by the National Institute for Standards and Technology. These were joined by the National Institute for Metrology (NIM) from China, the All Russian Institute of Physical-Technical and Radio-Technical Measurements (VNIIFTRI), the Council for Scientific and Industrial Research from South Africa (CSIR), and the National Physical Laboratory from the UK (NPL). The participants are listed in Table I along with their country and relevant regional metrology organization (RMO). These are EUROMET (the metrology organization for Europe), APMP (Asia Pacific Metrology Programme), COOMET (Coopération Métrologique), SADCMET (South African Development Community Cooperation in Measurement Traceability), and SIM (Sistema Interamericano de Metrologia). Also listed in Table I are the country codes used to identify the results in the graphical representations later in this paper. The full contact details of each participant are listed with the author affiliations.

TABLE I. Participants in the comparison.

Institute	Country	Country code	Regional metrology organization (RMO)
DRDC	Canada	CA	SIM
NIM	China	CN	APMP
WTD	Germany	DE	EUROMET
VNIIFTRI	Russia	RU	COOMET
CSIR	South Africa	ZA	SADCMET
NPL	United Kingdom	UK	EUROMET
USRD/NUWC	United States	US	SIM

The pilot laboratory for the project was NPL, which undertook the initial assessment and calibration of the hydrophones, prepared the protocol document for the comparison, and performed checks on the hydrophone sensitivities between the calibrations by participants to ensure that the hydrophone sensitivities were stable. The hydrophones were circulated to the participants in a round-robin fashion, with the devices being returned to NPL after each set of calibrations by the participants.

B. The hydrophones

The hydrophones chosen for the comparison were an H52 hydrophone manufactured by USRD-NUWC in the USA; a B&K8104 hydrophone manufactured by Brüel & Kjær in Denmark; and a TC4034 hydrophone manufactured by Reson in Denmark. The devices used for the calibration are listed in Table II along with the frequency over which calibrations were undertaken. Each participant calibrated the three hydrophones at approximately 40 discrete acoustic frequencies for each hydrophone in the range 1 to 500 kHz. These hydrophones were chosen because they were routinely in use as standard measuring hydrophones within laboratories.

TABLE II. Details of the three hydrophones used in the comparison.

Hydrophone type	Manufacturer	Frequency range (kHz)	Nominal sensitivity (1 kHz) (dB re: 1 V/μPa)	Integral preamplifier
H52	USRD	1–100	−177	Yes
8104	Brüel & Kjær	10–150	−205	No
4034	Reason	100–500	−218	No

TABLE III. The range of overall standard uncertainties quoted by participants (expressed as relative uncertainties in percent). Note that the high value for participant ZA is only applicable for the highest frequencies close to 500 kHz.

UK	DE	US	RU	CN	CA	ZA
2.5–3.9	5.0	2.0–4.6	2.2–3.5	2.7–4.1	3.6–6.3	5.9–28.8

C. Calibration methods

Participants were asked to perform an absolute measurement of the end-of-cable free-field open-circuit sensitivity for each hydrophone using their own in-house methods and procedures for the calibrations. The "left-hand" *XYZ* coordinate system suggested by IEC 565 was adopted,[16] with each participant asked to align the hydrophones such that an alignment mark on the hydrophone body pointed in a direction parallel to the direction of propagation of the incoming acoustic wave. NPL prepared and circulated a protocol document describing the measurements required.

The method of calibration used by participants was the method of three-transducer spherical-wave reciprocity. By use of this method, which is described in IEC 565,[16] a hydrophone may be calibrated absolutely by making purely electrical measurements. Most commonly, participants used laboratory tank facilities of varying sizes, the largest being $15\times7.5\times7$ m and the smallest dimension of any of the test tanks used being 4.5 m. All of the tanks had a framework or traversing system used for mounting and positioning the transducers. One participant used an open-water facility on a lake, which had a water depth of 11 m, a laboratory platform being created using a pier or pontoon-based structure from which transducers may be lowered into the water. For all participants, discrete-frequency tone-burst signals were employed, with reflections isolated from the direct-path signal by use of gating and time-windowing techniques.

D. Uncertainties

Each participant was requested to provide a value for the overall uncertainty for the calibrations assessed according to the ISO document *Guide to the Expression of Uncertainty in Measurement*.[17] These are summarized in Table III. In addition, each participant was requested to provide a breakdown of their uncertainties in two categories: type A components which are estimated by statistical methods through repeated calibrations (sometimes termed the repeatability or "random" uncertainty); and type B components which cannot be estimated by repeated calibrations (often these components can be considered to express any systematic "bias" in the calibrations). Since each participant had used a slightly different implementation of the reciprocity method, the sources of type B uncertainty and the values of the individual components varied. In general, the values quoted varied with frequency, the full range of uncertainty values being shown in the table.

TABLE IV. Variations in the results of check calibrations performed by NPL for the hydrophones. The type A uncertainty is the repeatability of the results expressed for a confidence level of 95%.

Frequency (kHz)	Overall mean (dB re: 1 V/μPa)	Maximum deviation (dB)	Minimum deviation (dB)	Mean type A uncertainty (dB)
	H52 hydrophone			
5	−177.43	0.14	−0.18	0.21
10	−177.94	0.10	−0.20	0.19
20	−178.28	0.12	−0.24	0.19
50	−177.66	0.19	−0.23	0.21
100	−179.64	0.24	−0.18	0.22
	B&K 8104 hydrophone			
10	−206.31	0.10	−0.14	0.25
20	−206.68	0.18	−0.21	0.25
50	−203.65	0.20	−0.22	0.27
100	−211.31	0.18	−0.16	0.21
150	−218.71	0.23	−0.24	0.22
	TC4034 hydrophone			
100	−218.71	0.12	−0.22	0.21
200	−218.36	0.19	−0.30	0.21
300	−216.04	0.20	−0.22	0.21
400	−219.26	0.27	−0.27	0.25
500	−230.00	0.32	−0.37	0.31

III. RESULTS

A. Hydrophone stability

As coordinating laboratory, NPL undertook "check calibrations" on each hydrophone in between the calibrations of the other participants in order to monitor the stability of the hydrophones. The check calibrations were undertaken using the same procedure for full free-field calibrations of the hydrophones at NPL, but with measurements made only at selected frequencies (to save time). For each hydrophone, the results of the eight check calibrations are presented in tabular form in Table IV showing the maximum and minimum deviation from the mean occurring in the results of the check calibrations, and the mean repeatability (type A uncertainty) for the NPL calibrations. The typical type A standard uncertainty for the check calibrations was of the order of between 0.5% and 1%, calculated from at least four repeated calibrations. To express this for a confidence level of 95%, the standard uncertainty was multiplied by a coverage factor derived from the Student's *t*-factor for the appropriate degrees of freedom. The mean value for the uncertainty of all the check measurements was calculated and is shown in the table, expressed in decibels. This allows some judgment to be made regarding the significance of the variations observed in the check calibrations.

The results of the check calibrations showed that the hydrophones may be considered stable for the purposes of the comparison exercise. The maximum variation in the check calibrations at each frequency was generally within or of the same order as the repeatability of the NPL measurements, and although the deviation slightly exceeded the uncertainty at some of the frequencies, this was not considered significant. There was some evidence that there may have been a gradual increase in the sensitivity of the H52 of

0.01 dB per month during the comparison, but no corrections were applied since this was considered a marginal variation. A comprehensive description of all of the checks on stability is given in the Key Comparison Final Report.[15]

B. Analysis methodology

To facilitate a comparison between participants, the results provided by the participants were used to derive a reference value that provides a type of "overall grand mean" of results. This reference value was termed the Key Comparison reference value (KCRV). This was calculated at each acoustic frequency using a weighted mean approach.[18] In this approach, the results of the individual participants are weighted according to the inverse of the square of the stated uncertainty. For the results, x_i, from a given device with associated uncertainties, u_i, where i is the index for a particular laboratory, w_i is the weight applied to the participant results, and $i=1,\ldots,N$, then the weighted mean, y, is evaluated from

$$y = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}, \quad w_i = \frac{1}{u_i^2}, \tag{1}$$

with the associated uncertainty on the weighted mean, $u(y)$, determined from

$$\frac{1}{u^2(y)} = \sum_{i=1}^{N} \frac{1}{u_i^2}. \tag{2}$$

This analysis was applied to the results at each frequency for each hydrophone to derive the KCRVs for each device. For this comparison, the actual values of the KCRVs have little inherent value in themselves, being merely the sensitivities of some arbitrarily chosen hydrophones. Their value is purely in their role in evaluating the degrees of equivalence.

The weighted mean has the advantage that it makes use of all of the data provided by the participants, including the quoted uncertainties as well as the sensitivity values. A test was performed to determine whether the observed value of the chi-squared statistic was significant at the 95% confidence level. The test demonstrated overall consistency of the data with the weighted mean model, demonstrating that the weighted mean was an acceptable model to use for this data set. The test was passed for all except 11 of the 117 frequency points measured, with five of the failures occurring at the five highest frequencies of calibration for the H52 hydrophone. In this range of 80 to 100 kHz, the hydrophone is being used at the limit of its useful range, and this led to exclusion of these results from the final analysis. For comparison, the KCRVs were also calculated using two other estimators: the median and the unweighted mean.[15] The results showed close agreement, with the median and unweighted mean agreeing with the weighted mean to within the uncertainties of the weighted mean at all frequencies for all hydrophones. This shows that for the data in this comparison, the values of KCRV were not highly sensitive to the choice of reference value estimator.[15] None of the results provided by participants were classified as outliers.

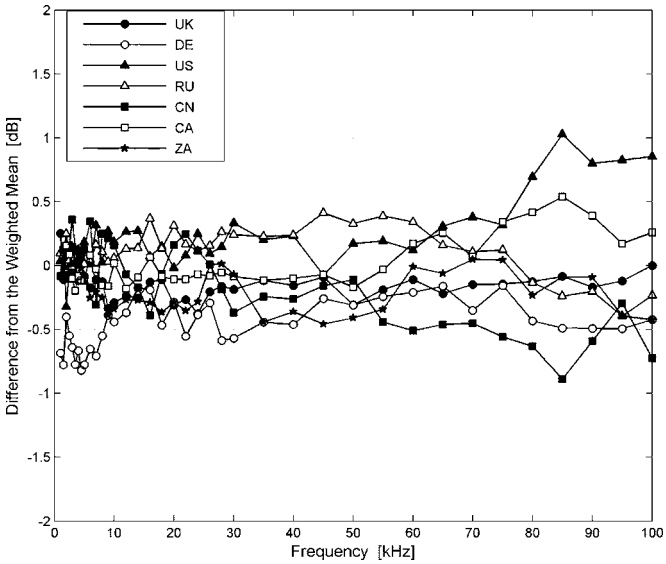


FIG. 1. The differences from the KCRV (weighted mean) for the H52 hydrophone.

C. Differences from the reference values

The results of the calibrations are presented in Fig. 1, Fig. 2, and Fig. 3 as differences from the KCRV for each hydrophone. The agreement between the results was generally encouraging, with the calibration values reported by the laboratories agreeing within quoted uncertainties over the majority of the frequency range, and the results generally lying within a ±0.5-dB band for frequencies up to 300 kHz, a factor of 2 improvement on the spread of results obtained in the 1998 EUROMET comparison.[13]

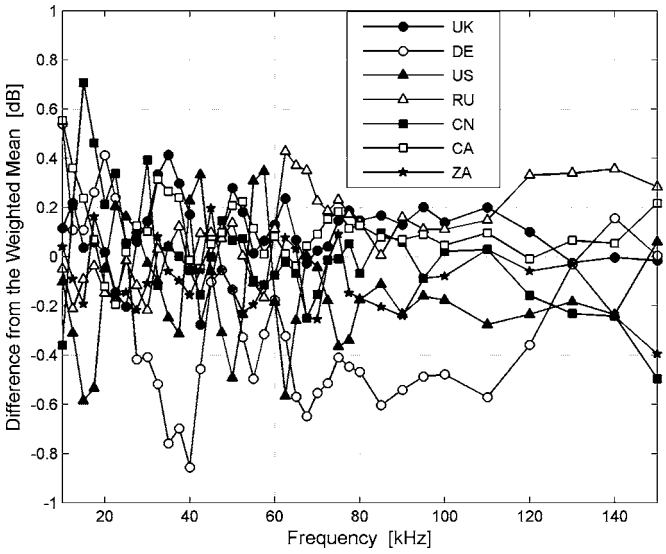


FIG. 2. The differences from the KCRV (weighted mean) for the B&K8104 hydrophone.

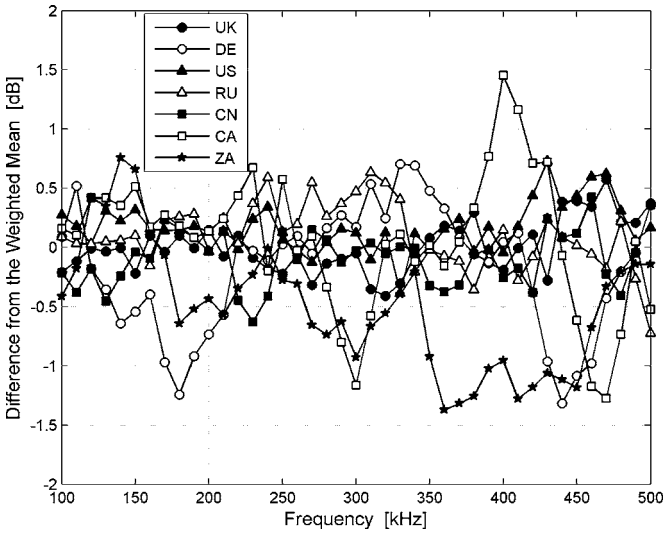


FIG. 3. The differences from the KCRV (weighted mean) for the TC4034 hydrophone.

D. Equivalence between participants

To estimate the equivalence between the calibrations of different laboratories requires that some measure of the difference in their results be calculated. This was done by calculating the degree of equivalence for each participant. For the purposes of this comparison, the term *degree of equivalence* of measurement standards is taken to mean the degree to which a standard is consistent with the reference value. The degree of equivalence of each national measurement standard is expressed quantitatively by two terms: its deviation from the reference value, and the uncertainty of this deviation (at the 95% level of confidence).

The degrees of equivalence (DoEs) of laboratory i were evaluated from

$$d_i = x_i - y, \tag{3}$$

where x_i and y are defined in Eqs. (1) and (2). The associated standard uncertainty $u(d_i)$ is determined from

$$u^2(d_i) = u^2(x_i) - u^2(y). \tag{4}$$

For frequencies at which more than one hydrophone was calibrated, an estimate of the DoE was available for each of the hydrophones used. Of the 94 distinct frequencies of measurement covered by the three devices, a calibration of more than one device had been undertaken at only 18 of these frequencies (less than 20% of the frequency points). It was necessary to combine these multiple estimates to provide a single value of the degree of equivalence at each frequency point.

To do this, values of the relative DoE were used since the reference values for each hydrophone varied appreciably in magnitude from hydrophone to hydrophone. For example, at 100 kHz the reference value for the H52 hydrophone was 1315.7 μV/Pa, whereas for the B&K8104 it was 48.2 μV/Pa, and for the TC4034 it was 11.8 μV/Pa. The relative DoE, r_{ij}, is therefore expressed as $r_{ij}=d_{ij}/y_j$, where i denotes the index of the laboratory and j the index of the hydrophone. The method used to perform the combination was to calculate the weighted mean r_i of the estimates r_{ij} with allowance made for the correlation between those estimates arising from mutual dependencies between the measurements made of the different hydrophones by the same laboratory. The estimate r_i was obtained as the solution to the generalized least-squares problem,

$$\min_{r_i}(\mathbf{r}_i - Ar_i)^T V_i^{-1}(\mathbf{r}_i - Ar_i), \tag{5}$$

in which, for the case of two hydrophones, $\mathbf{r}_i=(r_{i1}, r_{i2})^T$ contains the estimates for the hydrophones with associated uncertainty (covariance) matrix V_i of size 2×2, and $A=(1,1)^T$. The uncertainty budgets provided by each laboratory were used to identify those components of uncertainty that described systematic effects that were common to its calibration of the different hydrophones. These uncertainties were then used to evaluate the covariances associated with the measurements (and, hence, to derive the off-diagonal elements of covariance matrix V_i).

Figure 4 shows the combined DOEs for 8 selected frequencies out of a total of 94 frequencies in the range 1 to 500 kHz, namely: 3, 10, 50, 100, 200, and 350 kHz. Note that Canada did not undertake calibrations at frequencies less than 2 kHz, and South Africa did not undertake calibrations at frequencies less than 3 kHz, so it was not possible to calculate DOEs for these participants at frequencies below these limits.

IV. DISCUSSION

A number of factors may influence the results of calibrations and contribute to the variations in results between participants, either because of unavoidable differences in environmental conditions, or because of differences in the procedures used by the participants. Potential influences include: poor alignment, lack of acoustic far-field conditions, the length of wetting and soaking time for the hydrophones, the lack of steady-state conditions, interference from boundary reflections, the influence of noise, and electrical loading by cables/amplifiers. Other influences include the effect of the mounting or rigging used with the hydrophones, and the variation of water temperature and depth of immersion between participant calibrations.

Some of the influences listed above are related to the calibration methodology and were controlled by specifying the procedure for measurements very carefully in the protocol document. Each participant produced an uncertainty budget for the calibrations that included estimates for the uncertainty contributions due to these influences. However, some of the influences listed relate to environmental effects.

It is known that the mounting configuration used can affect the measured sensitivity for some hydrophones, and this may contribute to the variation in results. Ideally, the mount should not cause any reflections or reverberations in the acoustic signal, but should be rigid enough to allow precise positioning of the hydrophones. Fortunately, there was a general similarity between the mounting arrangements used by participants, with many using some form of free-flooded tube made of metal or plastic. However, the variations in the mounting almost certainly contributed to the discrepancies

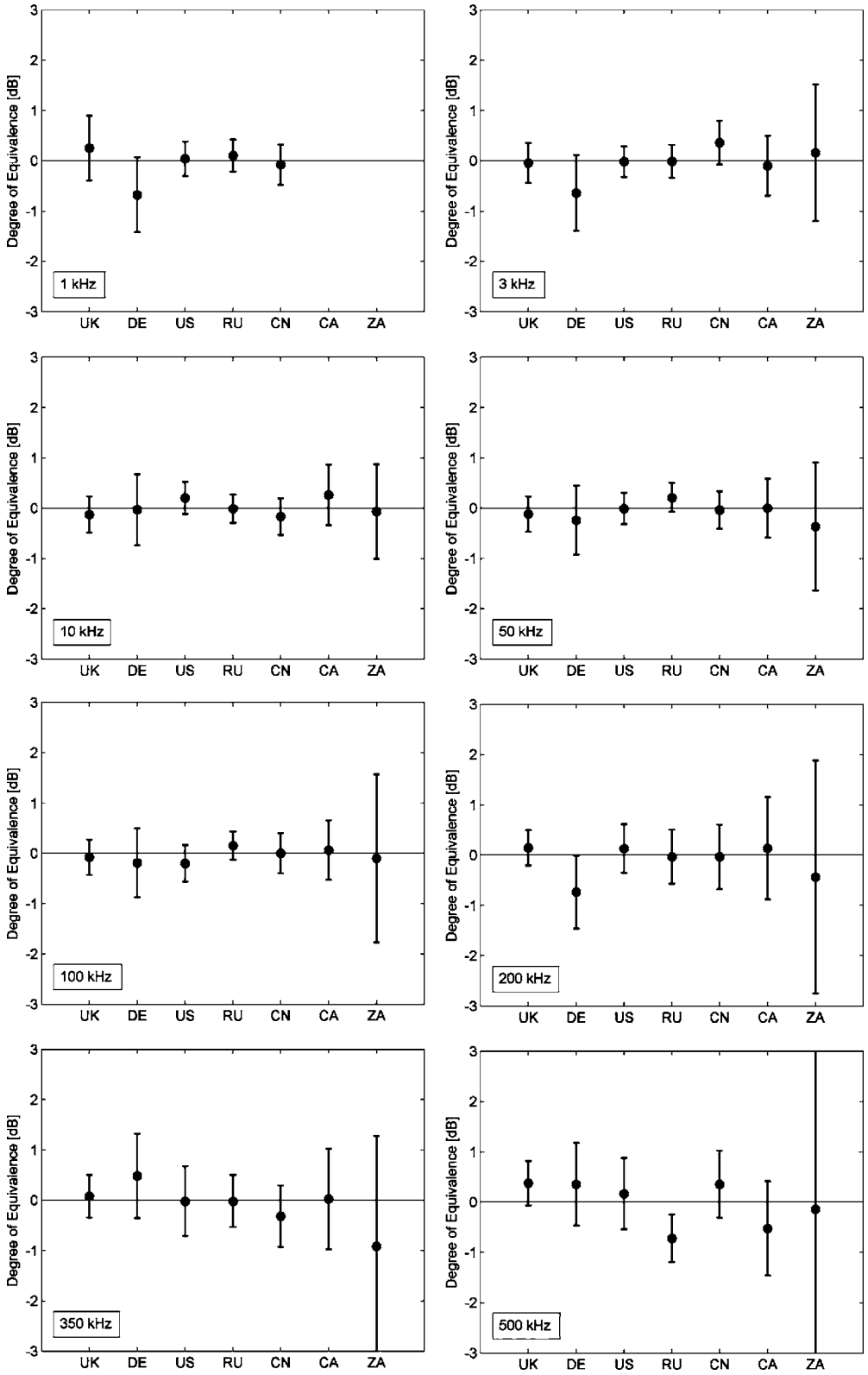


FIG. 4. Degrees of equivalence for frequencies of 1, 3, 10, 50, 100, 200, 350, and 500 kHz. Values are stated in dB and expanded uncertainties are expressed for a confidence level of 95%.

observed between the results of participants. In future comparisons, it may be better to prescribe the type of mount to be used, or perhaps circulate the mounts with the hydrophones under test.

There is a general trend for increasing spread in the results as the frequency is increased from 200 to 500 kHz. This was expected and is mostly likely due to the increased difficulty of undertaking free-field calibrations at higher frequencies. This occurs because the devices become more directional, leading to more difficulties with alignment; in addition, devices used beyond their main resonance can become less reciprocal in their behavior, a necessary condi-

tion of the calibration method; finally, noise can become issue if the transmitting response of the device rolls off at very high frequencies.

For this comparison, the depths of immersion during calibrations by different participants ranged between 1.8 and 4.0 m and the water temperatures ranged from 14.0 °C to 21.1 °C. For the range of depths employed by the participants in this comparison, it is highly unlikely that the depth variation has significantly influenced the results.

However, the water temperature is likely to have affected the results and contributed to some of the discrepancies observed. The majority of the measurements by participants were undertaken within a ±2 °C range centered around 19.0 °C, with the exception of two participants who undertook a portion of their measurements at lower temperatures, leading to an extreme case of a ±3.5 °C variation in the temperature (for a limited portion of the data on the hydrophones). Some data were available on the typical variation in response of the hydrophone models used here from measurements made at NPL[19] and at USRD,[20] and extra measurements were made at NPL during this work.[15] This extensive evaluation showed that the TC4034 and the H52 hydrophones were remarkably stable. For these hydrophones, even taking the most extreme variation in water temperature, the worst case variation in the responses of the hydrophones at any frequency was only 0.2 dB for the H52 hydrophone, and 0.25 dB for the TC4034 hydrophone. However, the results for the variation in response of the B&K8104 with temperature showed less stability. The results were highly complex, depending strongly on the acoustic frequency, and showing some evidence of nonlinear behavior with temperature. For this hydrophone, taking the most extreme variation in water temperature, the worst case variation of response at any frequency was 0.5 dB. However, this a pessimistic estimate based on a worst case—for the majority of the results the effect is likely to be much smaller. The participants' calibration data were not corrected for temperature because the participants felt that the variation of sensitivity with temperature was not known accurately enough, this being the first time such data have been measured as a function of frequency. In future comparisons of this type, it is recommended that particular attention be given to both controlling the temperature during calibrations (so that it falls within a narrow range for all participants), and to choosing hydrophones which are stable with temperature. The results of the work to measure the variation in response with temperature have been published elsewhere[15,19,20] and may be the subject of further publications.

Although there was some general agreement in the value of the overall uncertainty, the values ascribed to the components of uncertainty within the uncertainty budgets of participants varied significantly. In the view of the participants, further work to identify and quantify these uncertainties is justified.

V. CONCLUSION

The comparison has been extremely valuable, with much confidence gained in the performance of primary standards. This is the first such Key Comparison exercise in the field of underwater acoustic calibration and measurement, with expert laboratories from UK, Germany, USA, Russia, China, Canada, and South Africa participating. The results obtained represent the current state of the art for hydrophone calibrations in the frequency range 1 to 500 kHz. The calibration values reported by the laboratories agreed within quoted uncertainties over the majority of the frequency range, and the results generally lay within a ±0.5-dB band for frequencies up to 300 kHz, a factor of 2 improvement on the spread of results obtained in the 1998 EUROMET comparison. At higher frequencies, the spread of results increases to almost ±1 dB at 500 kHz. The uncertainties quoted by participants were typically of the order of 0.5 dB when expressed for a confidence level of 95%, increasing somewhat at high frequencies. Although most participants quoted similar uncertainties, the values of individual components varied considerably. The effect of water temperature variation is likely to have contributed to some of the discrepancies observed, particularly for the B&K8104 hydrophone.

The results demonstrate equivalence (within the quoted uncertainties) between the measurement standards maintained by the individual countries. The results represent a significant step forward in establishing the infrastructure for metrology in underwater acoustics, a field where metrology is relatively immature compared to airborne sound. The coverage of the comparison may be extended by conducting regional comparisons which include as a participant one of the laboratories taking part in the Key Comparison. These regional comparisons may then be linked to the Key Comparison via the common participant(s).

The results available in the database on the BIPM website also include tabulated data on the bilateral degrees of equivalence between the countries along with associated uncertainties. The data generated by the comparison are available on the BIPM website at ⟨http://kcdb.bipm.org/appendixB⟩.

ACKNOWLEDGMENTS

NPL acknowledges the support of the National Measurement System Directorate of the UK Department of Trade and Industry for the financial support of this work. Of course, the comparison would not have been possible without the effort of participants, and the authors would like to express gratitude to all those who took part. In addition, the support and guidance of the CCAUV (in particular Dr. Penny Allisy-Roberts, Secretary, CCAUV) and the BIPM is gratefully acknowledged.

[1]BIPM, *Mutual Recognition of National Measurement Standards and of Calibration and Measurement Certificates Issued by National Metrology Institutes* (Bureau International des Poids et Mesures, Paris, October 1999).

[2]R. G. Barham, "A BIPM/CIPM key comparison on microphone calibration—defining the state of the art," J. Acoust. Soc. Am. **112**, 2340 (2002).

[3]R. G. Barham, "Key Comparison: Report on key comparison CCAUV.A-K1," Metrologia **40**, 9002 (2003).

[4]V. Nedzelnitsky, "Issues concerning international comparison of free-field calibrations of acoustical standards," J. Acoust. Soc. Am. **112**, 2340 (2002).

[5]G. S. K. Wong and L. Wu, "Microphone interlaboratory comparison in the Americas," J. Acoust. Soc. Am. **112**, 2340 (2002).
[6]G. S. K. Wong and L. Wu, "Interlaboratory comparison of microphone calibration," J. Acoust. Soc. Am. **115**, 680–682 (2004).
[7]B. Zeqiri and N. D. Lee, "A BIPM/CIPM key comparison covering the calibration of ultrasonic hydrophones over the frequency range 1 MHz to 15 MHz," J. Acoust. Soc. Am. **112**, 2342 (2002).
[8]R. C. Preston and S. P. Robinson, "European comparison of ultrasonic hydrophone calibrations," Metrologia **36**(4), 345–349 (1999).
[9]R. C. Preston, D. R. Bacon, S. S. Corbett III, G. R. Harris, P. A. Lewin, J. A. McGregor, W. D. O'Brien Jr, and T. L. Szabo, "Interlaboratory comparison of hydrophone calibrations," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 206–213 (1988).
[10]W. J. Trott, "International standardization in underwater sound measurements," Acustica **20**, 169–181 (1968).
[11]S. P. Robinson, R. C. Preston, and G. J. Green, "An intercomparison of hydrophone calibrations within Europe," J. Acoust. Soc. Am. **103**(4), 2755 (1998).
[12]S. P. Robinson, R. C. Preston, and G. J. Green, "An intercomparison of hydrophone calibrations within Europe," in Proceedings of the 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America, Seattle, WA, **1**, pp. 59–60 (1998).
[13]S. P. Robinson, G. J. Green, R. C. Preston, L. Peirlinckx, L. P. Kofoed, C. Skodborg, A. Roy, Y. Mori, A. Brenner, D. Krüger, S. Buogo, G. Cannelli, L. Troiano, C. Runborg, and G. Gooch, "International comparison of free-field hydrophone calibrations in the frequency range 10–315 kHz," Metrologia, **36**, 287–296 (1999).
[14]A. M. Enyakov, S. M. Likhatchev, V. A. Platonov, W. J. Yuan, Y. B. Wang, and J. Q. Li, "A Russian-Chinese international comparison of hydrophone calibration methods," Metrologia **36**, 297–303 (1999).
[15]S. P. Robinson, "Final Report for Key Comparison CCAUV.W-K1: Calibration of hydrophones in the frequency range from 1 kHz to 500 kHz," NPL Report DQL-AC 009 (2004).
[16]IEC 565:1977, "The calibration of hydrophones," (International Electrotechnical Commission, Geneva, 1977).
[17]BIPM, IEC, IFCC, ISO, IUPAC, and OIML, *Guide to the Expression of Uncertainty in Measurement*, 2nd ed. (International Organization for Standardization, Geneva, (1995).
[18]M. G. Cox, "The evaluation of key comparison data," Metrologia **39**, 589–595 (2002).
[19]G. A. Beamiss, S. P. Robinson, G. Hayman, and T. J. Esward, "Determination of the variation in free-field hydrophone response with temperature and depth," Acust. Acta Acust. 88, 799–802 (2002).
[20]A. L. Van Buren, R. M. Drake, and A. E. Paolero, "Temperature dependence of the sensitivity of hydrophone standards used in international comparisons," Metrologia, **36**(4), 281–285 (1999).

Experimental investigation of coupled vibrations in piezoelectric cylindrical shells[a)]

Boris S. Aronov, David A. Brown,[b)] and Sundar Regmi
BTech Acoustics, LLC and Acoustics Research Laboratory, Advanced Technology and Manufacturing Center, Electrical Engineering, University of Massachusetts Dartmouth, 151 Martine Street, Fall River, Massachusetts 02723

(Received 15 September 2005; revised 15 June 2006; accepted 23 June 2006)

An experimental investigation of the electromechanical properties of radially polarized thin-walled piezoelectric ceramic cylinders as a function of the height-to-diameter aspect ratio is presented. The focus of this investigation is on the lowest order axially symmetric modes of vibration for cylinders with free ends having aspect ratio less than 3 with emphasis on the region of strongest coupling around aspect ratio $\pi/2$. The resonance frequencies and coupling coefficients are presented vs aspect ratio and quantify the lower, upper, and an intermediate resonance branch. It is shown that the intermediate branch is flexural in origin and clearly crosses the so-called "dead zone," which is an inadequacy of the membrane theory applied to the vibration of thin shells [Junger and Rosato, J. Acoust. Soc. Am., **26**, 709–713 (1954)]. The coupling coefficient for the lower branch is shown to drop to zero at $h/2a=\pi/2$, whereas the coupling coefficient associated with the upper branch reaches a maximum at an aspect ratio near this region. As an example of the practical application of this study, the performance of two underwater electroacoustic transducers comprised of cylindrical elements having different aspect ratios is reported. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2228611]

PACS number(s): 43.38.Ar, 43.38.Fx, 43.38.Pf, 43.40.At, 43.40.Ey [AJZ] Pages: 1374–1380

I. INTRODUCTION

Thin-walled piezoelectric cylinders are employed in a broad range of applications and are available in a variety of length-to-diameter aspect ratios and thicknesses. A great deal of attention has been given to the analysis of the coupled longitudinal and radial vibrations in cylindrical shell elements, as the electromechanical properties depend on their dimensions. A theoretical and experimental study on ferroelectric cylindrical shells was first published by Haskins and Walsh (1957).[1] Their analytical treatment was based on the two-dimensional (membrane) theory, which had been previously established by Love[2] for finite-length elastic shells with free-boundary conditions. This work[1] still represents the most comprehensive experimental investigation published on the subject to date, although their experimental results support the physically improper conclusion inherent in the membrane theory that predicts a "dead zone," a band in which no natural resonance modes can exist. However, a prior publication by Junger and Rosato[3] showed theoretically that the apparent dead zone was a result of neglecting the flexural vibrations of the shell's wall, the contribution of which becomes especially important in the region of strongest coupling between the longitudinal and radial modes at the values of aspect ratios close to $h/2a=\pi/2$. Subsequently, many papers were devoted to different issues related to the coupled vibrations in hollow piezoelectric cylinders. Most of these publications (e.g., Refs. 4–6) remained in the framework of membrane theory and some[7,8] considered the coupled vibrations in piezoelectric cylindrical shells taking into account their finite thickness, i.e., considered flexural stresses along with membrane stresses in the shells. Nevertheless, these papers do not provide experimental data to explicitly disprove the dead zone. Moreover, the experimental data on the resonance frequencies[7] are reported to be in good agreement with the results of Haskins and Walsh. They show the same two lower and upper resonance branches for the gravest modes of vibration but do not explicitly discuss an intermediate resonant branch of flexural origin. Also, the previous literature does not present experimental or analytical results on the effective coupling coefficient and its dependence on aspect ratio. This brief review shows that, in spite of the widespread literature on coupled vibrations in cylindrical piezoelectric shells, there are still important theoretical issues related to improving our understanding of the nature of the vibrations and practical issues related to improving real devices employing cylindrical piezoelectric elements.

The objective of this paper is to address these issues by experimental investigation. Thus, in Sec. III the resonance frequencies of the lowest-order axial symmetric modes are presented for finite-height cylinders with free ends having aspect ratios up to $h/2a=3$, which represent the range for most practical applications. It is shown that, between the upper and lower frequency branches, which are adequately accounted for by the membrane theory, an intermediate branch exists, which has its origin in the class of flexural vibrations of the shell's wall. The flexural nature of this branch is confirmed by results of vibration mode shape mea-

[a)]Reported in part at the 147th meeting of the Acoustical Society of America, 115(5), Pt.2, May 2004.
[b)]Author to whom correspondence should be addressed; electronic mail: dbAcoustics@cox.net

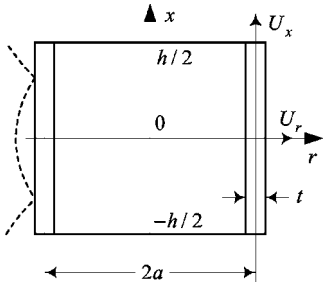


FIG. 1. Cylinder geometry and notations. An illustrative vibration mode shape (modulus) for the intermediate mode in the radial direction is shown by the dotted line.

surements. The dependence of the effective coupling coefficients vs aspect ratio corresponding to the three resonant branches is presented and discussed in another subsection. Section IV presents a case study of underwater acoustic transducers made from the piezoelectric cylindrical elements having aspect ratios $h/2a=0.54$ and $h/2a=1.1$. This study shows the effects of coupled vibration on electroacoustic performance of the transducers.

II. DESCRIPTION OF THE SAMPLES AND MEASUREMENT TECHNIQUES

The majority of results was obtained with PZT-4 ceramic cylinders having mean diameter $2a=35$ mm, wall thickness $t=3.2$ mm, and various height-to-mean-diameter ratios (aspect ratios) $h/2a$ in the range of 0.2 to 3.0. The geometry of the cylinders is shown in Fig. 1. A supply of 19 mm tall cylinders was cut and bonded as needed to obtain various heights. In addition, many measurements were made on PZT-5 cylinders of diameter $2a=11.3$ mm and $t=1.2$ mm and on several rings of different dimensions and material. The data obtained for different samples were compared by normalizing the results to the resonance frequency and coupling coefficients obtained on the corresponding radial vibrations of the shortest rings investigated. The results obtained on samples of the same geometry were repeatable and were in good agreement with the normalized results obtained on samples of different geometry whether comprised of ring-bonded elements or single cylinders.

The resonance frequencies of the piezoelectric cylindrical elements and the effective electromechanical coupling coefficients, k_{eff}, were determined by measurement of the electrical admittance of the unloaded cylinders using an HP 4192 impedance analyzer. Representative plots of the moduli of the admittances, $|Y|$, as functions of frequency are depicted for a selection of samples in Fig. 2, illustrating how the electrical properties change for samples of varying aspect ratio. It can be seen from Fig. 2 that in some cases the admittance ratio, $|Y|_{\max}/|Y|_{\min}$, is too small to accurately determine the frequencies of mechanical resonance and to apply the simple formula for calculating the coupling coefficient, $k_{\text{eff}}^2=1-(f_r/f_{ar})^2$. In these cases, more accurate results were obtained from measurement of the dynamic capacitance and conductance of the samples (C_p and G measurement) by using the measurement standard procedures described in

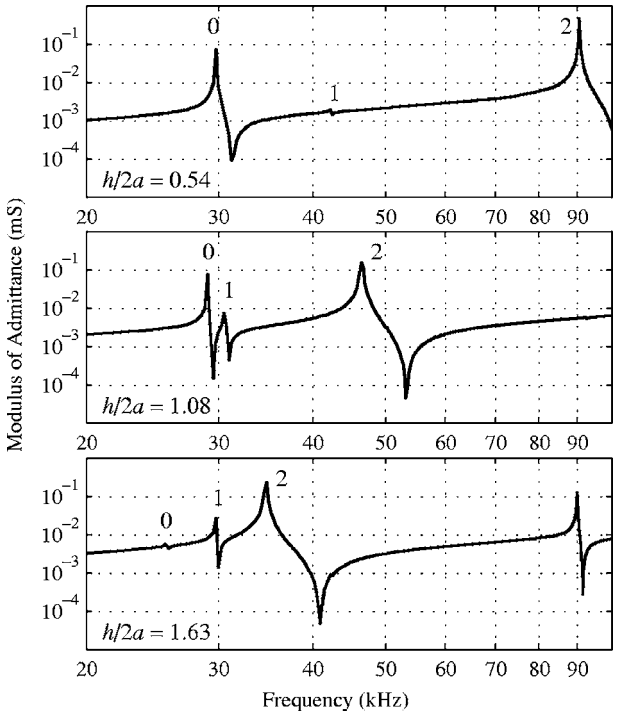


FIG. 2. Admittance of the PZT-4 cylinders with various aspect ratio $h/2a=0.54$, 1.08, and 1.63 as indicated on the plots. The frequencies of resonance are labeled as 0, 1, and 2 corresponding to the corresponding frequency branches. The mean diameter of the cylinders $2a=35$ mm, thickness $t=3.2$ mm.

Ref. 9. To determine the radial vibration mode shapes and the ratios of magnitudes of vibration velocity in the radial and axial directions, $|U_r|/|U_x|$, an optical displacement two-channel probe (MTI, photonic sensor) was used. The sample under investigation was driven from a spectrum analyzer in the frequency range of interest. The output of the sensor was measured at the frequencies corresponding to each resonance. This measurement was then repeated after the displacement sensor was repositioned on the cylinder surface. This point-by-point measurement method was utilized in increments of approximately 2 mm, and provided sufficient data density and accuracy to characterize a particular mode shape. Two optical probes were used simultaneously to determine the sign (phase) of the radial to axial velocity ratio U_r/U_x. The phase shift between the outputs of the two channels was determined with an oscilloscope. When investigating the characteristic mode shapes of vibration in a broad range of aspect ratios $h/2a$, difficulties arose associated with measuring very small radial displacements in the presence of large unwanted axial (transverse) displacements and vice versa. In order to overcome these complications the measurements were performed on samples having the same aspect ratios but different diameters and along different generating lines of a particular sample. The averaged data were then presented. The sufficiently large density and good agreement of the experimental data made it possible to represent the dependence of all the measured quantities vs aspect ratio with trend curves.

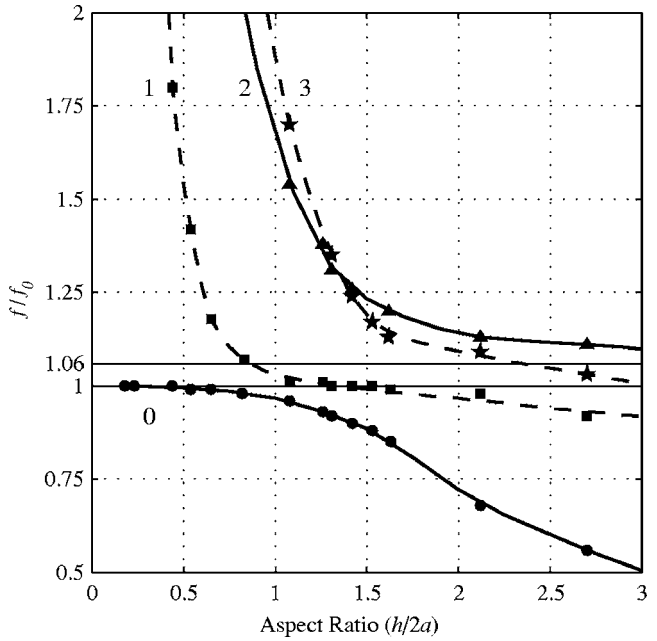


FIG. 3. Resonance frequencies of piezoelectric cylinders as a function of aspect ratio ($h/2a$). Results are normalized to the resonance frequency (f_0) of the radial vibration of a short ring (at $h/2a<0.2$). For the PZT-4 rings with a mean diameter of 35.0mm, the normalization value is $f_0=30.0$ kHz. The branches for the experimental data are drawn as trend (or best-fit) curves. The data related to branch 1 are valid for cylinders with $2a=35$ mm and $t=3.2$ mm. The branches labeled 1 and 3 are seen to cross the so called "dead zone," the space bounded by the lines $f_0/f=1$, and $f/f_0=1/\sqrt{(1-(\sigma_1^E)^2)}=1.06$.

III. EXPERIMENTAL RESULTS

A. Resonance frequencies and modes of vibration

The admittance curves, depicted for a selection of samples having different aspect ratios in Fig. 2, clearly show an existence of three lowest-order frequencies of resonance. They are labeled as 0, 1, and 2 corresponding to the order that they appear in the frequency spectrum. The results of a systematic measurement of these frequencies vs cylinder height-to-diameter aspect ratio are presented in Fig. 3 in normalized form and in Table I as the actual results obtained by measurement of the PZT-4 samples with $2a=35$ mm and $t=3.2$ mm.

The lowest branch (0) corresponds with frequencies f_0. (The choice of f_0 as a notation has its origin in the common labeling of the resonance frequency for the radial vibration of a short ring, i.e., at $h/2a\ll 1$.) At small aspect ratios this branch accounts for the predominantly radial vibrations of a cylinder. The highest branch (branch 2) corresponds with frequencies f_2. At small aspect ratios this branch accounts for the predominantly axial vibrations of a cylinder in its gravest mode. These two branches are well known. They were adequately predicted by membrane theory[1,2] and confirmed experimentally.[1,9] The theoretical data calculated by formulas from Ref. 1 for the same aspect ratios are presented in Table I and are in excellent agreement with our experimental data.

The intermediate branch that corresponds with frequencies f_1 is labeled as branch 1. A notable feature of this branch is that it intersects the so-called dead zone, which is situated between the lines $f_0/f=1$ and $f/f_0=\sqrt{1-(\sigma_1^E)^2}$ (the term $\sigma_1^E=-s_{12}^E/s_{11}^E$ is an analog of the Poisson's ratio for a piezoelectric ceramic), and thus provides experimental evidence that this zone is not dead. The intermediate branch appears to be flexural in nature, at least in the range of aspect ratios considered in this investigation. This was confirmed in particular by measuring the vibration mode shapes at the resonance frequencies belonging to all of the branches and at different aspect ratios. Examples of the mode shapes for the cylinders with aspect ratios $h/2a=0.54$, 1.08, and 2.70 are presented in Fig. 4. For all the aspect ratios, the mode shapes attributed to the frequency branch 1 have two nodes corresponding to nodal circles on the cylinder surface. The mode shapes at $h/2a=0.54$ and 1.08 are qualitatively similar to those of flexural vibration of a free-free beam, however, corresponding to the axially symmetric (tubular) flexural vibration of the shell. Additional evidence of the flexural origin of branch 1 is in the dependence of the corresponding resonance frequencies on the thickness of the cylinder wall for low aspect ratios. This was observed experimentally on samples of aspect ratios $h/2a< 1$. Note that, at the higher aspect ratio ($h/2a=2.70$), the mode shape for branch 1 shows a tendency to be representative of the expected distribution of the radial vibration component corresponding to the third longitudinal mode.

An important note must be made regarding the normalized presentation of the frequency branches in Fig. 3. The normalized extensional branches are quantitatively valid for the thin-walled cylinders regardless of their diameter and thickness. When applied to cylinders made of different materials, the results may depend on the corresponding value of σ_1^E, although no visible difference was observed between the experimental data obtained on samples made from PZT-4 ($\sigma_1^E\approx 0.33$) and PZT-5 ($\sigma_1^E\approx 0.35$). Thus, the related experimental points were used together for the branches 0 and 2 in Fig. 3. Also, note that branch 2 becomes asymptotic to the

TABLE I. Measured resonance frequency data as a function of height-to-diameter aspect ratio for the hollow piezoelectric cylinders of mean diameter 35.0 mm for the lowest four resonant modes branches. The theoretical data, based on the membrane theory presented by Haskins and Walsh (Ref. 1), is included for the frequencies on the lower (0) and upper (2) branches for comparison purposes.

Ratio	Measured (kHz)				(Ref. 1)	
$h/2a$	f_0	f_1	f_2	f_3	f_0	f_2
0.18	30.0		264		30.0	288
0.23	30.0	138	213		30.0	212
0.44	30.0	54.0	111		29.9	111
0.54	29.7	42.6	90.0		29.9	90.5
0.65	29.7	35.4	75.0		29.9	75.4
0.83	29.4	32.1	60.0	72.0	29.7	59.3
1.08	28.8	30.3	46.2	51.0	29.4	46.2
1.26	27.9	30.0	41.4		28.8	40.4
1.31	27.6	30.0	39.3	40.5	28.6	39.1
1.42	27.0	30.0	37.8	37.2	28.0	36.8
1.53	26.4	30.0		35.1	27.2	35.2
1.63	25.5	29.7	36.0	33.9	26.3	34.1
2.12	20.4	29.4	33.9	32.7	21.5	32.0
2.70	16.8	27.6	33.3	30.9	17.2	31.5
3.18	14.1	27.0	32.7	29.7	14.7	31.3

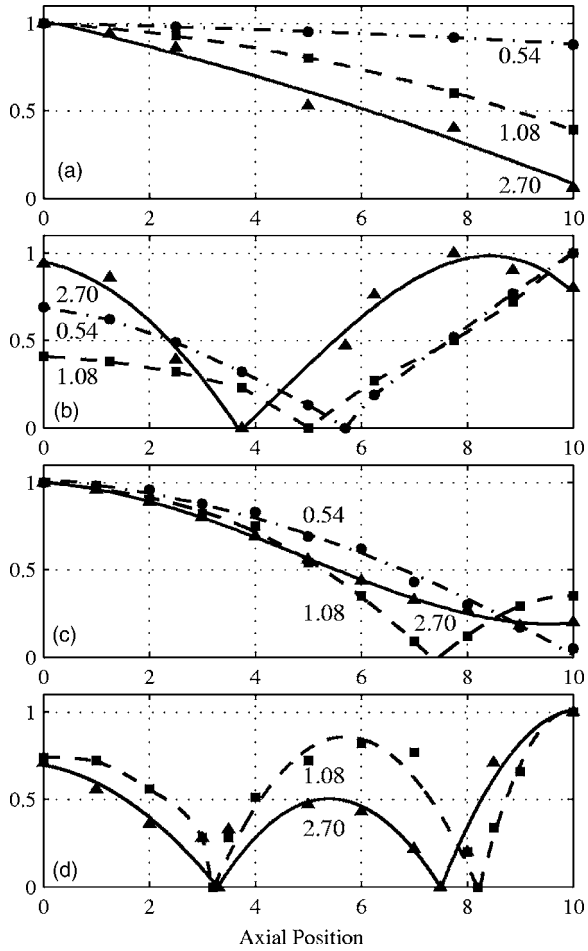


FIG. 4. Plots of the mode shapes measured at different aspect ratios at the corresponding resonance frequencies associated with each frequency branch: (a) branch 0; (b) branch 1; (c) branch 2; and (d) branch 3. The aspect ratios are labeled on the plots.

line $f/f_0=\sqrt{1-(\sigma_1^E)^2}$, corresponding to 1.06 and 1.07 for PZT-4 and PZT-5, respectively. The normalized flexural branch can only be used quantitatively for the experimental data presented in Fig. 3, which were obtained for the cylinders with dimensions $2a=35$ mm and $t=3.2$ mm. For PZT-4 cylinders having this particular combination of dimensions, the actual measured resonance frequencies are given in Table I. The normalization factor of $f=30.0$ kHz was used for the presentation in Fig. 3, which corresponds to the measured resonance of a short ring. For different cylinders with different combinations of diameter and thickness, the flexural branch may be considered as qualitatively representative.

Although the flexural branch could not be accounted for by the membrane theory approximation employed by Haskins and Walsh,[1] it is surprising that in the experimental part of their treatment, this intermediate branch was not detected at aspect ratios below approximately unity. Also, they interpreted the experimental data obtained on the approximately horizontal part of this branch as belonging to the radial mode coupled with the second ($n=2$) axial resonant mode.[1]

After the intermediate branch (1) was shown to cross the dead zone, a more careful examination of this frequency region was undertaken. The frequency bands beyond the frequencies of strong resonances were studied with greater resolution in order to look for weaker resonances, which otherwise would be undetected in the region of much stronger resonances. This search revealed more resonance frequencies, which were also regularly spaced. The corresponding experimental points are shown in Fig. 3 as forming the branch labeled as branch 3. Mode shapes were also measured for several aspect ratios on this branch, and the results for $h/2a=1.0$ and 2.87 are presented in Fig. 4(d). Judging by the number of nodes and their location, and the distribution of resonance frequencies, the frequency branch 3 can be associated with the second mode of axial symmetric flexural vibration of the cylinder. Thus, the dead zone appears to be populated by flexural modes of various orders.

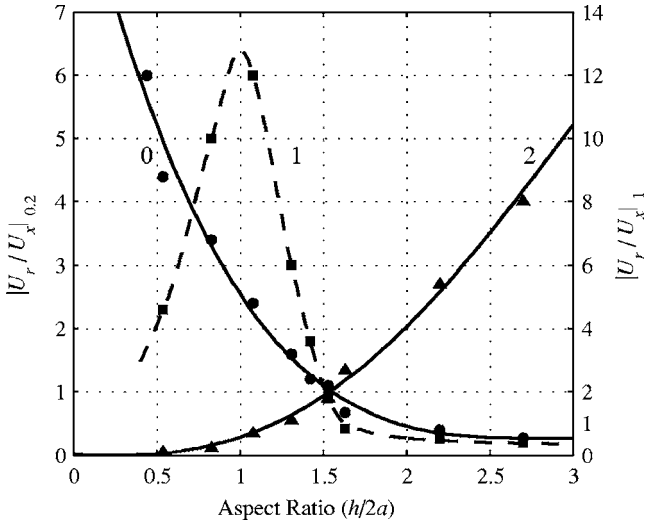


FIG. 5. Ratios of the radial to axial velocity $|U_r/U_x|$ as functions of $h/2a$ measured on the cylinders with $2a=35$ mm and $t=3.2$ mm. The trend curves for the experimental data are presented. The scale on the left is for modes 0 and 2 while the scale on the right is for mode 1.

It is well known from the literature, and clearly articulated in Ref. 8, that the vibrations associated with the lower branch 0 are predominantly radial at very small aspect ratios and gradually become predominantly axial at large aspect ratios. And, conversely, the vibrations associated with the upper branch (2) are axial at small aspect ratios and gradually become radial for long cylinders of large aspect ratios. In order to quantitatively estimate how this transition occurs, and to characterize the progression of the flexural mode along the intermediate branch (1), the ratio of magnitudes of vibration in the radial and axial directions U_r/U_x was measured vs aspect ratio for the three branches. The results obtained are presented in Fig. 5. It was observed that the vibrations in the radial and axial directions are out of phase along the lower branch (0) and are in phase along the upper branch (2) (with reference that the displacements outward to surface of the cylinder are considered to have the same sign). The trend curves for the experimental (modulus) data intersect approximately at the value $h/2a=\pi/2$, which is known as the occurrence of the strongest coupling, and at this point the radial to axial velocity is unity, $|U_r/U_x|=1$. These results are in accordance with the theoretical predictions made in Ref. 8 and calculations made based on that theory in Ref. 10.

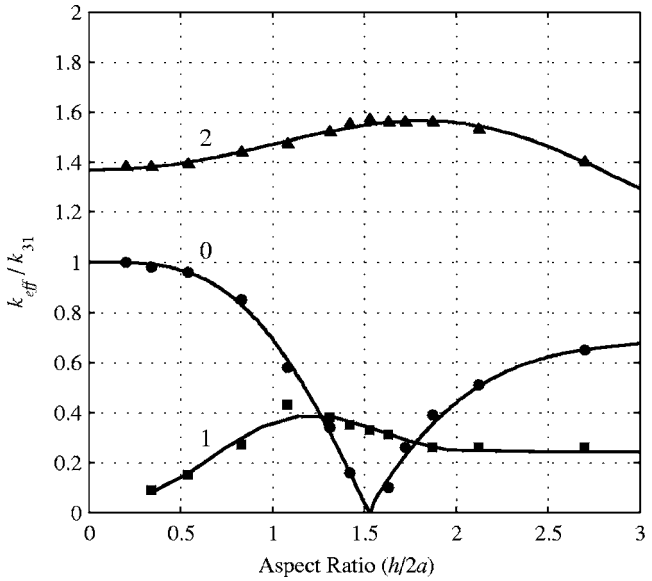


FIG. 6. The effective coupling coefficients normalized to $k_{31}=0.34$ of PZT-4 for the frequency branches as functions of aspect ratio. The data for the intermediate branch 1 are valid for cylinders with $2a=35$ mm and $t=3.2$ mm. The trend curves for the data are shown.

There exists a clear maximum of the radial to axial displacement ratio U_r/U_x for the flexural branch 1, and in some range of aspect ratios around this maximum the radial displacement is much greater than axial (note that even a different scale is used to present this ratio). This could be expected taking into consideration the nature of the flexural vibration. The vibrations in this mode become more longitudinal with increasing aspect ratio. This tendency is also illustrated by the mode shape shown in Fig. 4(b) at $h/2a=2.7$, as was previously noted. Therefore, it is more appropriate to regard this branch as "intermediate," although at small aspect ratios the flexural vibrations clearly dominate. We note that the experimental data related to the extensional branches 0 and 2 are numerically valid regardless of the diameter and thickness of the thin-walled cylinder. The data related to the intermediate branch (1) are valid numerically only for the cylinder with dimensions $2a=35$ mm and $t=3.2$ mm. Otherwise the data have to be considered as qualitative. In particular, the position of the maximum and the values of velocity ratios may change for different geometries.

B. The effective coupling coefficients

The experimental data on the dependence of the effective coupling coefficients vs $h/2a$ ratio for the frequency branches are presented in Fig. 6. The experimental data are normalized to the value of the material coupling coefficient of PZT-4, $k_{31}=0.34$, which in our particular case was measured on a ring with $h/2a=0.15$. The trend curves related to the extensional branches 0 and 2 are valid in general, regardless of the dimensions of the cylinder. In order to obtain an actual value, the normalized coefficient has to be multiplied by the coupling coefficient k_{31} of the particular piezoelectric ceramic material employed. The intermediate branch data are valid for cylinders with dimensions $2a=35$ mm and $t=3.2$ mm. Otherwise they qualitatively characterize the dependence of the coupling coefficient. For the lower branch, the effective coupling coefficient drops rapidly toward zero corresponding to the aspect ratio $h/2a=\pi/2$ (the trend curve clearly shows this tendency). In the vicinity of this aspect ratio, the coupling coefficient remains very small, and the gravest radial mode practically disappears (see the admittance curve at $h/2a=1.63$ in Fig. 2). For the upper branch the effective coupling coefficient remains large in all the range of $h/2a$ ratios and even has a maximum in the region near $h/2a=\pi/2$. The behavior of the effective coupling coefficients along the extensional frequency branches qualitatively correlates with the fact that the axial and radial vibrations are out of phase for the lower and in phase for the upper branches, as the electromechanically induced stresses in the axial and radial directions are of the same sign for the radially polarized cylinders. The results displayed in Fig. 6 are in agreement with predictions noted in Ref. 8.

IV. CASE STUDY: DESIGN EXAMPLE OF UNDERWATER ACOUSTIC TRANSDUCERS

In order to illustrate the effect of coupled vibrations on transducer performance and to estimate how this effect correlates with the experimental data on the parameters of the cylindrical piezoelectric elements vs their aspect ratio, two transducers of the same overall height were fabricated from PZT-4 cylinders with the mean diameter $2a=35$ mm and thickness $t=3.2$ mm. One of them was made from the single cylinder with aspect ratio $h/2a=1.1$, and the other from two cylinders with aspect ratio $h/2a=0.54$. The two smaller cylinders were connected electrically in parallel. The transducers were built in an air-backed configuration with piezoelectric ceramic cylinders mechanically isolated from the caps, and from each other, which practically achieves the free-free boundary conditions.

The comparative characteristics of the two transducers are illustrated in Figs. 7–9. Parameters of the input admittances of transducers (conductance and capacitance) are shown in Fig. 7. There is a significant difference between the frequency dependence of the parameters, which should be taken into account when matching the transducer with a power amplifier to maximize operation in a broad frequency range. The frequency dependence of the dynamical capacitance and conductance of the transducers in air [Figs. 7(a) and 7(b)] clearly correlates with the admittance curves presented in Fig. 2 for the corresponding aspect ratios. The frequency dependencies measured in water [Figs. 7(c) and 7(d)] are also influenced by a combined effect of the differences in the vibration mode shapes and in the coupling coefficients that are illustrated by Figs. 4 and 6.

The power factor ($\cos\varphi$) comparison is depicted in Figs. 8(a) and 8(b) for the case of the transducer that is not tuned and for the case that is tuned at its resonance frequency, respectively. It is seen that the transducer comprised of two cylinders has a corresponding greater power factor and greater operational bandwidth for both without tuning and in the case of tuning, than that of the single taller cylinder. The

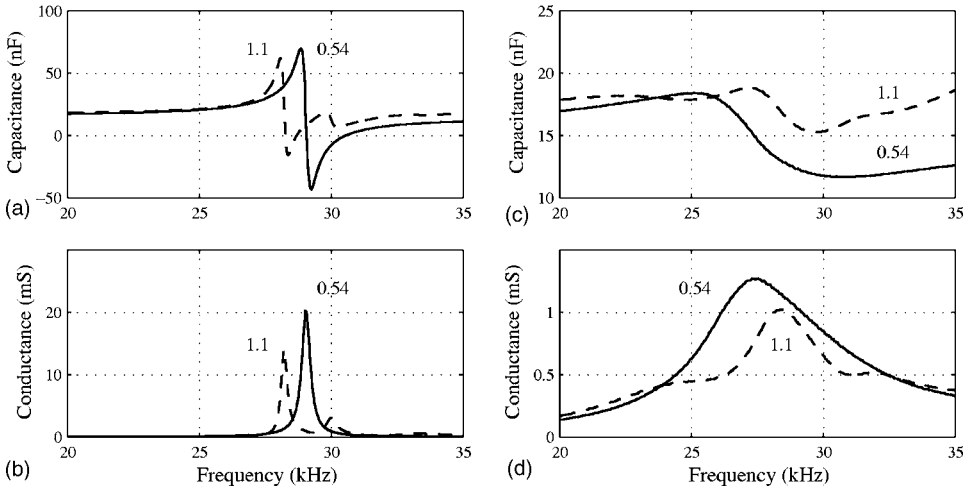


FIG. 7. Admittance plots for the transducers made from the single cylinder with $h/2a=1.1$ (curves 1) and from the two cylinders with $h/2a=0.54$ (curves 2): (a) and (b) measured in air; (c) and (d) measured in water.

operational bandwidth for the tuned transducer is greater for the two-ring transducer by as much as a factor of 2, which is in direct correspondence with the increase of the effective coupling coefficient for aspect ratio $h/2a=0.54$, as compared with aspect ratio $h/2a=1.1$ shown for the branch 0 in Fig. 6.

The transmit voltage response (TVR) and the receive free-field voltage sensitivity (FFVS) of the two transducers measured without tuning are depicted in Figs. 9(a) and 9(b). Both of these quantify an advantage of the two-cylinder transducer in the operational frequency range around the radial resonance.

Qualitatively the results of this case study could be predicted based on the experimental data regarding the effective coupling coefficients and mode shapes of vibration for the case where the operational resonance frequencies belong to the lower branch and are at aspect ratios below the point of the strongest coupling. For more quantitative and analytical predictions of the performance of cylindrical electroacoustic transducers vs aspect ratio of comprising piezoelectric elements, a treatment of the coupled piezoelectric-elastic-acoustic problem has to be made (or should be). An example of such treatment for a free-flooded piezoelectric cylinder transducer is given in Ref. 11, in which the vibration of a finite-length cylinder *in vacuo* was considered based on the Haskins and Walsh solution.[1]

V. CONCLUSIONS

The results of an experimental investigation of the resonance frequencies, corresponding modes of vibration, and effective coupling coefficients vs the height-to-diameter aspect ratio for radially polarized cylindrical cylinders were presented. The investigation revealed an intermediate resonance frequency branch, which lives between the established lower and upper frequency branches and crosses the so-

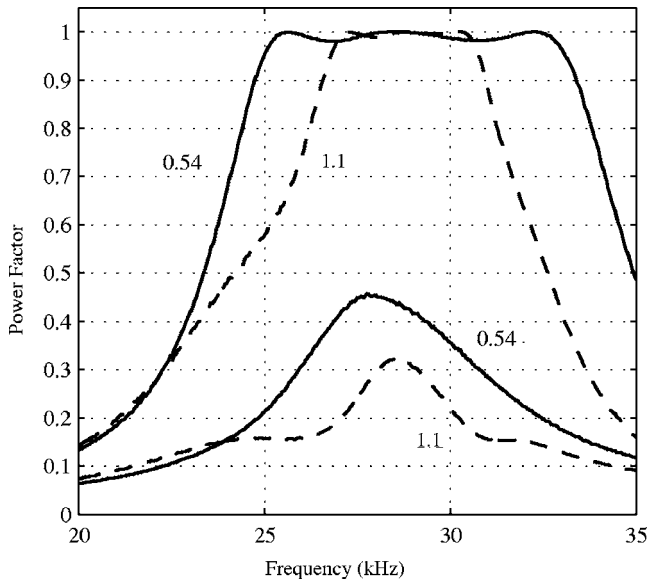


FIG. 8. Power factor curves of the transducers made from the single cylinder with $h/2a=1.1$ and two cylinders with $h/2a=0.54$ (a) without tuning (lower two curves), and (b) tuned at the resonance frequencies.

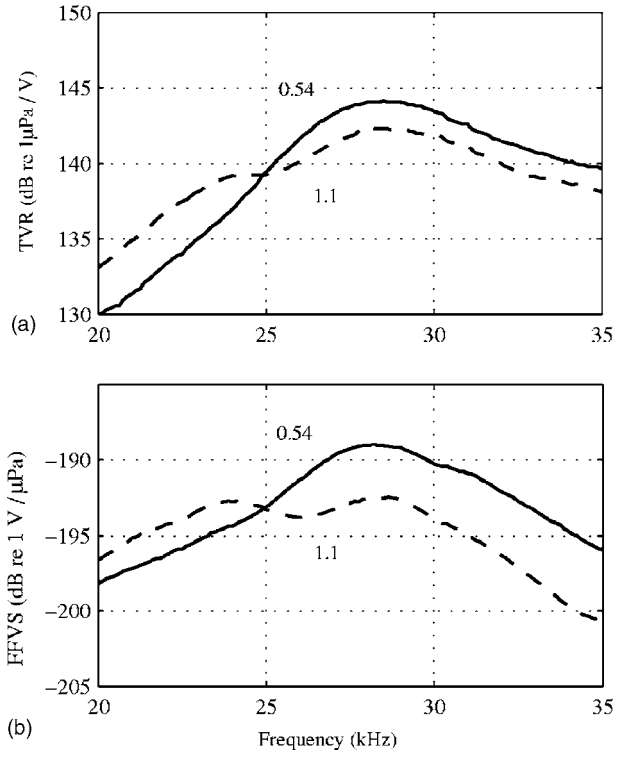


FIG. 9. Frequency response of the transducers made of the single cylinder with $h/2a=1.1$ and two cylinders with $h/2a=0.54$: (a) transmit voltage response (TVR); (b) receive free-field voltage sensitivity (FFVS).

called dead zone. In the lower range of height-to-diameter aspect ratios, this mode is clearly dominated by the flexural vibrations of the cylinder wall. This was verified by measurements of the corresponding vibration mode shapes of the cylinders. Measurements of the radial and axial vibration magnitudes showed the transformation from predominantly radial to longitudinal motion for the lower branch and from predominantly axial to radial for the upper branch as a function of height-to-diameter aspect ratio. The ratio of radial to axial displacement for the intermediate branch showed a peak near the region of strongest coupling. The experimental results on the effective coupling coefficients were obtained for the three frequency branches. Together with the more expanded investigation of the resonance frequencies presented, these results may be used for improving the performance of electroacoustic transducers. To illustrate this, a case study of two underwater transducers composed of cylindrical piezoelectric elements having different $h/2a$ ratios illustrates the effects of coupled vibrations on their performance. Based on the results obtained, it can be concluded that the performance of the electromechanical and electroacoustic transducers may be improved in terms of a substantial increase in the effective electromechanical coupling coefficient, transmit and receive response, power factor, and practical usable bandwidth by an informed choice of the height-to-diameter aspect ratio of the comprising piezoelectric elements.

ACKNOWLEDGMENTS

This work was supported by BTech Acoustics, LLC, and assisted in part by related work funded by the Office of Naval Research. The authors acknowledge the assistance of Sheng Liu, who assisted with the measurements of the vibration mode shapes, as well as Corey Bachand and Stephen Murray, who assisted with the preparation of figures. The contributing author and research assistant, Sundar Regmi, made preliminary measurements of samples in 2004 and has since relocated to the Transducer Development Group, Olympus-NDT, 48 Wored Ave., Waltham, MA 02453.

[1]J. F. Haskins and J. L. Walsh, "Vibration of ferroelectric cylindrical shells with transverse isotropy. I. Radially polarized case," J. Acoust. Soc. Am. **29**, 729–734 (1957).

[2]A. E. H. Love, *Mathematical Theory of Elasticity*, 4th ed. (Dover, New York, 1944), p. 546.

[3]M. C. Junger and F. G. Rosato, "The propagation of elastic waves in thin-walled cylindrical shells," J. Acoust. Soc. Am. **26**, 709–713 (1954).

[4]J. A. Martin, "Vibrations of longitudinally polarized ferroelectric cylindrical tubes," J. Acoust. Soc. Am. **35**, 510–520 (1963).

[5]H. Wang, "On the tangentially and radially polarized piezoceramic thin cylindrical tube transducers," J. Acoust. Soc. Am. **79**, 164–176 (1986).

[6]D. D. Ebenezer and A. Pushba, "Eigenfunction analysis of radially polarized cylindrical piezoelectric shells of finite length," J. Acoust. Soc. Am. **102**, 1549–1558 (1997).

[7]D. D. Ebenezer and A. Pushba, "Piezoelectric thin shell theoretical model and eigenfunction analysis of radially polarized ceramic cylinders," J. Acoust. Soc. Am. **105**(1), 154–163 (1999).

[8]D. S. Drumheller and A. Kalnins, "Dynamic shell theory for ferroelectric ceramic," J. Acoust. Soc. Am. **47**, 1343–1353 (1970).

[9]ANSI/IEEE Standard on Piezoelectricity - Std. 176 - 1987, Sec.6, The Institute of Electrical and Electronics Engineers, Inc., New York (1987).

[10]L. J. Eriksson, "On the vibrations of thin cylindrical shells," J. Acoust. Soc. Am. **49**, 1672–1673 (1971).

[11]P. H. Rogers, "Mathematical model for a free-flooded piezoelectric cylinder transducer," J. Acoust. Soc. Am. **80**(1), 13–18 (1986).

Coupling between thermoacoustic resonance pipes and piezoelectric loudspeakers studied by equivalent circuit method

Li Fan, Shu-yi Zhang, and Ben-ren Wang
Lab of Modern Acoustics, Institute of Acoustics, Nanjing University, Nanjing 210093, P. R. China

(Received 7 October 2005; revised 15 June 2006; accepted 20 June 2006)

In order to miniaturize thermoacoustic refrigerators, piezoelectric loudspeakers are used as acoustic sources for operating at high resonance frequencies. An integrated equivalent circuit model of a thermoacoustic resonance pipe driven by a piezoelectric loudspeaker is presented to investigate the coupling conditions between the piezoelectric loudspeaker and thermoacoustic resonance pipe. Using the equivalent circuit model the optimized coupling conditions can be obtained, with which the highest electroacoustic transfer efficiency and the largest acoustic output power of a piezoelectric acoustic source can be achieved. The theoretical simulations are taken to optimize the structures and the operating frequencies of the thermoacoustic systems. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2225643]

PACS number(s): 43.38.Ja, 43.35.Ud, 43.38.Fx [AJZ] Pages: 1381–1387

I. INTRODUCTION

In order to miniaturize thermoacoustic refrigerators, Hofler *et al.* reported a new thermoacoustic refrigerator driven by a piezoelectric loudspeaker, which has much higher resonance frequency than that driven by an electrodynamic loudspeaker.[1] Then the piezoelectric thermoacoustic refrigerators with small sizes operating at the frequency of about 4000 Hz are presented.[2,3] Meanwhile, Lihoreau *et al.* set up a thermoacoustic system driven by a piezoelectric disk stuck on a passive backing plate (without vibration diagram) and presented a corresponding equivalent circuit model of the system.[4] Lihoreau *et al.* also calculated the equivalent parameters of the piezoelectric transducer with rear and front loads and obtained the frequency response of the acoustic pressure in the pipe, which has been verified by their experimental measurements. The acoustic pressure, particle velocity, and power flux in the resonance pipe with and without stacks are also calculated to evaluate the influence of the stack on the acoustic transmission in the pipe. However, the results are not suitable to illustrate the performance of Hofler's system driven by an integrated loudspeaker including a vibration diaphragm. Furthermore, they focused on the optimized operating frequency in the view of the acoustic pressure and power flow, but ignored the efficiency of the system and did not optimize the equivalent parameters of the piezoelectric loudspeaker itself.

In order to investigate and optimize the kind of thermoacoustic refrigerators reported by Hofler *et al.*, an integrated equivalent circuit model composed of the equivalent parameters of the piezoelectric disk, vibration diaphragm, coupling component, rear cavity, and acoustic load (resonance pipe) is presented in this paper. With the theoretical model, two reactance matching conditions of the loudspeaker, i.e., the coupling component to the piezoelectric disk and the component to the vibration diaphragm, are demonstrated, which are the basic conditions to obtain large electroacoustic transfer efficiency and acoustic output power. The influences of the stiffness of the coupling component on the strokes of the piezoelectric disk and the vibration diaphragm and also on the electroacoustic transfer efficiency and acoustic output power are presented. Accordingly, the optimized input resistance of the resonance pipe, i.e., the loaded resistance of the loudspeaker, determined by the stiffness of the coupling component and the resistances of the loudspeaker, are demonstrated. Then, the trade-off between the performance optimization and the system miniaturization are discussed. Additionally, it is also illustrated that the optimized operating frequency of the system must be close to the resonance frequency of the resonance pipe in order to meet the impedance matching conditions, as well as reduce the system volume. Therefore, the highest transfer efficiency and the largest acoustic output power of the piezoelectric loudspeaker can be obtained, and the optimization of the system can be achieved.

II. COUPLING BETWEEN RESONANCE PIPE AND PIEZOELECTRIC LOUDSPEAKER

A. Equivalent circuit of thermoacoustic refrigerator driven by piezoelectric loudspeaker

Simple models of a piezoelectric loudspeaker and a thermoacoustic refrigerator are shown in Figs. 1(a) and 1(b), respectively. The integrated equivalent circuit model of the thermoacoustic system is shown in Fig. 2(a), where e is the driving voltage, C_E' is the static capacitance of the piezoelectric disk, and C_p, M_p, and R_p are the equivalent mechanical compliance, mass, and resistance of the piezoelectric disk, respectively. C_d, M_d and R_d are the mechanical compliance, mass, and resistance of the vibration diaphragm, respectively. C_R and C_ε are the mechanical compliances of the rear cavity and coupling component, respectively. The transformation coefficients from the electrical and acoustical ends to the mechanical end are N and S, respectively. Z_{AR} is the acoustic load impedance of the piezoelectric loudspeaker. Thus, removing the transformers and transforming the electrical and acoustical ends to the mechanical end, the equivalent circuit Fig. 2(a) is simplified to Fig. 2(b). In the mechanical equiva-

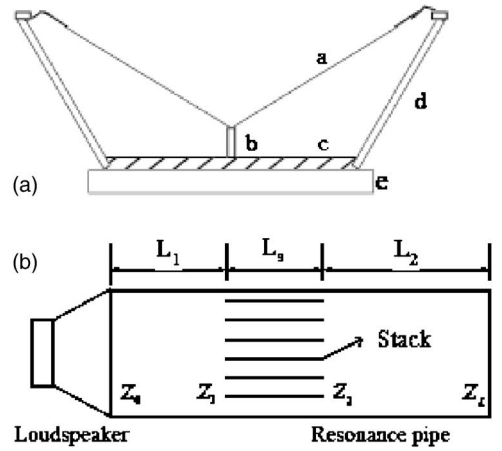


FIG. 1. (a) Schematic diagram of the piezoelectric loudspeaker: a: vibration diaphragm, b: coupling component, c: piezoelectric disk, d: supporting component, and e: substrate. (b) Schematic diagram of a thermoacoustic refrigerator.

lent circuit of Fig. 2(b), F_e and F_A are the equivalent driving forces on the whole system and the mechanical load Z_{MR}, respectively.

B. Expressions of electroacoustic transfer efficiency and acoustic output power

For convenience, define [see Fig. 2(b)]

$$C_0 = C_d C_R/(C_d + C_R),$$

$$Z_d = R_d + j(\omega M_d - 1/\omega C_0) = R_d + jX_d,$$

$$Z_M = Z_d + Z_{MR} = R_M + jX_M,$$

$$Z_{MR} = R_{MR} + jX_{MR},$$

$$Z_\varepsilon = 1/j\omega C_\varepsilon,$$

$$Z_R = Z_M Z_\varepsilon/(Z_M + Z_\varepsilon) = R_R + jX_R,$$

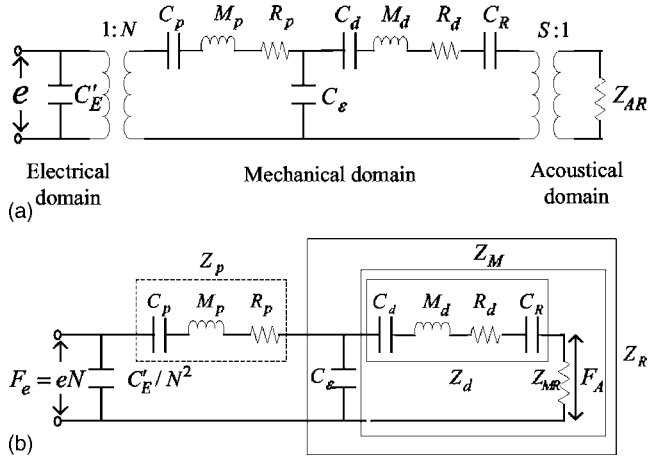


FIG. 2. Equivalent circuit: (a) thermoacoustic refrigerator system and (b) removing transformers of (a).

$$Z_p = R_p + j(\omega M_p - 1/\omega C_p) = R_p + jX_p, \quad (1)$$

where ω is the angle frequency of the electric source. Thus, the electric power P_E applied to the loudspeaker is

$$P_E = \tfrac{1}{2}|F_e|^2(R_p + R_R)/|Z_p + Z_R|^2, \quad (2)$$

and the acoustic output power P_A of the loudspeaker is

$$P_A = \tfrac{1}{2}\,\mathrm{Re}[F_A v_d^*] = \tfrac{1}{2}(|F_A|^2/|Z_{MR}|^2)\mathrm{Re}[Z_{MR}], \quad (3)$$

where $v_d = F_A/Z_{MR}$ is the vibration velocity of the diaphragm and * indicates the complex conjugate.

Using Eq. (1), the acoustic power can be obtained from Eq. (3):

$$P_A = \frac{1}{2}|F_e|^2 \frac{R_{MR}}{|Z_R + Z_p|^2(|Z_M + Z_\varepsilon|/|Z_\varepsilon|)^2}. \quad (4)$$

Then, the electroacoustic transfer efficiency η can be obtained:

$$\frac{1}{\eta} = \frac{R_M}{R_{MR}} + \frac{\omega^2 C_\varepsilon^2 R_p[R_M^2 + (X_M - 1/\omega C_\varepsilon)^2]}{R_{MR}}. \quad (5)$$

Define

$$\alpha = |Z_M + Z_\varepsilon|^2/|Z_\varepsilon|^2. \quad (6)$$

Then η and P_A can be expressed as

$$\eta = \frac{R_{MR}}{\alpha(R_R + R_p)}, \quad (7)$$

$$P_A = \frac{1}{2}|F_e|^2 \frac{R_{MR}}{\alpha|Z_R + Z_p|^2}. \quad (8)$$

If the stiffness of the coupling component is high enough, the corresponding mechanical compliances C_ε can be taken as an open circuit, i.e., the mechanical impedance $|Z_\varepsilon| \to \infty$ and $a = 1$. Then η and P_A can be simplified as η_l and P_{Al}:

$$\eta_l = R_{MR}/(R_p + R_d + R_{MR}), \quad (9)$$

$$P_{Al} = \frac{1}{2}|F_e|^2 \frac{R_{MR}}{(R_p + R_d + R_{MR})^2 + (X_p + X_d + X_{MR})^2}. \quad (10)$$

III. RESULTS AND DISCUSSIONS

A. Electroacoustic transfer efficiency

When the resonance condition of the parallel branch Z_R [see Fig. 2(b)] is satisfied, i.e., $X_M - 1/\omega C_\varepsilon = 0$, from Eq. (5), η becomes η_r:

$$\eta_r = R_{MR}/(R_M + R_p R_M^2/|Z_\varepsilon|^2). \quad (11)$$

Expanding Eq. (11) using Eq. (1), one can obtain

$$\eta_r = 1 \Bigg/ \left[\left(\frac{R_p}{|Z_\varepsilon|^2} \right) R_{MR} + \left(R_d + \frac{R_p R_d^2}{|Z_\varepsilon|^2} \right) \frac{1}{R_{MR}} + \left(1 + \frac{2R_p R_d}{|Z_\varepsilon|^2} \right) \right]. \quad (12)$$

To obtain an optimized electroacoustic transfer efficiency

η_{opt}, the mechanical resistance R_{MR} equals $R_{MR\eta}$:

$$R_{MR\eta} = \sqrt{(R_d|Z_\varepsilon|^2 + R_p R_d^2)/R_p}. \quad (13)$$

Then η_{opt} is

$$\eta_{opt} = 1 \Bigg/ \left[2\sqrt{\frac{R_p R_d}{|Z_\varepsilon|^2} + \frac{R_p^2 R_d^2}{|Z_\varepsilon|^4}} + \left(1 + \frac{2R_p R_d}{|Z_\varepsilon|^2}\right) \right]. \quad (14)$$

It shows that the small resistances of the piezoelectric disk R_p and the vibration diaphragm R_d are favorable to increasing η_{opt}, meanwhile a large $|Z_\varepsilon|$, i.e., high stiffness of the coupling component, can also increase η_{opt}.

When C_ε can be taken as an open circuit, as shown in Eq. (9), the electroacoustic transfer efficiency can be increased by $R_{MR} \gg R_d + R_p$.

Under the parallel resonance condition, $\alpha = R_M^2/X_M^2 = R_M^2/|Z_\varepsilon|^2$, from Eq. (11), one can get

$$\eta_r = R_{MR}/(\alpha R_p + R_d + R_{MR}). \quad (15)$$

Comparing Eq. (9) with Eq. (15), one can see that the coefficient α indicates the influence of C_ε on η. Obviously, a small α can induce a large η. When $\alpha = 1$, η equals η_l. When $|Z_\varepsilon|^2 \gg R_M^2$, i.e., $\alpha \approx 0$, the influence of R_p on the efficiency nearly completely vanishes and then $\eta = R_{MR}/(R_{MR} + R_d)$.

B. Acoustic output power

Using Eq. (1), Eq. (8) can be expanded as

$$P_A = \frac{1}{2}|F_e|^2 R_{MR} \Bigg/ \left(|Z_M + Z_\varepsilon|^2/|Z_\varepsilon|^2 \right) \times \left| \frac{(R_M + jX_M)/j\omega C_\varepsilon}{R_M + j(X_M - 1/\omega C_\varepsilon)} + (R_p + jX_p) \right|^2. \quad (16)$$

When the system is operating under the reactance matching condition between the vibration diaphragm and coupling component $X_M - 1/\omega C_\varepsilon = 0$, Eq. (16) can be simplified to be

$$P_{Ar} = \frac{1}{2} \frac{|F_e|^2 R_{MR}}{(R_M/|Z_\varepsilon|)^2 [(R_p + |Z_\varepsilon|^2/R_M)^2 + (X_p - 1/\omega C_\varepsilon)^2]}. \quad (17)$$

When the reactance matching condition between the piezoelectric disk and coupling component is also satisfied, i.e., $X_p - 1/\omega C_\varepsilon = 0$, P_{Ar} becomes P_{Am}:

$$P_{Am} = \frac{1}{2}|F_e|^2 \frac{R_{MR}}{(|Z_\varepsilon| + R_M R_p/|Z_\varepsilon|)^2}. \quad (18)$$

Expanding Eq. (18), one can obtain

$$P_{Am} = \frac{1}{2}|F_e|^2 \Bigg/ \left[\frac{(|Z_\varepsilon|^2 + 2R_p R_d + R_p^2 R_d^2/|Z_\varepsilon|^2)}{R_{MR}} + \frac{R_p^2}{|Z_\varepsilon|^2} R_{MR} + 2\left(R_p + \frac{R_p^2 R_d}{|Z_\varepsilon|^2}\right) \right]. \quad (19)$$

Therefore, to obtain an optimized output acoustic power P_{Aopt}, R_{MR} equals R_{MRP}:

$$R_{MRP} = R_d + |Z_\varepsilon|^2/R_p. \quad (20)$$

Then P_{Aopt} is

$$P_{Aopt} = \frac{|F_e|^2}{8(R_p + R_p^2 R_d/|Z_\varepsilon|^2)}. \quad (21)$$

It can be seen that the small resistances R_p, R_d and large $|Z_\varepsilon|$ are favorable to increasing P_{Aopt}, which is in agreement with the optimization conditions of η_{opt} as shown in Eq. (14).

When C_ε can be taken as an open circuit, the optimized output acoustic power P_{Alopt} can be obtained when $X_{MR} = -(X_p + X_d)$ and $R_{MR} = R_d + R_d$:

$$P_{Alopt} = |F_e|^2/8(R_p + R_d). \quad (22)$$

The impedance matching conditions are similar to that in a thermoacoustic system driven by an electrodynamic loudspeaker.[5] The acoustic output power can be increased by decreasing R_p and R_d, which is the same as the optimization conditions of η_l. It can be seen that, when $X_{MR} = -(X_p + X_d)$, the optimization conditions of η_l and P_{Al} are the same as those of a pure electric resistance R_{MR} driven by a voltage source with the internal resistance $R_p + R_d$.

The influences of C_ε on the optimized acoustic output power can be evaluated by comparing Eqs. (21) and (22). In fact, C_ε induces a coefficient $\beta = R_p^2/|Z_\varepsilon|^2$ of R_d in Eq. (21). Therefore, when $\beta < 1$, $P_{Aopt} > P_{Alopt}$; when $\beta = 1$, $P_{Aopt} = P_{Alopt}$; and when $\beta > 1$, $P_{Aopt} < P_{Alopt}$.

C. Strokes of piezoelectric disk and vibration diaphragm

The optimized conditions described above can be explained by the strokes of the piezoelectric disk and vibration diaphragm. The power consumption in the whole system can be divided into three parts: consumptions on R_p, R_d, and R_{MR}, which are indicated as P_p, P_d, and P_A, respectively. P_p is proportional to $|v_p|^2$, v_p is the vibration velocity of the piezoelectric disk, and P_d, P_A are both proportional to $|v_d|^2$. When the system is operating at an angle frequency ω, the strokes of the piezoelectric disk and diaphragm are $s_p = 2|v_p|/\omega$ and $s_d = 2|v_d|/\omega$, respectively, so the power consumptions are also proportional to the squares of the strokes of the piezoelectric disk and diaphragm. Using the current distribution equation of a parallel circuit, one can obtain the ratio A of the stroke of the diaphragm to that of the piezoelectric disk:

$$A = \frac{|Z_\varepsilon|}{|Z_M + Z_\varepsilon|} = \frac{1}{\sqrt{\alpha}} \quad (23)$$

or

$$A = \frac{|Z_\varepsilon|}{\sqrt{R_M^2 + (X_M - 1/\omega C_\varepsilon)^2}}. \quad (24)$$

Here, A can be defined as a stroke amplification coefficient.

Under the optimized conditions for η, $X_M - 1/\omega C_\varepsilon = 0$ and $R_{MR} = R_{MR\eta}$,

$$A = \frac{|Z_\varepsilon|}{R_{MR\eta}} = \frac{1}{R_d/|Z_\varepsilon| + \sqrt{R_d/R_p + R_d^2/|Z_\varepsilon|^2}}. \quad (25)$$

It can be seen that the optimized conditions and a large $|Z_\varepsilon|$ induce a large A, which ensures that the stroke of the dia-

phragm is much larger than that of the piezoelectric.

On the other hand, it can be derived from Eq. (15) that

$$\eta_r = \frac{R_{MR}}{R_{MR} + R_d + R_p/A^2}. \tag{26}$$

Obviously, a large stroke amplification coefficient A greatly decreases the unfavorable influence of R_p on η_r, i.e., it decreases the ratio of the power loss on the piezoelectric disk to that on the whole system, therefore producing a large η_r. Additionally, if $R_p/A^2 \ll 1$, $\eta_r \approx R_{MR}/(R_{MR}+R_d)$. In this case, a large R_{MR} can induce a large η_r but decreases A simultaneously. Therefore, there exists an optimized R_{MR} with which the highest efficiency can be obtained, as shown in Eq. (13).

Since the parallel resonance condition and a large $|Z_\varepsilon|$ can induce a large stroke amplification coefficient A, a large stroke s_p of the piezoelectric disk must be obtained for achieving a large absolute stroke s_d of the diaphragm, i.e., a large acoustic output power. The stroke of the piezoelectric disk is

$$s_p = \frac{(2F_e/\omega)}{||Z_\varepsilon|^2/(R_d + R_{MR}) + R_p + j(X_p - 1/\omega C_\varepsilon)|}. \tag{27}$$

It can be seen that the reactance matching condition between the piezoelectric disk and coupling component, $X_p - 1/\omega C_\varepsilon = 0$, can ensure a large stroke $s_p = 2F_e/\omega[|Z_\varepsilon|^2/(R_d+R_{MR}) + R_p]$. Meanwhile, a large $|Z_\varepsilon|$ decreases s_p although it is favorable to A. Therefore, a larger R_{MR} is required to reduce the unfavorable influence of $|Z_\varepsilon|$ on s_p. Additionally, $P_A = s_d^2\omega^2 R_{MR}/8$ shows that a larger R_{MR} is favorable to P_A at certain ω and s_d. Therefore, R_{MRP} is larger than $R_{MR\eta}$,which can be obtained from Eqs. (13) and (20):

$$R_{MRP}^2 - R_{MR\eta}^2 = |Z_\varepsilon|^4/R_p^2 + R_d|Z_\varepsilon|^2/R_p. \tag{28}$$

Equation (28) shows that a large $|Z_\varepsilon|^2$ and a small R_p are favorable to increasing η_{opt} and P_{Aopt}, but also increase the difference between $R_{MR\eta}$ and R_{MRP}, which makes more difficulties to optimize the system performance by considering both η and P_A simultaneously.

D. Optimized operating frequency

The calculated mechanical impedance and efficiency of the system are shown in Fig. 3. The intersections of the dotted lines 1, 2 ($|Z_\varepsilon|$) and the solid line (X_M) in Fig. 3(a) indicate the parallel resonance condition $X_M - 1/\omega C_\varepsilon = 0$. There are three different cases:

(1) When C_ε and M_d are properly chosen, the intersections (a, b, c) of the dotted line 1 and the solid line are close to or even at the resonance frequency f_r of the resonance pipe. Therefore, there are possibilities to obtain large η at the matching frequencies [see solid line 1 in Fig. 3(b)].

(2) The intersection (d) on dotted line 2 at f_{r1} is far from the resonance frequency f_r. Because the load resistance R_{MR} is nearly zero at frequencies far from f_r, the efficiency η is very small at the frequency f_{r1} [see Fig. 3(b)].

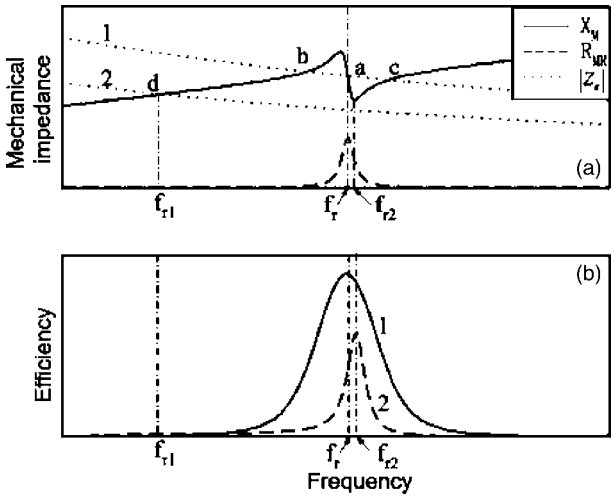


FIG. 3. Mechanical impedance and electroacoustic transfer efficiency versus frequency: (a) mechanical impedance and (b) efficiency.

(3) When $|X_M - 1/\omega C_\varepsilon|$ is minimized, but not zero, at the frequency f_{r2}, which is close to f_r, a peak of η can also be obtained [dashed line 2 in Fig. 3(b)], but the peak value of η is smaller than that of the case (1).

Furthermore, it can also be seen from Eq. (19) that a R_{MR} near zero cannot supply a large P_A. Therefore, the case (1) is the best choice. Then the operating frequency must be chosen near the resonance frequency of the resonance pipe.

E. Optimization of system structure

1. Optimization of equivalent impedance of loudspeaker

As discussed above, small resistances R_d and R_p and large $|Z_\varepsilon|$ are favorable to increasing η and P_A. Besides, other equivalent parameters, such as M_d, C_d, M_p, and C_p, must be properly chosen accordingly to satisfy both reactance matching conditions for obtaining the optimized system performance. Therefore, in order to increase η and P_A, suitable materials and sizes of the piezoelectric disk, the diaphragm, as well as a coupling component with high stiffness, are strongly recommended.

In order to calculate the influence of the parameters of the system, the loudspeaker with the equivalent parameters $C_p = 2.5\times10^{-4}$ m/N, $M_p = 1.76\times10^{-5}$ kg, $R_p = 34$ Ns/m, $C_d = 4.76\times10^{-4}$ m/N, and $R_d = 0.3$ Ns/m, but without a rear cavity, is taken as an example. The lengths of the three parts of the resonance pipe are cold end length 3.0 cm, stack length 0.4 cm, and hot end length 0.8 cm; the radius of the pipe cross-section is 4 mm. The operating gas is the air under the normal temperature and pressure. The mechanical input impedances of the pipe calculated with the fluid impedance formula in pipes with and without stacks[4,6–9] are shown in Fig. 4.

The distributions of η vs. ω and M_d are calculated, as shown in Fig. 5. In Fig. 5(a), taking $C_\varepsilon = 10^{-2}$ m/N, when $M_d = 2.2\times10^{-5}$ kg, $f = 4084$ Hz, and $\alpha = 2.15\times10^4$, η is maximized as 3.5×10^{-7}, which is very small, while in Fig. 5(b), taking $C_\varepsilon = 10^{-6}$ m/N, when $M_d = 1.6\times10^{-3}$ kg, $f = 4043$ Hz, and $\alpha = 2.4\times10^{-3}$, η is maximized about 0.7,

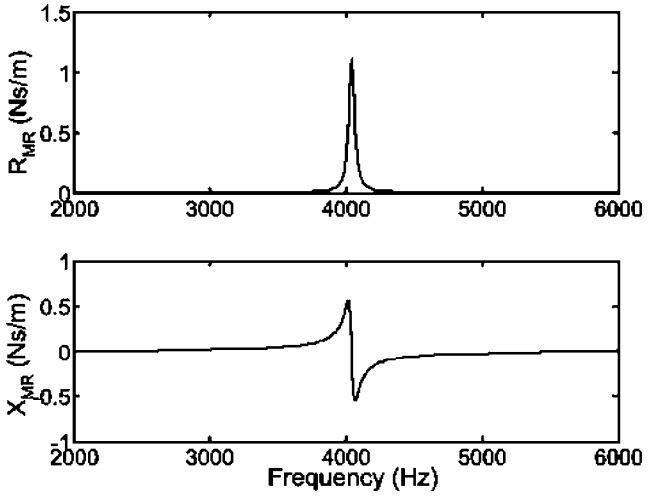


FIG. 4. Input mechanical impedance of resonance pipe.

which is an acceptable efficiency. Comparing both figures demonstrates that the frequencies for the maximum efficiencies are very close to the resonance frequency of the pipe even if C_ε and M_d have great variations. Meanwhile, as C_ε decreases, M_d must be increased in order to satisfy the parallel resonance condition $X_M - 1/\omega C_\varepsilon = 0$. Similar results are obtained by calculating the acoustic output power using the parameters above. Taking $C_\varepsilon = 10^{-2}$ m/N, when $\alpha = 2.22 \times 10^4$, the maximum acoustic power is only about 10^{-9} W, while taking $C_\varepsilon = 10^{-6}$ m/N, when $\alpha = 0.33$, the maximum acoustic power reaches 1.4×10^{-3} W. Therefore, a small C_ε must be selected to obtain larger η and P_A.

Similarly, to obtain an optimized acoustic output power, the reactance matching condition $X_p - 1/\omega C_\varepsilon = 0$ must be also satisfied, so proper equivalent mass M_p and compliance C_p of the piezoelectric disk must be also properly selected according to C_ε.

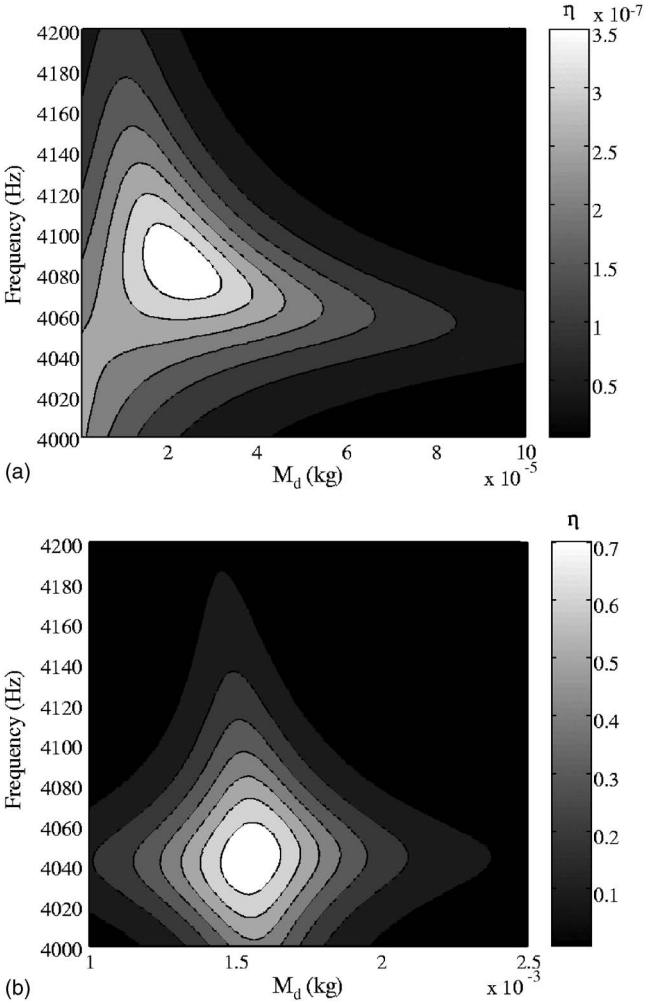


FIG. 5. Distributions of electroacoustic transfer efficiency versus frequency and equivalent mass of vibration diaphragm as $C_d = 4.76 \times 10^{-4}$ m/N: (a) $C_\varepsilon = 10^{-2}$ m/N and (b) $C_\varepsilon = 10^{-6}$ m/N.

2. Optimization of cross-sectional area of resonance pipe

In order to achieve the maximum η or P_A, optimized values of $R_{MR\eta}$ and R_{MRp} shown in Eqs. (13) and (20) must be obtained. Based on $R_{MR} = S^2 R_{AR}$, the cross-sectional area of the resonance pipe can be adjusted to achieve optimized values $R_{MR\eta}$ and R_{MRP} when the temperature, pressure, and gas type in the resonance pipe have been determined. Then the optimized area for η or P_A are

$$S_\eta = \sqrt{R_{MR\eta}/R_{AR}}, \tag{29}$$

$$S_P = \sqrt{R_{MRP}/R_{AR}}. \tag{30}$$

Unfortunately, in a miniaturized thermoacoustic refrigerator, the sizes of pipe cross section and loudspeaker diaphragm are both limited, therefore, in order to obtain $R_{MR\eta}$ and R_{MRP} with small system volume, the resonance frequency of the resonance pipe, which induces the maximum load acoustic resistance R_{AR}, is the most proper operating frequency because η and P_A increase monotonously with R_{MR} when it is smaller than the optimized values $R_{MR\eta}$ or R_{MRP}, which can be seen from Eqs. (12) and (19). In this case, the maximum η and P_A can simultaneously be obtained with the resonance frequency of the pipe f_r:

$$\eta' = 1 \Bigg/ \left[\left(\frac{R_p}{|Z_\varepsilon|^2} \right) R_{MRm} + \left(R_d + \frac{R_p R_d^2}{|Z_\varepsilon|^2} \right) \frac{1}{R_{MRm}} + \left(1 + \frac{2R_p R_d}{|Z_\varepsilon|^2} \right) \right], \tag{31}$$

$$P_A' = \frac{1}{2} |F_e|^2 \frac{R_{MRm}}{(|Z_\epsilon| + R_M R_p / |Z_\varepsilon|)^2}, \tag{32}$$

where R_{MRm} is the value of R_{MR} at the resonance frequency of the pipe.

3. Summary

The electroacoustic transfer efficiency η and acoustic output power P_A under both reactance matching conditions and at the resonance frequency f_r are shown in Fig. 6, which shows the influences of the stiffness of the coupling component and the radius of the resonance pipe r on the system performance. The system parameters are the same as those of Fig. 5. In Figs. 6(a)–6(d), C_ε equals 1×10^{-5}, 1×10^{-6}, 1×10^{-7}, and 1×10^{-8} m/N, respectively.

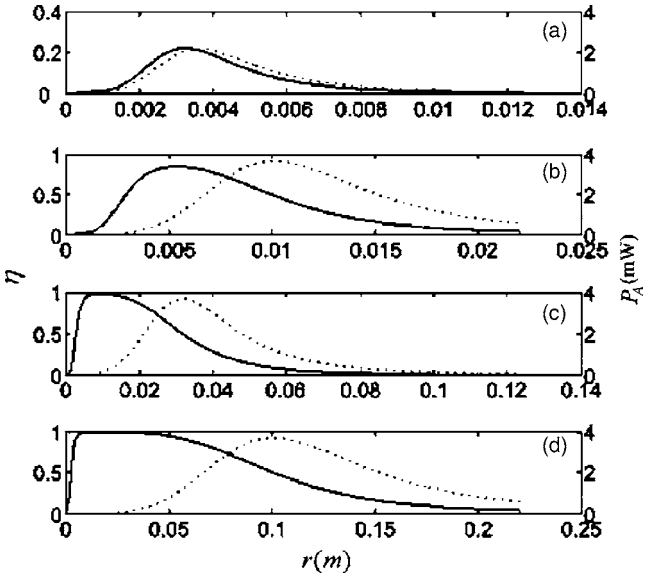


FIG. 6. Electroacoustic transfer efficiency (solid line) and acoustic output power (dotted line) versus radius of resonance pipe: (a) $C_\varepsilon=10^{-5}$ m/N; (b) $C_\varepsilon=10^{-6}$ m/N; (c) $C_\varepsilon=10^{-7}$ m/N; and (d) $C_\varepsilon=10^{-8}$ m/N.

In Fig. 6(a), with a small $|Z_\varepsilon|$, the optimized pipe radii r_η and r_P for both η and P_A, respectively, are almost the same because $R_{MR\eta}$ and R_{MRP} are close to each other, which is in agreement with that shown in Eq. (28). However, both η_{opt} and P_{Aopt} are small because of a small $|Z_\varepsilon|$.

In Fig. 6(d), with a very large $|Z_\varepsilon|$, both η_{opt} and P_{Aopt} are large, but r_η and r_P are very different because $R_{MR\eta}$ is much smaller than R_{MRP}. It can be seen that a large η can be obtained with a wide range of radius (from about 1 to 10 cm), however, P_{Aopt} is obtained at $r_P \simeq 10$ cm and becomes very small when r decreases to about 5 cm, which is still too large in a miniaturized refrigerator. Furthermore, in order to obtain both reactance matching conditions, M_d and M_p must be 0.15 kg, which is also unavailable in the miniaturized system. Thus, although a coupling component with large stiffness is favorable to increasing both η and P_A, it simultaneously requires large M_d, M_p, $R_{MR\eta}$, and R_{MRP} to achieve the optimized conditions, i.e., "heavy" diaphragm and piezoelectric disk, and a large cross-sectional area of the resonance pipe and diaphragm are required, which is unacceptable in the fabrication of a miniaturized thermoacoustic refrigerator. Therefore, a coupling component with extremely large stiffness or even direct connection of the piezoelectric disk to the diaphragm is not a good choice, because proper M_d, M_p, r_η, and r_P are all unavailable in this case.

In Fig. 6(b), $\eta_{opt}=0.85$ with $r_\eta=0.5$ cm and $P_{Aopt}=3.65\times10^{-3}$ W with $r_P=1$ cm; in Fig. 6(c), $\eta_{opt}=0.98$ with $r_\eta=1$ cm and $P_{Aopt}=3.68\times10^{-3}$ W with $r_P=3$ cm. The figures show that, with proper $|Z_\varepsilon|$, both large efficiency and power can be obtained with acceptable pipe radius. However, the optimized radii r_η and r_P cannot be obtained simultaneously. A trade-off between η and P_A should be taken, such as in Fig. 6(b). The proper radius should be about 0.7 cm, while in Fig. 6(c), the proper radius is about 2.5 cm. It must be stated that using an operating gas with large acoustic impedance and increasing the operating pressure can increase R_{AR} and subsequently decrease r_η and r_P. Therefore, a larger stiffness of the coupling component is available and better system performance can be achieved.

Generally, a proper stiffness of the coupling component is significant to the system performance, and a trade-off between the system performance and miniaturization must be considered.

IV. CONCLUSIONS

An integrated equivalent circuit model is presented to analyze the performances of thermoacoustic refrigerators driven by piezoelectric loudspeakers. The expressions of the electroacoustic transfer efficiency and acoustic output power of the piezoelectric acoustic source are obtained using the equivalent circuit theory. The influences of the equivalent parameters of the loudspeaker, the resonance pipe, and the operating frequency on the electroacoustic transfer efficiency and acoustic output power are evaluated and optimization conditions are subsequently achieved:

(1) A reactance matching condition $X_M-1/\omega C_\varepsilon=0$ of the parallel branch Z_R must be firstly satisfied to obtain a high efficiency and, furthermore, a reactance matching condition of the piezoelectric disk $X_p-1/\omega C_\varepsilon=0$ should also be satisfied in order to achieve an optimized acoustic power. Both reactance matching conditions can be obtained by choosing the proper piezoelectric disk, vibration diaphragm, rear cavity, and coupling component.
(2) In a miniaturized thermoacoustic refrigerator, a small radius of the resonance pipe or the diaphragm of the loudspeaker are required; the proper operating frequency must be close to the resonance frequency of the pipe, at which the acoustic load resistance is large, which can decrease the required cross-sectional areas of the resonance pipe and the diaphragm. Furthermore, an operating gas with large impedance and a large operating pressure are both favorable to the system miniaturization and performance.
(3) Generally, a coupling component with a high stiffness, a piezoelectric disk, and a vibration diaphragm with low resistances are favorable to achieving a high electroacoustic transfer efficiency and a large acoustic output power. However, they increase the difference between the optimized load acoustic resistances for getting large efficiency and power; then a trade-off between the electroacoustic transfer efficiency and the acoustic output power must be selected.
(4) For a coupling component of high stiffness, which increases the stroke of the vibration diaphragm and then the electroacoustic transfer efficiency and acoustic output power, large equivalent masses of the vibration diaphragm, piezoelectric disk, and large areas of the pipe cross section and vibration diaphragm are all required to obtain the optimized conditions, which are unfavorable to the system miniaturization. Therefore, there is a trade-off between the performance and the miniaturization of the system.

ACKNOWLEDGMENT

The project is supported by the National Natural Science Foundation of China, Grant No. 10374051.

[1]T. J. Hofler and J. A. Adeff, "A miniature thermoacoustic refrigerator for ICs," in *Proceedings of 17th International Congress on Acoustics*, Roma, Italy, 2001.
[2]O. G. Symko and A. R. Ehab, "High frequency thermoacoustic refrigerator," US Patent No. 6,574,968 (2003).
[3]O. G. Symko, A. R. Ehab, D. J. Zhang, and T. Klein, "High frequency thermoacoustic refrigerator," US Patent No. 6,804,967 (2004).
[4]B. Lihoreau, P. Lotton, M. Bruneau, and V. Gusev, "Piezoelectric source exciting thermoacoustic resonator: analytical modelling and experiment," Acta. Acust. Acust. **88**, 986–997 (2002).
[5]R. S. Wakeland, "Use of electrodynamic drivers in thermoacoustic refrigerators," J. Acoust. Soc. Am. **107**, 827–832 (2000).
[6]W. P. Arnott, H. E. Bass, and R. Raspet, "General formulation of the thermoacoustics for stacks having arbitrarily shaped pore cross sections," J. Acoust. Soc. Am. **90**, 3228–3237 (1991).
[7]R. Raspet, J. Brewster, and H. E. Bass, "A new approximation method for thermoacoustic calculations," J. Acoust. Soc. Am. **103**, 2395–2402 (1998).
[8]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1977), Chap. 9, pp. 200–216.
[9]L. Fan, B. R. Wang, T. Jin, H. Zhang, and S. Y. Zhang, "Performance optimization of thermoacoustic systems by measurement with two-microphones method," Acta. Acust. Acust. **91**, 831–839 (2005).

Brake squeal as dynamic instability: An experimental investigation

Francesco Massi[a)]
Department of Mechanics and Aeronautics, University of Rome "La Sapienza", Rome Italy, and Contact and Solid Mechanics Laboratory, INSA of Lyon, Lyon, France

Oliviero Giannini
Department of Mechanics and Aeronautics, University of Rome "La Sapienza", Rome Italy

Laurent Baillet
LGIT, Laboratory of Geophysics and Tectonophysics, University Joseph Fourier, Grenoble, France

(Received 24 February 2006; revised 29 May 2006; accepted 24 June 2006)

This paper presents an experimental analysis performed on a simplified brake apparatus. In past years a common approach for squeal prediction was the complex eigenvalues analysis. The squeal phenomenon is treated like a dynamic instability: when two modes of the brake system couple at the same frequency, one of them becomes unstable, leading to increasing vibration. The presented experimental analysis is focused on correlating squeal characteristics with the dynamic behavior of the system. The experimental modal identification of the setup is performed and different squeal conditions and frequencies are reproduced and analyzed. Squeal events are correlated with the modal behavior of the system as a function of the main parameters. A clear distinction between squeal events involving the dynamics of the pad and squeal events involving the dynamics of the caliper is performed. The effect of the adding of damping is also investigated on the squeal phenomenon. Two opposite roles of the modal damping are described: a large modal damping can either prevent the rise of squeal instabilities or enlarge the squeal propensity of the brake apparatus. The robustness of the obtained squeal events permits a further analysis on the triggering mechanism of the squeal instability during braking. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2228745]

PACS number(s): 43.40.At, 43.50.Lj, 43.20.Ks [DF] Pages: 1388–1398

I. INTRODUCTION

Disk brake noise continues to be an object of investigation for automotive manufacturers and researchers.[1] Because of the complexity of the problem and the need for estimating the squeal tendency during brake design, many analytical and numerical[2,3] approaches have been proposed. Mills[4] and Fosberry and Holubecki[5,6] tried to correlate the occurrence of squeal with a negative slope of the friction coefficient-relative velocity curve, while Spurr[7] proposed his sprag-slip theory, followed later by Earles *et al.*[8,9] and Jarvis and Mills.[10] In 1972, North[11,12] published the first experimental work on a real brake apparatus. Akay *et al.*,[13] Tuchinda *et al.*,[14,15] Allgaier *et al.*,[16,17] and Tarter[18] conducted extensive analyses of the beam-on disk setup. Nowadays, the modal coupling[11–13] (mode lock-in) between two system modes is one of the most accepted theories. The complex eigenvalues analysis is a popular numerical tool for squeal instability prediction.[19,20] This paper presents an experimental analysis aimed at verifying if this approach can predict efficiently the squeal occurrence in a brake. Brake squeal is a nonlinear phenomenon, characterized by friction material and contact nonlinearities. Thus, the modal approach can only predict the onset of the squeal instability, when it is still in linear conditions.

The experimental analysis is performed on a simplified setup, named TriboBrake COLRIS (COllaboration Lyon-Rome for Investigation on Squeal), that has three main advantages.

(1) it can generate squeal noise easily;
(2) it has a dynamic behavior that can be easily measured; and
(3) it has a dynamic behavior that can be easily adjusted.

The third point is particularly important to correlate the dynamics of the system and the squeal in a "what-if" analysis.

The modal behavior of the setup is first presented. The dynamics of the system is studied by considering three main substructures of the brake (the caliper, the rotor, and the pad). Different approaches, studied to shift the natural frequencies of the system, are used to find different instability conditions, and the squeal conditions are related to a specific dynamic configuration of the system with specific values of the parameters, so that they are easily reproducible.

The experimental analysis shows that squeal occurrence in the experiments is always correlated to a modal coupling between one of the modes of the rotor and a mode of either the pad or the caliper. Therefore, to find an instability condition it is necessary to shift a mode of the system so that it falls close to another. In the paper the coincidence between two modes is called "modal tuning."

[a)]Electronic mail: francesco.massi@uniromal.it

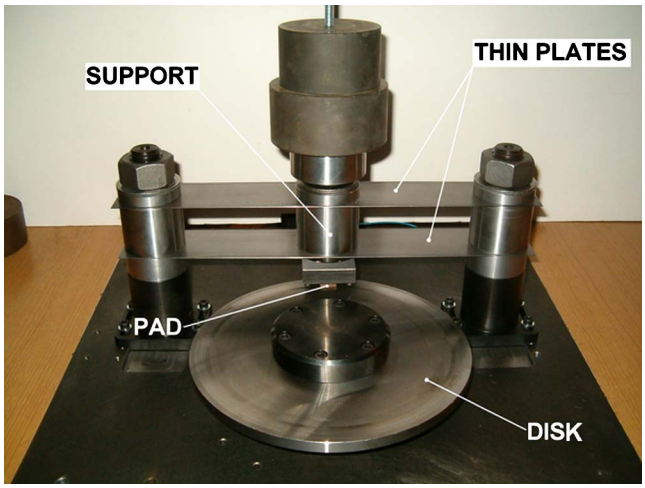


FIG. 1. (Color online) Experimental setup.

A further analysis on the effect of the modal damping on the squeal propensity is performed. Some comments and concluding remarks are made.

II. EXPERIMENTAL SETUP

Since a real brake apparatus is characterized by geometry and dynamics that can be hardly controlled or understood, an experimental and theoretical study of a simplified experimental setup is preferred. The setup consists of a rotating-disk (the disk brake rotor) and a small friction pad pushed against the disk by weighs positioned on a rigid support (Fig. 1). The disk is made of steel (internal diameter 100 mm, external diameter 240 mm, thickness 10 mm) and is assembled with the shaft by two hubs of large thickness that assure a rigid behavior of the connection in the frequency range of interest. The velocity of the motor can be adjusted to have a disk velocity between 5 and 100 rpm. The brake pad is made of commercial brake friction material, obtained by machining standard brake pads. Reduced pad dimensions are adopted to simplify and control easily its dynamics. The support may host pads with different dimensions; when not otherwise specified the dimensions are $10 \times 10 \times 10$ mm^3.

The support (the central cylindrical body in the figure) is also made of steel and its shape is chosen to simplify its dynamic behavior. The normal force between pad and disk (braking load) can be adjusted by adding weights on the top of the support, between 25 (the weight of the support) and 250 N. Two thin plates hold the pad support in the tangential direction while allowing a negligible stiffness in the normal direction. Adjusting the normal load with weights (Fig. 1) allows the pad surface to follow the disk oscillations, due to the possible not-perfect planarity of the disk, and assures a constant value of the contact normal force. A triaxial force transducer is placed between the pad and the support. The transducer allows the time history of the normal and friction forces to be measured. It is important to note that these forces are not measured on the real contact surface, but above the brake pad. Thus, the pad dynamics influences the measured forces.

III. SETUP DYNAMICS

The main objective of this paper is to show how the dynamics of a brake system influences squeal occurrence and squeal frequencies. Therefore, the investigation of the setup dynamics is the first step of the work. Particular attention is focused on the bending modes of the disk (in the normal direction) and the bending modes of the support and the pad (in the tangential direction). It was shown in previous experiments[21] that these are the modes involved in squeal phenomena. Moreover, both a finite-element mode, (FEM) analysis and measured frequency response function (FRFs) in the plane direction show that, in this simplified setup, the in plane modes of the disk are not involved in the squeal phenomenon.

Three different substructures are considered in the analysis: the disk, the support, and the pad. In the dynamics of the assembled system it is possible to recognize the combination of the dynamics of the disk and of the support (because of the reduced contact surface between the two substructures and the consequent low coupling between them). Therefore, "disk modes" refers to the modes involving bending vibration of the disk, being the largest part of the energy concentrated in the disk. As well, "support modes" refers to the modes involving tangential vibration of the support. A further analysis allows one to recognize the influence of the pad dynamics on the dynamics of the assembled system.

The disk modes are characterized by nodal diameters and nodal circumferences: the (n,m) mode of the disk is characterized by n nodal circumferences and m nodal diameters. The disk is characterized by an axial symmetry: therefore, the modes of the disk are generally double modes. Due to the contact with the pads, the disk loses its axial symmetry and the double modes of the disk split at two different frequencies (Fig. 2):

(i) mode $(n,m-)$, a nodal diameter falls in the contact area (sine mode); and
(ii) mode $(n,m+)$, an antinode falls in the contact area (cosine mode).

The support modes are analyzed by an SIMO (single-input–multi-output) analysis, exciting the support in the tangential direction, close to the contact area. In the frequency range of interest two rigid and three bending tangential modes of the support are recognized. This analysis is performed with the coupled system, when the disk does not rotate. However, the same peaks in frequency and same deformed shapes are found during brake simulations.[22] Table I lists the natural frequencies of the system obtained by experimental modal analysis (EMA), with a contact load equal to 25 N, and a friction pad with 10×10 mm contact surface is mounted.

The third substructure investigated is the friction pad. Its dynamics is easily recognized in the assembled dynamics. Figure 3 shows the PSD (power spectral density) of the pad acceleration in the tangential direction during brake simulation (gray line). The first three peaks in frequency correspond to three support modes. The others two peaks at 4 and 11.1 kHz correspond to modes of the pad. A second test was made by dragging the pad, disassembled from the disk and

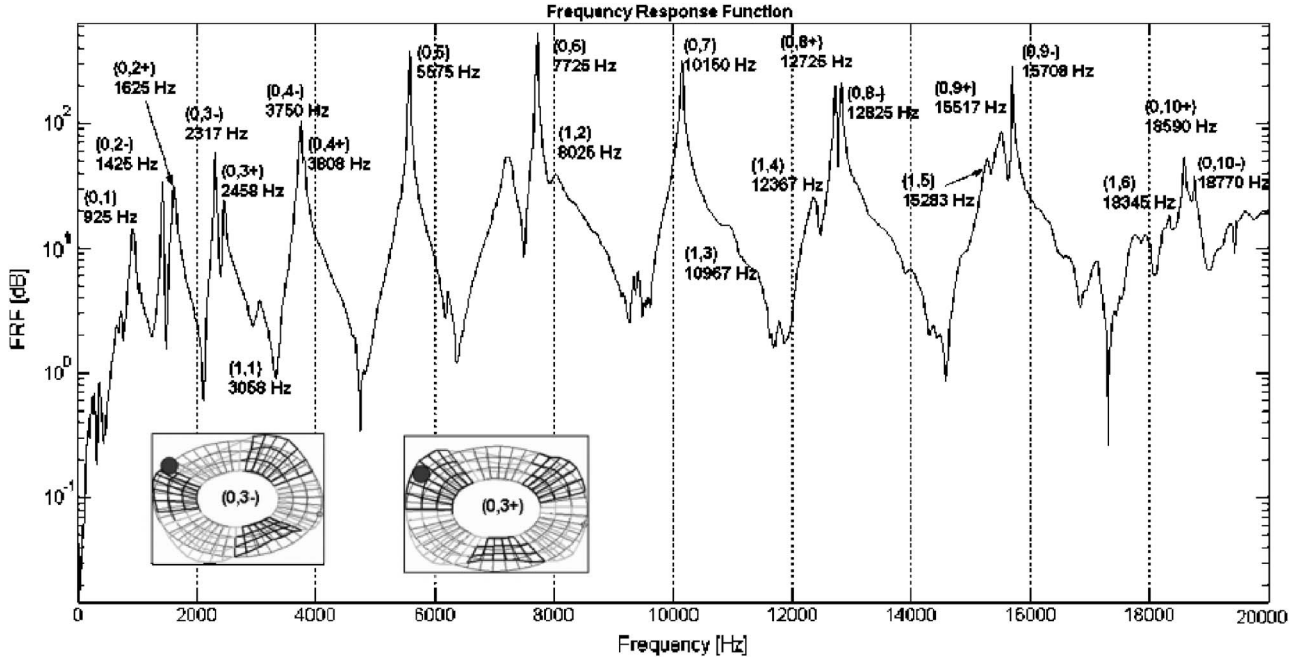


FIG. 2. FRF of the setup measured on the disk periphery when the pad is in contact with the disk (coupled system) and samples of the disk deformed shapes (0,3) and (0,3+).

the support, on a rigid surface. The black line in Fig. 3 shows the acceleration PSD during this test. Only the two peaks related to the pad modes appear. An FE modal analysis allows us to identify two pad modes characterized by the pad deformation in the tangential direction and large tangential oscillations at the contact surface, at the same frequencies obtained experimentally.

The proposed design allows us to control the dynamics of the setup by changing only a few parameters. The variation of the normal load applied on the top of the support shifts the natural frequencies of the system. In particular, the double modes of the disk split at two different frequencies, and the split increases when the normal contact load increases. The dimension of the pad influences mostly its natural frequencies. The stiffness of the thin plates affects the natural frequencies of the support. The insertion of damping material between the thin plates and the support introduces modal damping and affects the natural frequencies of the pad.

TABLE I. System natural frequencies and modal damping of the setup with normal load equal to 25 N.

Mode	Frequency [Hz]	Hysterical damping %
I support	489	7,27
(0,1+)	925	5,36
(0,2−)	1 425	1,67
(0,2+)	1 625	3,69
II support	2 091	0,72
(0,3−)	2 317	1,18
(0,3+)	2 458	2,02
III support	2 912	3,99
(1,0)	3 058	2,11
(0,4−)	3 750	1,31
(0,4+)	3 808	0,67
IV support	5 146	2,09
(0,5−)	5 575	0,49
(0,5+)	5 589	0,32
(1,1)	7 217	2,35
V support	7 717	0,75
(0,6)	7 725	0,26
(1,2)	8 025	3,06
(0,7+)	10 088	0,37
(0,7−)	10 141	0,38
(1,4)	12 367	1,07
(0,8+)	12 725	0,27
(0,8−)	12 825	0,17
(1,5)	15 283	0,58
(0,9+)	15 517	0,54
(0,9−)	15 708	0,13

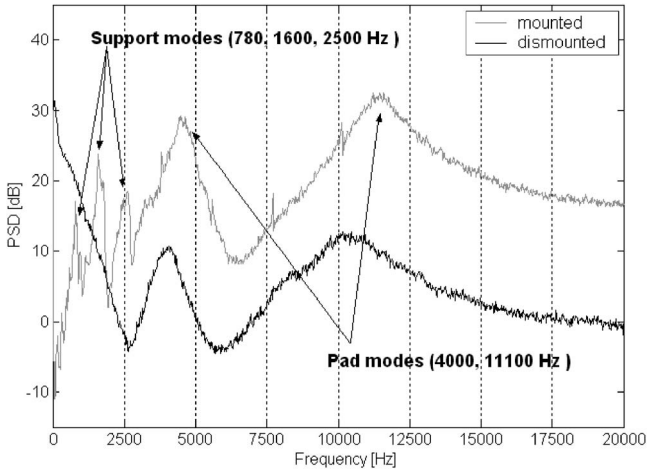


FIG. 3. PSD of the pad acceleration during brake simulation (gray), and in disassembled condition (black). Three peaks are due to the modes of the support and two peaks are due to the modes of the pad.

Because of the aim of this paper, it is useful to reproduce many different squeal events. The proposed experimental setup achieves this goal: in fact, by controlling the driving parameters,[23] it allows different squeal conditions and reproduces them easily.

IV. SQUEAL PHENOMENA

This section presents the results of an extensive experimental investigation aimed to find as many squeal frequencies as possible by changing the operational parameters of the setup. During this investigation the dynamics of the system was monitored to relate its variation to the rising of instabilities.[24] Five different squeal frequencies are found: 1566, 2467, 3767, 7850, and 10 150 Hz. These squeal conditions are obtained for defined values of the driving parameters, and all of them are easily reproducible.

Both the pad and the support present modes characterized by tangential vibrations along the contact surface; experiments show that the disk dynamics can couple either with the dynamics of the pad or with the dynamics of the support, bringing in both cases squeal instabilities.

The experimental acquisitions are performed by laying the support over the rotating disk. The disk velocity is maintained at 10 rpm. The global normal and tangential forces are acquired between the pad and support. The tangential acceleration on one side of the pad is measured. A microphone placed at 200 mm from the contact point measures the sound-pressure level. The squeal events are recognized by the increase of vibrations of the system and by the harmonic sound emission that may reach 100 dB.

A. Pad-disk squeal coupling

Each braking phase starts by laying the support over the disk. The normal and friction forces show a starting ramp due to the initial contact between pad and disk.[25] After, the system vibrations and the sound pressure start to increase. The PSD of the pad acceleration and the sound-pressure level reveal the harmonic nature of the system vibrations. After a small time (less than a second) the system reaches its limit cycle.

Figure 4 presents a dynamic characterization of the pad when squeal occurs at 3767 Hz: the gray lines are the PSD of the pad acceleration in the tangential direction for different values of the normal load from 250 to 45 N, while the dashed line is the measured FRF of the disk, obtained with normal load equal to 45 N. The amplitude of the peak related to the pad mode decreases with the decrease of normal load because the pad excitation in the tangential direction, due to the friction force, decreases proportionally to the normal load. By lowering the normal load, the first mode of the pad moves to lower frequencies and it gets close to the $(0,4+)$ mode of the disk. This is caused by the lower contact stiffness between the pad and the disk.

It is important to notice that only when the natural frequency of the pad is close enough to the one of the disk, squeal happens.[26]

In Fig. 4 only the black line (PSD for 45 N of normal load) presents the frequency spectrum characteristic of the

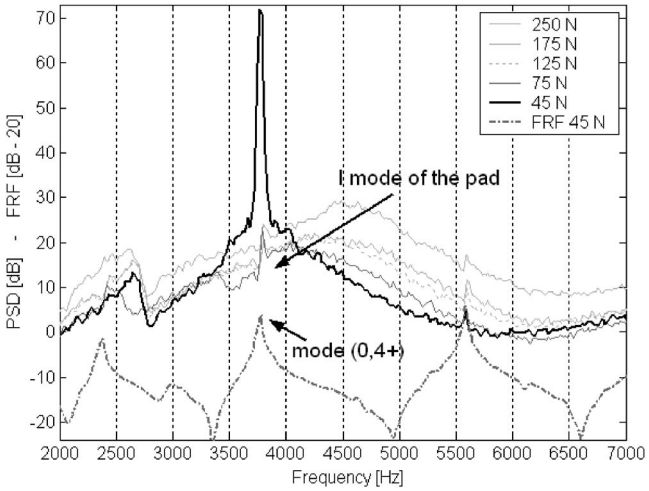


FIG. 4. PSD of pad acceleration in function of the normal load, and FRF at the disk surface. By decreasing the load the natural frequency of the pad mode decreases up to the tuning with the natural frequency of the disk mode (squeal frequency).

squeal phenomenon. Similar plots can be obtained by varying other parameters. This means that squeal instability can occur only when a coincidence of two system modes is obtained, as predicted by the complex modal analysis and by the lock-in theory. Two other squeal events due to the coincidence between a natural frequency of the pad and a natural frequency of the disk are obtained at 7850 Hz when the frequency second mode of the pad tunes with the frequency of the $(0,6+)$ mode, and at 10 150 Hz when the frequency of the second mode of the pad tunes with the frequency of the $(0,7+)$ mode. The former is obtained when the normal load is equal to 25 N, the dimensions of the pad is equal to 8 $\times 8$, and the thickness of the thin plates is equal to 0.5 mm; the second is obtained with a normal load equal to 45 N when introducing a thin layer of rubber between the support and the thin plates. With this expedient the tangential stiffness of the pad mode decreases and its natural frequency shifts to a lower frequency, becoming close to the $(0,6+)$ mode.

B. Support-disk squeal coupling

Other squeal frequencies are obtained by the coupling of modes of the support characterized by tangential deformation and modes of the disk characterized by bending vibrations in the normal direction. Like in the previous cases, squeal arises only with the tuning between the two natural frequencies. The modes of the support involved in squeal events are the second and the third mode. In fact, these modes are characterized by the largest vibration at the contact end, where coupling between the tangential vibrations of the pad surface and the normal vibrations of the disk happens. Figure 5(a) shows the PSD of the tangential acceleration of the pad when squeal at 1566 Hz occurs. The normal load is fixed at 200 N; the dimensions of the contact surface is 10×10 mm^2. In these conditions the natural frequency of the second mode of the support tunes with the natural frequency of the $(0,2+)$ mode of the disk, while there is no tuning between the third mode of the support and the $(0,3+)$ mode of the disk. There-

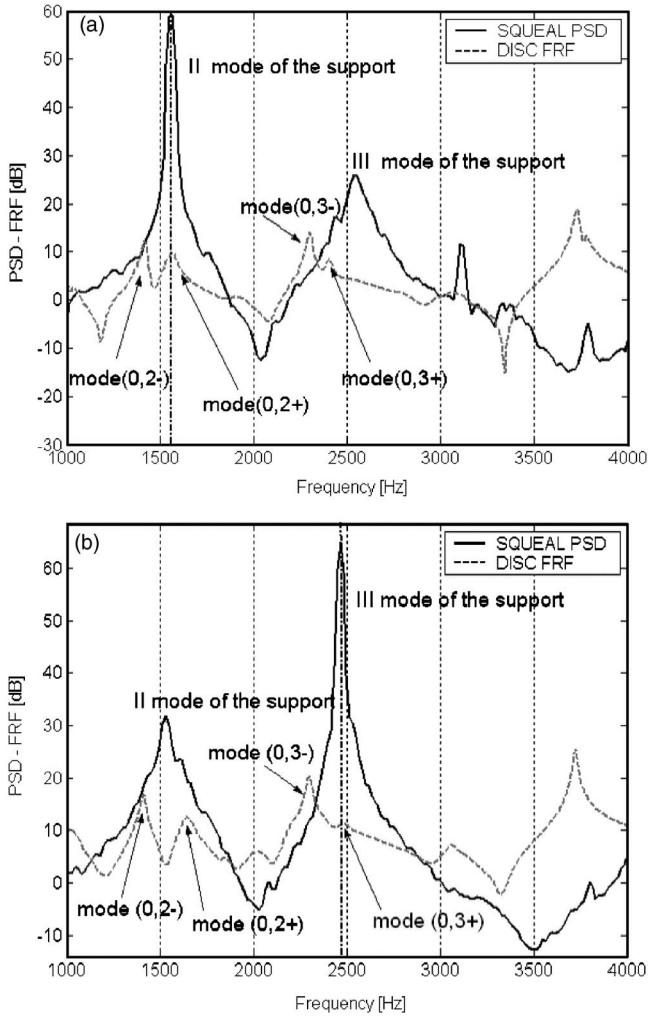


FIG. 5. (a) PSD of the pad acceleration during squeal at 1566 Hz and FRF of the disk (tuning between the second mode of the support and the $(0,2+)$ mode of the disk); (b) PSD of the pad acceleration during squeal at 2467 Hz and FRF of the disk (tuning between the third mode of the support and the $(0,3+)$ mode of the disk).

fore squeal occurs at the frequency of coincidence between the former modes (the dotted vertical line in the graphics).

Notice that squeal at 2467 Hz [Fig. 5(b)] occurs when the normal load is equal to 250 N. Increasing the load from 200 to 250 N, the natural frequencies of the support decrease and tune with the $(0,3+)$ mode of the disk. Consequently, these two modes couple together and cause squeal, while the second mode of the support and the $(0,2+)$ of the disk cannot couple together anymore because they are not close enough in frequency.

Squeal events involving the support disk coupling occur only when an exact tuning between the disk and the support natural frequencies occurs. In fact, just by changing the weight configuration slightly (e.g., positioning them just a little off the support axis) it is possible to obtain the results shown in Fig. 6. The third natural frequency of the support moves slightly (about 40 Hz) away from the $(0,3+)$ mode of the disk, and squeal does not occur.

Figure 7 shows the disk modes involved in squeal events during experiments, along with the frequency ranges covered by the support and pad modes when changing the driving

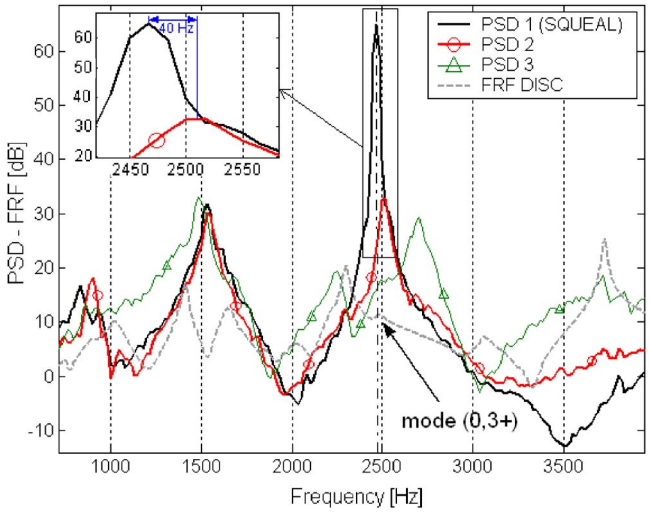


FIG. 6. (Color online) PSDs of the acceleration of the pad with three different weight dispositions and FRF measured on the disk. Squeal appears only for precise tuning between the natural frequencies. For differences above 40 Hz squeal does not occur.

parameters. The bending modes of the disk in the range of frequencies covered by the support and pad modes may be involved in instability, and they become unstable when their frequencies tune with the corresponding support or pad natural frequencies. The $(0,5+)$ mode frequency of the disk does not fall in any range covered by the tangential modes of the pad or support and, consequently, squeal does not occur.

V. MODAL DAMPING EFFECTS ON SQUEAL INSTABILITY

This section describes two different, almost opposite, roles that the modal damping plays on squeal phenomenon. Experiments show that a higher modal damping can either facilitate the rise of squeal or reduce the instability amplitude and suppress it completely.

Moreover, a further proof that squeal is a dynamic instability, related to the dynamics of the mechanism, is brought by triggering its rise with an external impulse.

A. Squeal suppression with damping

In order to stop or prevent squeal events thin layers of rubber are introduced between the support and the thin plates. The rubber introduces a high modal damping to the third mode of the support (2.7 kHz) that is characterized by its bending deformation. Figure 8(a) shows the PSD of the acceleration of the pad[24] when the values of the parameters are chosen to avoid squeal. As mentioned, the frequency peak at 2.7 kHz is extremely reduced by introducing rubber layers (from 28 to 8 dB). Figure 8(b) suggests that the introduction of modal damping prevents the unstable coupling. In fact, once the damping layer is introduced, it is not possible to obtain this squeal condition by any change in the parameters.

On the contrary, when the values of the parameters are set to have coupling between the first mode of the pad and the $(0,4+)$ mode of the disk, the introduction of the rubber layers does not affect the rise of squeal, that occurs as easily

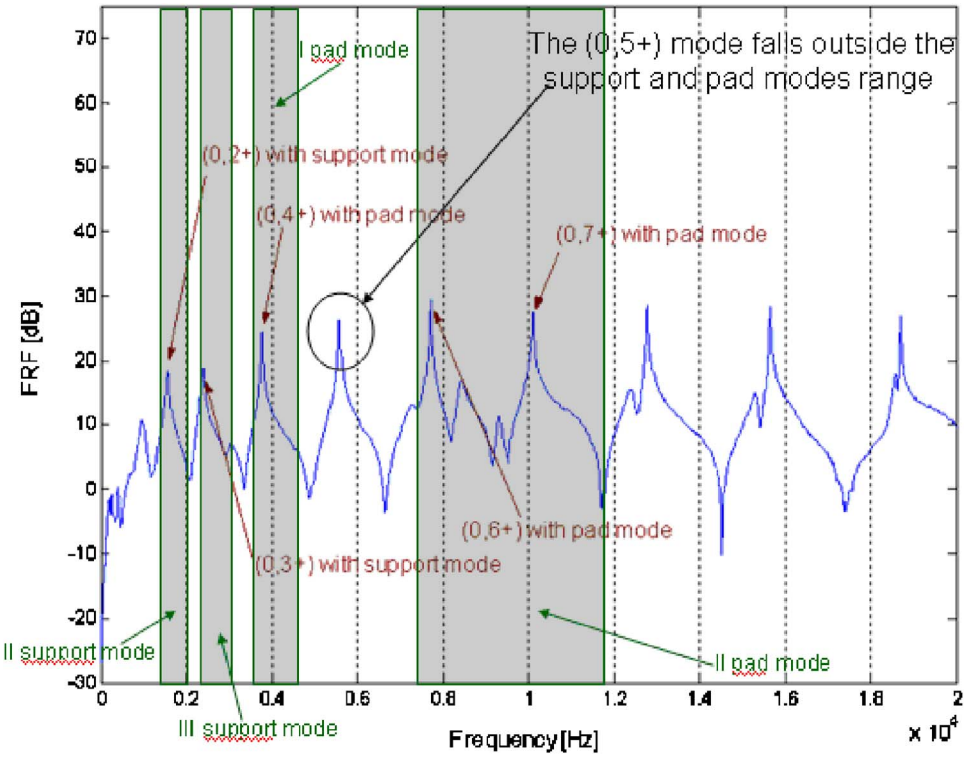


FIG. 7. (Color online) Modes involved in squeal instabilities. All the modes $(0,m+)$ of the disk that fall in the ranges of frequencies covered by the tangential modes of the pad and support are involved in squeal.

as without rubber [Fig. 8(c)]. This behavior of the setup is a demonstration of the fact that the disk modes can couple with modes of either the pad or the caliper. Introducing solutions to avoid squeal caused by the coupling between modes of two substructures (support and disk in this case) can be useless to prevent squeal that can arise by the coupling between modes of two other substructures (pad and disk). Thus, brake design should take into account the dynamics of both the caliper and the pad. The introduction of thin plates between the back plate of the pad and the piston, adopted on commercial brakes to prevent squeal, can introduce high contact damping[27] in the tangential direction that avoids the modes of the caliper to couple with the modes of the disk, but cannot affect the role of the dynamics of the pad.

A different effect of the introduction of the rubber layer is the shift of the mode of the pad toward lower frequencies, due to the decrease of the tangential stiffness at the connection between the pad and the support. This effect is used to shift the natural frequency of the second mode of the pad from the coupling frequency with the $(0,7+)$ mode of the disk (squeal at 10 150 Hz) into the coupling frequency with the $(0,6+)$ mode of the disk (squeal at 8850 Hz), as shown in Fig. 8(d).

B. Increase of squeal probability with modal damping

The previous section shows how an increase of modal damping can prevent the damped mode to participate in the squeal coupling. The following experiments will demonstrate also that a mode characterized by a large modal damping can be more easily involved in squeal than a mode characterized by a small modal damping.

The modes of the support are characterized by a smaller modal damping than the modal damping of the pad. This is due to the large damping coefficient of the friction materials and appears on the PSDs of the pad acceleration showing two highly damped peaks (~4 and 11.1 kHz), corresponding to the two tangential modes of the pad (Fig. 3).

As reported in Sec. IV B, when squeal occurs by coupling a mode of the support with a mode of the disk, the instability rises only if there is an exact tuning between the two natural frequencies (Fig. 6). On the contrary, when a mode of the pad is involved in squeal (coupled with a mode of the disk) the behavior of the instability is different and it can occur still when there is not exact tuning.

When there is exact tuning between the two modes,[24] squeal occurs as soon as the disk starts to rotate; however, if the two modes are not perfectly tuned, squeal can still occur. First, the parameters of the setup are set to have exact tuning between the $(0,4+)$ mode of the disk and the first mode of the pad (squeal at 3767 Hz). Then, the normal load is changed from the tune-in value (45 N) and the brake simulation starts in "silent conditions," i.e., without squeal emission.

In "silent conditions" a hammer impulse is provided to the disk in the normal direction. Depending on the tuning level between the two modes (i.e., the distance between the two natural frequencies), three behaviors, described below, are found.

When the normal load increases from 45 to 48 N (the two modes are off-tuning of about 40 Hz), squeal starts only after the impulse is applied to the disk, as shown in Fig. 9 and, after the impulse, the squeal does not stop. This means that with a highly damped mode (i.e., the mode of the pad),

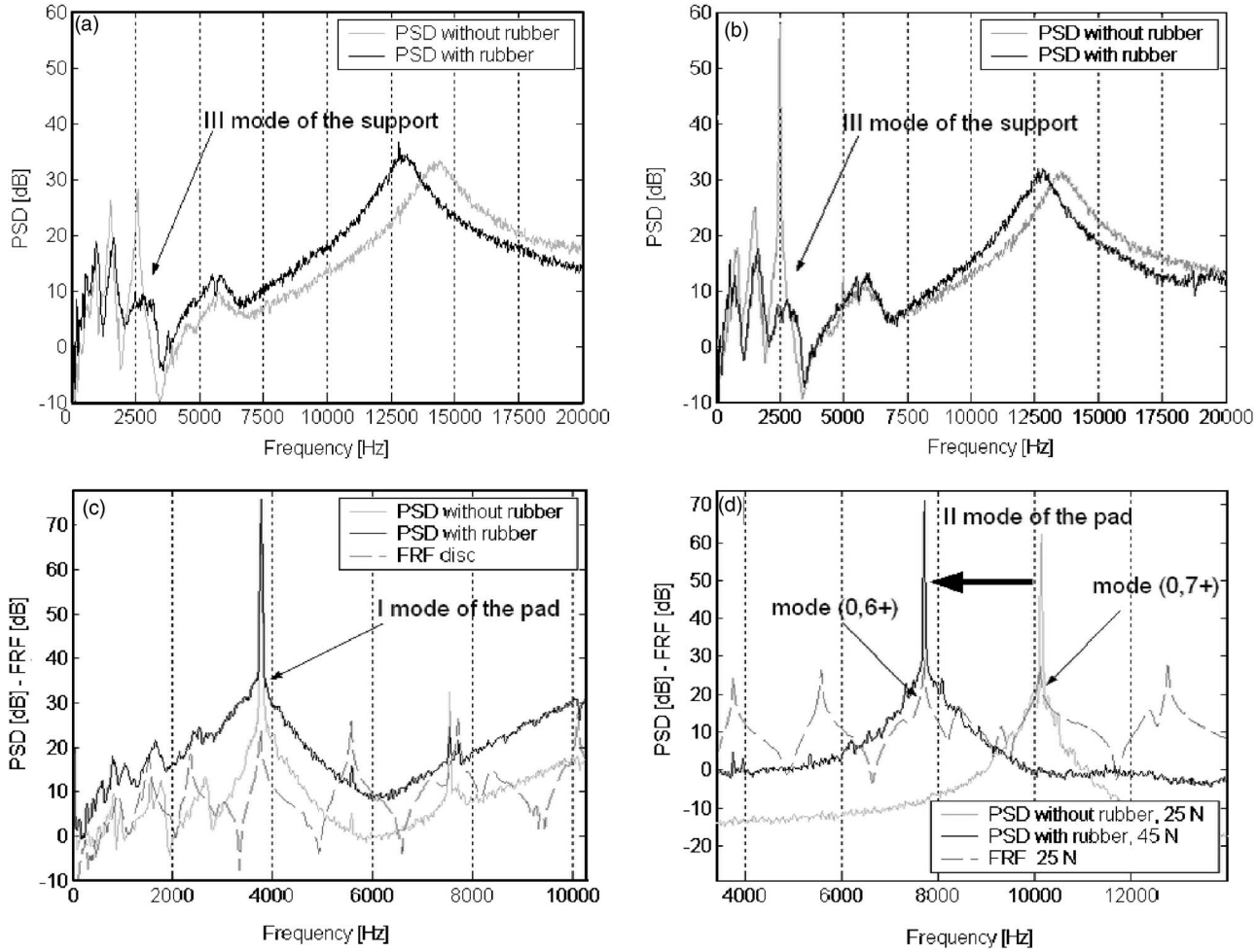


FIG. 8. PSD of the pad acceleration with and without rubber layers, respectively, when (a) parameters are set to not have squeal; (b) parameters are set to have coupling between the third mode of the support and the $(0,3+)$ mode of the disk (squeal at 2467 Hz); (c) parameters are set to have coupling between the first mode of the pad and the $(0,4+)$ mode of the disk (squeal at 3767 Hz); (d) PSD of the pad velocity when parameters are set to have coupling between the second mode of the pad and the $(0,6+)$ mode of the disk (squeal at 7850 Hz, black), with rubber; and when parameters are set to have coupling between the second mode of the pad and the $(0,7+)$ mode of the disk (squeal at 10 150 Hz, gray), without rubber.

the tune-in range of the mode is wide and squeal can still occur even if the two natural frequencies do not coincide exactly.

The impulse is given in the opposite side of the contact with the pad, in order to excite all the modes with an antinode at the contact point. Figure 9 shows that, after the impact, the squeal starts and goes on until the disk stopped.

Figure 10 shows these results and three different key steps are identified: the gray band on the time history underlines the time range of the acceleration where the FFT of the tangential acceleration of the pad is calculated.

(i) Phase (1) before the impulse: a wideband noise with only two peaks related to the support modes and two highly damped peaks due to the pad modes is present [Fig. 10(a)];
(ii) Phase (2) during the impulse: the FFT shows all the peaks due to the bending modes of the disk with an antinode at the contact surface [Fig. 10(b)];
(iii) Phase (3) after the impact: only the frequency peak (squeal at 3767 Hz), coincident with the $(0,4+)$ bending mode of the disk tuned with the first mode

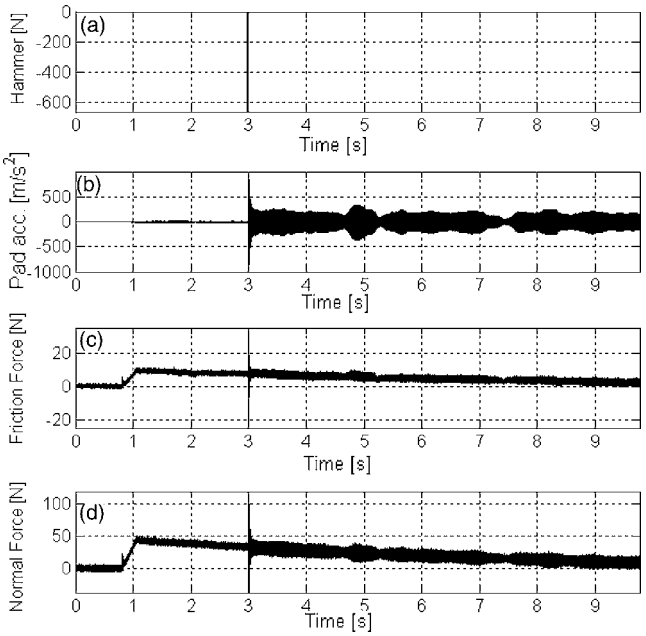


FIG. 9. Squeal vibration starts after an impulse normal to the disk surface: (a) hammer impulse; (b) tangential acceleration of the pad; (c) friction force; (d) normal force.

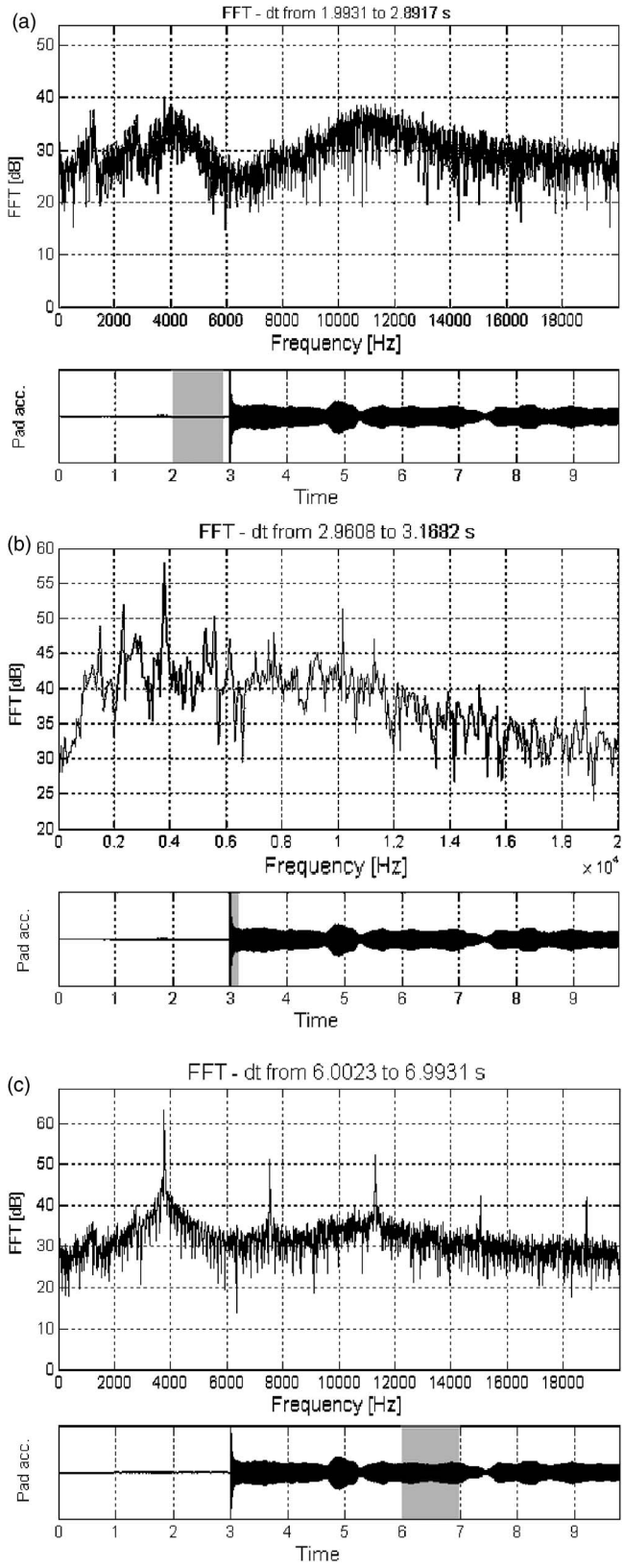


FIG. 10. FFT of the pad acceleration during the hammer impulse test: (a) before the impulse a stable behavior is measured; (b) the impulse excites all the bending modes of the disk; (c) after the impulse squeal occurs at the "tuning" frequency.

of the pad, and its harmonics are observed [Fig. 10(c)].

A different response of the system to the hammer impulse is obtained by increasing the normal load to 70 N and the first mode of the pad shifts about 300 Hz away from the (0,4+) disk mode. After every hammer impulse, squeal starts, and stops after few seconds.

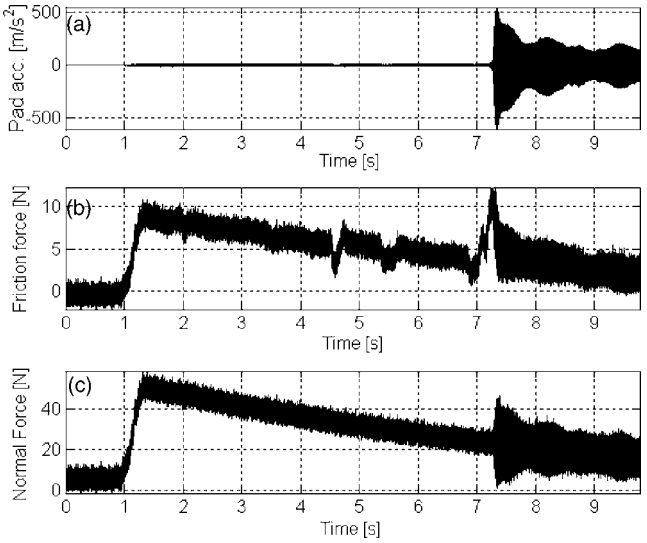


FIG. 11. Squeal vibration starts after an impulse in the friction force provided by a perturbation in the friction coefficient.

If the normal load is equal to 120 N, so that the first mode of the pad shifts more than 400 Hz away from the (0,4+) disk mode, the hammer impulse does not trigger the squeal anymore. Same results can be obtained by exciting the pad in the tangential direction by touching the disk surface in a point of the contact circumference with a humid wad. In this way it is possible to obtain an impulse in the friction force (Fig. 11) when the humid point passes under the pad. Using this kind of excitation, phases 1 and 3, previously described, remain unchanged [Figs. 10(a) and 10(c)], while in phase 2 only the frequency peak related to the disk mode, which is close to the pad mode, increases. Thus, the cause of instability is the excitation of one of the two modes that couple together. The impulse on the friction force excites the mode of the pad that couples with the (0,4+) bending mode of the disk. The hammer impulse on the disk surface excites the (0,4+) mode that couples with the mode of the pad.

C. Phase of the squeal vibrations

The major effect of a large modal damping is the slow ramp of the phase response of the system around the natural frequency. Larger is the modal damping and more gradual is the shift of the phase of the system response at its natural frequency. The authors believe that the dynamic instability is due to a phase relationship between the exciting force (friction force) and the response of the system (vibration of the pad surface in the tangential direction and vibration of the disk in the normal direction) that causes a self-excitation mechanism and brings squeal. A correlation between the squeal and the phase differences between the vibrations of the system was already presented by the authors.[21] Figure 12 shows the phase difference[28] between the measured normal and tangential forces in the different cases analyzed.

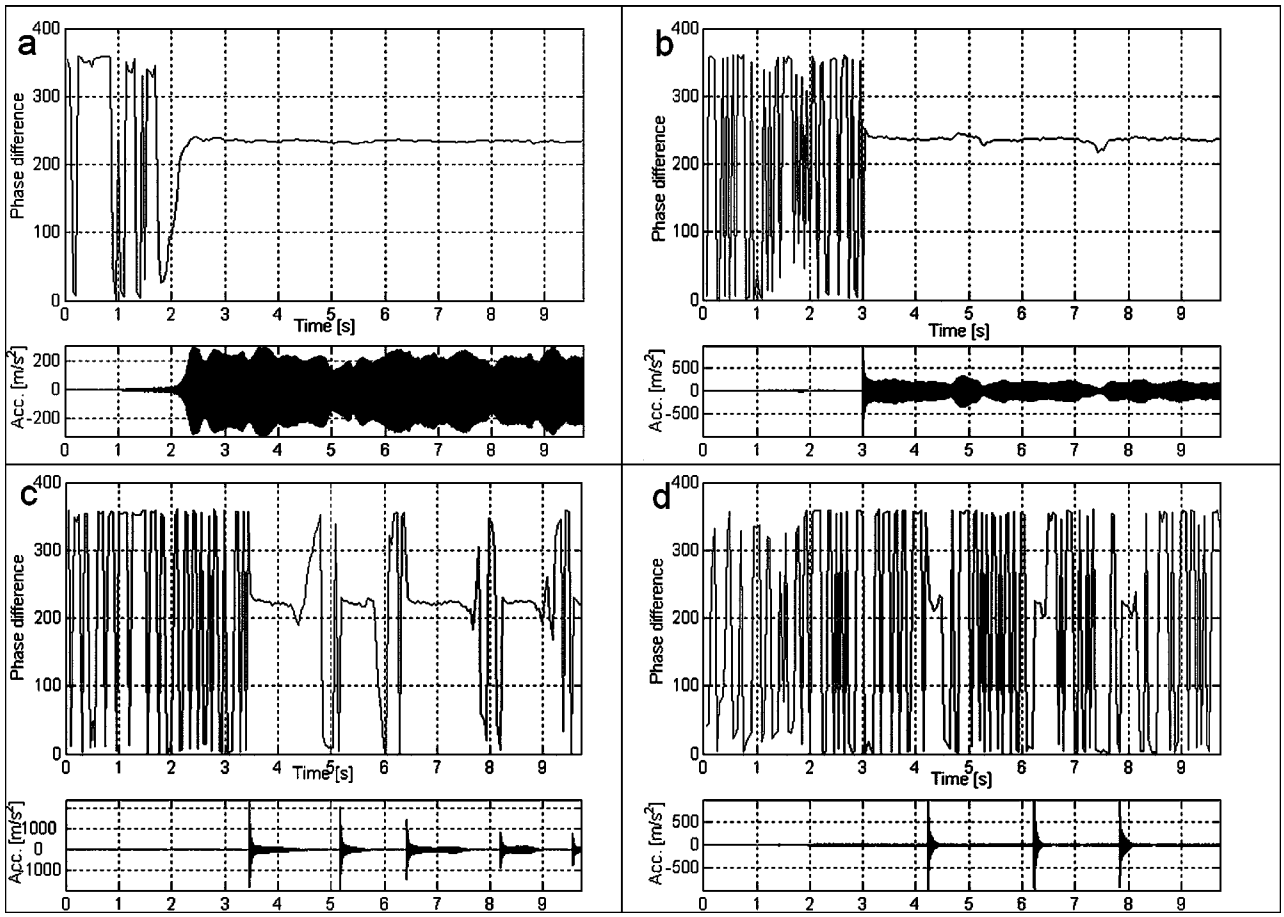


FIG. 12. Behavior of the phase difference between normal and friction forces calculated at the squeal frequencies (3767) over time, when the normal load is, respectively, equal to (a) 45 N with constant squeal; (b) 48 N with constant squeal after the impulse; (c) 70 N with momentary squeal after the impulses; (d) 120 N without squeal.

At the contact surface the normal and friction forces are in phase and proportional. The measured tangential force presents a difference of phase with the normal load because it is measured between the pad and the support. Thus, the dynamics of the pad is in between the contact point and the measured force. When the pad is excited in the tangential direction by the friction force, at a frequency close to its natural frequency, the pad vibrates with a maximum phase difference between these vibrations and the friction force.

When normal load is equal to 45 N and there is tuning between the two natural frequencies of the disk and pad, as soon as the pad gets in contact with the disk, squeal occurs and a constant phase difference of 240° is measured [Fig. 12(a)].

Changing the normal load from 45 to 48 N, a continuous squeal can still occur but an external excitation is necessary to trigger the squeal. Once excited, the squeal vibrations of the setup show a constant phase difference equal to 240° [Fig. 12(b)].

When the normal load is equal to 70 N it is not possible to have continuous squeal, but only temporary squeal occurs once the system is excited with an impulse; in these conditions, a phase difference equal to 230° is measured during the squeal emission [Fig. 12(c)]. With this value of normal load the two natural frequencies do not coincide and the squeal occurs at the natural frequency of the mode $(0,4+)$.

When the normal load is equal to 120 N [Fig. 12(d)] the two natural frequencies are well separated, so that when the system is excited with an impulse, squeal does not start and the phase difference is equal to 215°.

Note that when there is not exact coincidence between the two natural frequencies, squeal happens at the frequency of the disk mode because of its lower modal damping.

These experiments show that a large modal damping can increase the possibility of coupling between two modes of the system because it enlarges the tune-in range of the mode. Therefore, while it is useful to add damping material to reduce the modal response of the system, attention should be paid to the extended range of frequencies where the damped mode can couple with other modes. This explains why the disk-pad instability is characterized by a wider tune-in range between the modes with respect to the disk support instability.

VI. CONCLUSIONS

The proposed experimental setup simplifies the geometry of a real brake extremely so that it is possible to have repeatable and consistent measurements. However, as a consequence, the squeal events occurring during the experiments can only reflect a subset of squeal instabilities that may occur in real brakes.

More precisely:

(i) because of the reduced dimension of the pad, the disk-pad interaction is closer to that occurring in low-frequency squeal, characterized by a wavelength of the disk deformed shape longer than the pad length; and

(ii) because of the geometry of the disk, it is not possible to study squeal events characterized by modes of the disk with a consistent in-plane deformation.

Nevertheless, the setup is able to reproduce, "at will," squeal events that are qualitatively analogous to some real brake squeals, and therefore can provide a reliable experimental insight into the squeal physics.

The simplified geometry of the TriboBrake allows also for an easy modeling by FE to predict and reproduce squeal events.[29]

The measurements presented in this paper confirm the squeal phenomenon as a dynamic instability of the brake system. The dynamic analysis, together with the analysis of the squeal characteristics, relates the squeal phenomenon to a modal instability. In particular, squeal happens when a mode, characterized by large tangential vibrations of the pad surface, couples with a mode characterized by large bending vibration of the disk at the contact point.

These results agree with the mode lock-in theory and provide an experimental basis supporting the complex eigenvalues analysis as a tool for squeal prediction. The presented analysis shows also that squeal can be easily triggered when the system dynamics is favorable, i.e., when there is tuning between two appropriate modes of the system. It is important to note that a brake apparatus is subjected to several circumstances that can trigger the instability.

An important distinction between dynamics of the pad and the support should be highlighted, because the experiments show that squeal can be obtained from the modal coupling between disk and either the pad or the support. During brake design for squeal suppression, this aspect should be taken in account.

Two different roles of the modal damping are described: a large modal damping can reduce the response of the damped mode and consequently prevent its participation in squeal coupling; however, a high damped mode has more probability to couple with others modes that fall close to its natural frequencies because its tune-in range is larger. This suggests careful use of the addition of damping in order to reduce squeal.

To conclude, it is important to note that the experiments do not show any direct dependence between squeal occurrence and the variation of a single parameter. In fact, since the squeal depends on the tuning between two modes of the brake apparatus, it is the modal distribution of the whole brake system that may or may not induce the squeal instability. Moreover, the modal distribution depends on several global and local characteristics of the system in a nonmonotonic way. Therefore, in order to avoid squeal occurrence, the effectiveness of the changing of a single parameter must be evaluated in relation to the actual modal distribution, with the aim of eliminating a possible tuning without causing another one at a different frequency.

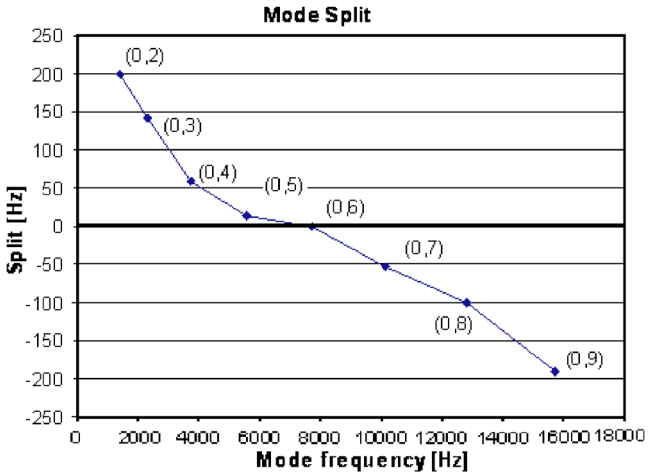


FIG. 13. Split of the disk modes with 225 N of normal load.

APPENDIX: NATURAL FREQUENCIES MODULATION FOR "MODAL TUNING"

In order to tune the dynamics of the setup to have squeal, different driving parameters are chosen to shift the natural frequencies of the system: the load applied on the top of the support, the friction pad dimensions, the stiffness of the thin plates, insertion of damping material between the thin plates and the support (Fig. 1).

Figure 13 shows the split values of the disk modes when a mass of 20 Kg is placed on the top of the support (225 N of normal load). The split is defined as positive when the $(n,m+)$ mode is at higher frequency than the $(n,m-)$ mode. The FRF of the disk presents a single peak when there is no contact with the pad; the separation of the double peaks increases by increasing the normal load. It is interesting to notice that the modal split is positive for low-frequency modes, almost zero for the mode with six nodal diameters, and negative for high-frequency modes. This behavior is due to the mass and stiffness effects introduced by the contact with the pad. The $(n,m-)$ mode has the contact area belonging to the nodal diameter so that it is not influenced by the contact and its frequency remains almost constant. The $(n,m+)$ mode has the contact area in the antinode of the disk. For low-frequency modes the addition of stiffness due to the contact stiffness with the pad and support has more influence than the addition of mass, and the natural frequency increases. The influence of the mass effect increases by increasing the frequency, and for high-frequency modes the natural frequency decreases with the increase of the load.

Adding weights on the top of the support changes the support natural frequencies. Moreover, because of the dependence of the properties of the friction material on the stress state of the material itself, and because of the increase of the contact stiffness with the load, the natural frequencies of the pad increase with the applied load. Table II shows the frequency ranges where the modes of the support and pad fall,

TABLE II. Support and pad frequency range for different load conditions.

Mode	Frequency range [Hz]
II support mode	1 500–2 100
III support mode	2 400–3 150
I pad mode	3 750–4 500
II pad mode	10 000–11 800

the normal load varied from 25 to 225 N. The contact surface of the pad is equal to 10×10 mm, and the thickness of the thin plates is equal to 0.5 mm.

Table II shows only the second and third modes of the support, because only these two modes, which have a large vibration at the bottom end where the support is in contact with the disk, are involved in the squeal phenomena. Since the dynamic instability is due to the coupling between one normal mode of the disk and one tangential mode of the support or pad, these modes are those characterized by highest coupling with the bending dynamics of the disk.

The thickness of the thin plates does not affect the frequencies of the disk and pad modes. By using different dimensions for the pad (8×8, 10×10, 10×15, 10×20 mm^2), the frequencies of the pad modes can be varied between 3700 and 5000 Hz for the first mode and between 10 000 and 14 000 for the second one.

[1] A. Akay, "Acoustic of friction," J. Acoust. Soc. Am. **111**(4), 1525–1548 (2002).

[2] N. M. Kinkaid, O. M. O'Reilly, and P. Papadopoulos, "Automotive disc brake squeal," J. Sound Vib. **267**, 105–166 (2003).

[3] H. Ouyang, W. Nack, Y. Yuan, and F. Chen, "Numerical analysis of automotive disc brake squeal: A review," Int. J. Vehicle Noise and Vibration **1**, Nos. 3/4, 207–231 (2005).

[4] H. R. Mills, "Brake squeak," Technical Report 9000 B, Institution of Automobile Engineers(1938).

[5] R. A. C. Fosberry and Z. Holubecki, "Interim report on disc brake squeal," Technical Report 1959/4, Motor Industry Research Association, Warwickshire, England (1959).

[6] R. A. C. Fosberry and Z. Holubecki, "Disc brake squeal: Its mechanism and suppression," Technical Report 1961/1, Motor Industry Research Association, Warwickshire, England (1961).

[7] R. T. Spurr, "A theory of brake squeal," in *Proceedings of the Automobile Division, Institution of Mechanical Engineers*, 1961–1962, No. (1), pp. 33–52 (1961).

[8] S. W. E. Earles and G. B. Soar, "Squeal noise in disc brakes," in Vibration and Noise in Motor Vehicles, Institution of Mechanical Engineers, London, England, pp. 61–69 (Paper number C 101/71) (1971).

[9] S. W. E. Earles, "A mechanism of disc-brake squeal," Technical Report 770181, SAE, Warrendale, PA (1977).

[10] R. P. Jarvis and B. Mills, "Vibrations induced by friction," Proc. Inst. Mech. Eng. **178**(32), 847–857 (1963).

[11] M. R. North, "Disc brake squeal, a theoretical model," Technical Report 1972/5, Motor Industry Research Association, Warwickshire, England (1972).

[12] M. R. North, "Disc brake squeal," in *Braking of Road Vehicles*, Automobile Division of the Institution of Mechanical Engineers (Mechanical Engineering, London, England, 1976), pp. 169–176.

[13] A. Akay, J. Wickert, and Z. Xu, "Investigation of mode lock-in and friction interface," Final Report, Department of Mechanical Engineering, Carnegie Mellon University(2000).

[14] A. Tuchinda, N. P. Hoffmann, D. J. Ewins, and W. Keiper, "Effect of pin finite width on instability of pin-on-disc systems," in *Proceedings of the International Modal Analysis Conference—IMAC*, Vol. **1**, pp. 552–557 (2002).

[15] A. Tuchinda, N. P. Hoffmann, D. J. Ewins, and W. Keiper, "Mode lock-in characteristics and instability study of the pin-on-disc system," in *Proceedings of the International Modal Analysis Conference—IMAC*, Vol. **1**, pp. 71–77 (2001).

[16] R. Allgaier, L. Gaul, W. Keiper, K. Willnery, and N. Hoffmann, "A study on brake squeal using a beam on disc," *Proceedings of the Int. Modal Analysis Conference—IMAC*, Vol. **1**, pp. 528–534 (2002).

[17] R. Allgaier, "Experimentelle und numerische untersuchungen zum bremsenquietschen," (Experimental and numerical examinations on the brake-squeaks), Ph.D. thesis, University of Stuttgart, Germany.

[18] J. F. Tarter, "Instabilities in a beam-disc system due to friction," Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA (2004).

[19] Q. Cao, H. Ouyang, M. I. Friswell, and J. E. Mottershead, "Linear eigenvalue analysis of the disc-brake squeal problem," Int. J. Numer. Methods Eng. **61**, 1546–1563 (2004).

[20] O. Giannini and A. Sestieri, "Predictive model of squeal noise occurring on a laboratory brake," J. Sound Vib. **296**, 583–601 (2006).

[21] O. Giannini, A. Akay, and F. Massi, "Experimental analysis of brake squeal noise on a laboratory brake setup," J. Sound Vib. **292**(1–2), 1–20 (2006).

[22] The effect of increasing the rotation of the disk is a reduction of the natural frequencies. Lower frequencies can be explained considering that the thin plates holding the support have one end compressed during braking phase. This introduces a lower stiffness in the support substructure to lower frequencies.

[23] An extensive description of the modulation of the setup dynamics is reported in the Appendix.

[24] During experiments the natural frequencies of the pad and the support are measured by looking at the PSD of the pad acceleration in the tangential direction. In fact, the dynamics of the components already being identified, the PSD allows one to follow the real values of the natural frequencies of the pad and the support in the effective operating state, i.e., during the braking phase. On the contrary, the natural frequencies involving bending vibrations of the disk are measured to the brake under brake application without disk rotation. However, a preliminary test, performed under brake application with disk rotation, assured a negligible effect of the rotation on these modes, which are the only rotor modes involved in squeal in this setup.

[25] In Fig. 9 slow, nonphysical decrease of the mean values of the global forces is observed, and it is due to the discharging of the capacitive transducer.

[26] Only FRF for 45 N of normal load is plotted for clarity, but no coincidence between the natural frequencies is obtained for different values of the load.

[27] W. Chen and X. Deng, "Structural damping caused by micro-slip along frictional interfaces," Int. J. Mech. Sci. **47**, 1191–1211 (2005).

[28] The difference of phase is calculated by performing the FFT of the global normal and friction forces and plotting the difference of the phases at the frequency of squeal, which is the frequency of the dynamic instability.

[29] F. Mass, L. Baillet, O. Giannini, and A. Sestieri, "Brake squeal: Linear and nonlinear numerical approaches," Mechanical Systems and Signal Processing (submitted).

Analysis of Sabine and Eyring equations and their application to concert hall audience and chair absorption

Leo L. Beranek[a)]
975 Memorial Drive, No. 804, Cambridge, Massachusetts 02138

(Received 23 January 2006; revised 2 June 2006; accepted 2 June 2006)

Historically, two equations have been used for predicting reverberation times, Sabine and Eyring. A precise means is presented for determining Eyring absorption coefficients α_{eyring} when the Sabine coefficients α_{sabine} are known, and vice versa. Thus, either formula can be used provided the absorption coefficients for the Sabine formula are allowed to exceed 1.0. The Sabine formula is not an approximation to the Eyring equation and is not a shortcoming. Given low reverberation times, the ratio of α_{sabine} to α_{eyring} may become greater than 2.0. It is vital that, for correct prediction of reverberation times, the absorption coefficients used in either formula must have been determined in spaces similar in size and shape, with similar locations of high absorption (audience) areas, and with similar reverberation times. For concert halls, it is found that, when the audience area (fully occupied) and midfrequency reverberation time are postulated, the hall volume is directly proportional to the audience absorption coefficient. Approximately 6% greater room volumes are needed when choosing nonrectangular versus classical-rectangular shaped halls and approximately 10% greater volumes when choosing heavily upholstered versus medium upholstered chairs. Determinations of audience sound absorption coefficients are presented, based on published acoustical and architectural data for 20 halls. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2221392]

PACS number(s): 43.55.Br, 43.55.Dt, 43.55.Ev, 43.55.Fw, 43.55.Gx [NX] Pages: 1399–1410

I. INTRODUCTION

The prediction of reverberation times in concert halls began with the researches of Sabine (1900). His simple formula for relating reverberation time directly to the volume of a room and inversely to the absorbing power of the audience and other surfaces and objects in the room has found widespread use. With it, zero reverberation time in a room fully covered with the material requires an absorption coefficient of infinity, a fact that Sabine appeared to disregard when he stated (1900, 1906, 1915) that the absorbing power of an open window, meaning a surface with no reflected sound, is 1.000. By contrast, in a 1912 paper, he shows an absorption coefficient of 1.26 at 1024 Hz for a felt material and, in a 1915 paper, an absorption coefficient of 1.10 at 512 Hz for "upholstered settees" and 1.12 at 512 Hz for wood sheathing, 2 cm thick.

Eyring (1930), presented an alternate equation that calculates zero reverberation time for a room fully lined with a material having an absorption coefficient of 1.0. Both authors assumed in their derivation that the absorbing power is nearly uniformly distributed over all the surfaces in the room, and that the sound field is nearly diffuse so that the results are almost independent of a room's shape. For rooms where the sound field is not perfectly diffuse, a controversy has existed over which of the two reverberation equations is more accurate even for the case where the absorbing power is evenly distributed over the surfaces and the average sound absorption coefficient does not exceed 0.5.

Audience absorption has received attention in recent years (Bradley, 1992, 1996; Davis *et al.*, 1994; Kirdegaard, 1996; Beranek and Hidaka, 1998; Hidaka *et al.*, 2001; Barron and Coleman, 2001; Beranek, 1962, 1969, 1996, 2004). Two goals of this paper are to investigate where and how to use the Sabine/Eyring equations (particularly in concert halls) and to determine audience and chair absorption coefficients. The following topics are treated. (1) The Sabine and Eyring equations and a precise means for deriving Eyring audience absorption coefficients from Sabine coefficients. (2) Under what conditions will either equation calculate accurate reverberation times. (3) Choice of unit-area versus per-person method for specifying the sound absorption of the audience in a concert hall. (4) Residual (nonaudience) absorption coefficients in halls for music. (5) Audience absorption coefficients for 20 concert halls. 7. Chair absorption coefficients for 20 halls. (6) Effect of room shape and degree of upholstering on audience and chair absorptions. (7) Hall volume related to room shape and chair upholstering.

II. THE SABINE EQUATION

The Sabine equation at normal room temperature, 22 °C, is

$$T_{60} = 0.161V/(A + 4mV)\ \text{s}, \tag{1}$$

$$A = \alpha_{\text{tot}} S_{\text{tot}}\ \text{m}^2, \tag{2}$$

$$\alpha_{\text{tot}} = (\alpha_T S_T + \alpha_R S_R + \Sigma \alpha_i S_i)/S_{\text{tot}}, \tag{3}$$

and

a)Electronic mail: beranekleo@ieee.org

$$S_{\text{tot}} = S_T + S_R + \Sigma S_i \text{ m}^2. \quad (4)$$

The audience absorption coefficient is

$$\alpha_T = (\alpha_{\text{tot}} S_{\text{tot}} - \alpha_R S_R - \Sigma \alpha_i S_i)/S_T, \quad (5)$$

where V is the room volume in cubic meters; A is the total sound absorption in the room in square meters; S_T is the acoustical audience area (i.e., area beneath chairs plus areas of strips 0.5 m wide around audience blocks except for sides at balcony rails or walls, and measured on the slope) plus the area the orchestra sits over (for the orchestra no sloping and limited to a maximum of 180 m^2); S_i's are any areas of highly absorbing surfaces in the hall (sometimes introduced for echo control); and S_R (called "residual absorption" area) is the area of all other surfaces in the hall, including under-balcony areas. The α's are the Sabine sound absorption coefficients associated with their corresponding areas; and m is the energy attenuation constant for sound traveling through air in units of m^{-1}.

In this study, the absorption by objects, lighting fixtures, ventilating openings, cracks around doors in the room, thin carpeting in some of the aisles, etc., are included in the residual absorption. The absorption by the air itself is assumed here to be of importance only for the frequency bands of 2000 Hz and higher (4 m in this paper is taken to be 0.0089 m^{-1} at 2000 Hz and 0.0262 at 4000 Hz). In only three halls of this study, New York Carnegie, Sapporo Kitara, and Amsterdam Concertgebouw are there significant areas of high absorptivity besides the audience and they are accounted for in the calculations that follow.

III. THE EYRING EQUATION[1]

The Eyring equation, with the same assumptions as above, is

$$T_{60} = 0.161 V/(A' + 4mV) \text{ s}, \quad (6)$$

$$A' = S_{\text{tot}}[-2.30 \log_{10}(1 - \alpha_{ey})] \text{ m}^2, \quad (7)$$

where

$$\alpha_{\text{ey}} = (\acute{\alpha}_T S_T + \acute{\alpha}_R S_R + \Sigma \acute{\alpha}_i S_i)/S_{\text{tot}}, \quad (8)$$

and the $\acute{\alpha}$'s are the Eyring sound absorption coefficients associated with their corresponding areas.

IV. DERIVING EYRING COEFFICIENTS FROM SABINE COEFFICIENTS

A simple and precise means for transfer from the Sabine sound absorbing coefficients to the Eyring ones is possible because the same procedure for obtaining the average absorption coefficients in a room is followed in both Eqs. (5) and (8). Thus

$$(\alpha_{\text{ey}}/\alpha_{\text{tot}}) = (\acute{\alpha}_T S_T + \acute{\alpha}_R S_R + \Sigma \acute{\alpha}_i S_i)/(\alpha_T S_T + \alpha_R S_R + \Sigma \alpha_i S_i) \quad (9)$$

and

$$(\acute{\alpha}_T S_T + \acute{\alpha}_R S_R + \Sigma \acute{\alpha}_i S_i) = (\alpha_{ey}/\alpha_{\text{tot}})(\alpha_T S_T + \alpha_R S_R + \Sigma \alpha_i S_i),$$

hence,

$$\acute{\alpha}_T = (\alpha_{\text{ey}}/\alpha_{\text{tot}})\alpha_T, \quad (10a)$$

$$\acute{\alpha}_R = (\alpha_{\text{ey}}/\alpha_{\text{tot}})\alpha_R, \quad (10b)$$

$$\Sigma \acute{\alpha} = (\alpha_{\text{ey}}/\alpha_{\text{tot}})\Sigma \alpha_i. \quad (10c)$$

V. ABOUT THE EQUATIONS—INTERRELATIONS AND ACCURACY

The Sabine and Eyring equations were derived under different assumptions. The Sabine equation assumes that as a sound wave travels around a room it encounters surfaces "one after another." The Eyring equation assumes that all the surfaces are simultaneously impacted by the initial sound wave, and that successive simultaneous impacts, each diminished by the average room absorption coefficient, are separated by mean free paths. Cremer and Mueller (1982) call the former "one-after-another" and the other "side-by-side." In a real room, neither condition is met. Schroeder (1973) assumed a special distribution of free paths and he got Sabine's values exactly.

Joyce (1978, 1980) has developed an exact equation for the geometrical-acoustics (ray-tracing) value of reverberation time (RT) in a room of arbitrary shape and arbitrary distribution of angular and spatial absorptivity and arbitrary surface irregularities. This RT formula can be solved numerically using available Fredholm solvers (Joyce, 1980). Unfortunately, the calculations are exceedingly difficult, except for the cases where the geometry is easy to express mathematically, such as a two wall "enclosure" or a spherical or rectangular enclosure. In the two-wall case, half the surface area can be made sound absorbent and the other half perfectly reflecting. But, even in the case of mathematically simple enclosures, e.g., sphere or rectangular in shape, the calculations are simple only if all surfaces have the same absorption coefficients and irregularities.

As an *illustration* of how the reverberation time RT varies as a function of a wide range of surface absorptivities and surface irregularities, Joyce presents the exact solution for a spherical enclosure with (a) the same absorptivity and roughness over the entire inner surface, (b) variation of absorptivity from 0 to 1.0, and (c) variation of reflectivity (measure of roughness) from specular to random, i.e., $s=1$ for specular and $s=0$ for random. Joyce then defines a reference sound absorption coefficient α_{ref} to be used in the reverberation equations.

The Sabine equation for the spherical example becomes

$$T_{60} = \text{constant} \times V/(S\alpha_{\text{ref}}). \quad (11)$$

The Eyring equation for the spherical example is

$$T_{60} = \text{constant} \times V/(S\alpha_K), \text{ where}, \quad (12)$$

$$\alpha_K = -2.3 \log(1 - \alpha_{\text{ref}}). \quad (13)$$

The results are given in Fig. 1 for a range of α_{ref} from 0.05 to 0.5.

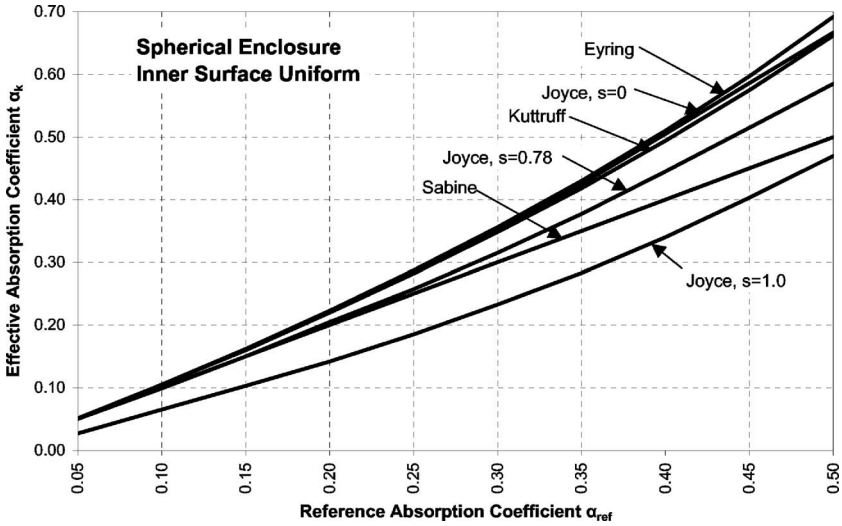


FIG. 1. Comparison of reverberation theories. Joyce labels the ordinate "Normalized decay rate." Here, it is called "Effective absorption coefficient." Joyce labels the abscissa "Surface absorptivity." Here, it is called "Reference absorption coefficient." From Joyce (1978).

Kuttruff (1973) derived a modification of the Eyring formula that assumed that the reflectivity of the surface was everywhere isotropically absorptive and also maximally diffusive

$$\alpha_{\text{kut}} = [-2.3 \log(1-\alpha_{\text{ref}})][1 + (2.3/16)\log(1-\alpha_{\text{ref}})]. \quad (14)$$

Otherwise, he made no assumption about the diffusivity of the sound field itself. Carrol and Chen (1977) obtained an analytic solution to the equation for a sphere with a spatially uniform absorptivity, i.e., Joyce's derivation with $s=0$. Kuttruff's equation is plotted in Fig. 1.

The effective absorption quotient from the Joyce theory for random reflection, $s=0$ is almost equal to the Eyring coefficient. Kuttruff's method gives a slightly lower absorption coefficient. Inspection of the Joyce curves for $s=0.78$ and 1.0, show that they lie closer to the Sabine curve than to the Eyring or Kuttruff curves. But surface considerations are not alone of importance. Joyce is careful to point out that "randomnizing" of the sound field is a function of the enclosure shape as well as the absorptivity/roughness characteristics of the material on its surfaces, and unless the sound field is nearly completely random there is a chance that the Sabine formula is even more correct than the Eyring or Kuttruff formulas. This statement is limited to values of α_{ref} that are less than about 0.4 (Joyce suggests 0.58, where the two bottom curves of Fig. 1 intersect).

There is another reason why one must beware of thinking that the sound field in a room is truly random. Joyce points out, and it is logical, that sound intensity tends to be greater for paths in the enclosure that mainly involve surfaces with low absorption coefficients. Or, as Joyce puts it, "there is a tendency of sound to accumulate on paths of longer life."

Hodgson (1993) in his Fig. 2 clearly shows that in actual room measurements, Eq. (10) is correct, but he incorrectly concludes, "At high frequencies the Sabine formula gives coefficients which are greater than 1.0—an apparently physically impossible result, indicating a short-coming of the Sabine theory," and Mange (2005) incorrectly states, "The standard Sabine equation was used…absorption coefficients greater than one are often measured. This is theoretically impossible." When using the Sabine theory, one must accept coefficients greater than one, which is neither theoretically impossible nor a shortcoming in the theory. If coefficients less than 1.0 are always desired, the Eyring alternate must be used. Hodgson gives measurements of the absorption coefficients in a test room with low reverberation time. At 8000 Hz he obtained an Eyring coefficient of 0.85 and a Sabine absorption coefficient of 1.90 and, i.e., $(\alpha_{\text{ey}}/\alpha_{\text{tot}}) = 0.45$, a legitimate result, not a shortcoming in the Sabine theory.

VI. PREMISE OF THIS PAPER

The question must be asked, "What is the measure of correctness?" Obviously, the answer ought to be that the chosen equation should predict the correct reverberation time for a room regardless of how much or where the principal absorbing material (with a constant absorption coefficient) is placed. But, is this possible? Look backwards at the problem. Start with the reverberation time and the complete physical characteristics of a rectangular enclosure. Using Joyce's exact theory, calculate the average sound absorption coefficient of the material for a particular area and location in the room. But, if this transfer function (RT to α) is now held constant, it well known that the absorption coefficient of a material will take on different values depending on where and how much of the material exists on the walls [Andre (1932)].

The premise of this paper is that if a particular reverberation equation (transfer function) is used and the reverberation time is held constant, the calculated absorption coefficient for the absorbing material will vary as it is moved around or made larger or smaller. Alternatively, if one requires that the absorption coefficient must remain the same regardless of its position or size, there is no *single* equation (transfer function) that will calculate the correct reverberation times. To explore this premise, this paper concentrates on the absorption of sound by a seated audience in concert halls.

Using an appropriate CAD program, a computer model of a room can be used to determine the reverberation time from a given audience absorption coefficient. The transfer

function (α_T to RT) will be different for every room shape and absorption configuration and may be near to or differ greatly from the Sabine or Eyring equation. Hidaka (2006) found that ta change in the mean-free-path length would enable the use of the Sabine equation for calculation of the RT's for non-shoebox type concert halls. The α_T's used therein were determined in shoebox type concert hall where he found MFP$=4V/S_{tot}$.

As will be demonstrated, the sound absorption by an audience depends on the shape of the enclosure, on the location of the audience in the enclosure, and on the nature of the sound field in the enclosure. Hence, for calculation of room reverberation times using a particular formula the audience absorption coefficients should have been determined from measurements in approximately the same shapes and sizes of enclosure, on the same position of the audience area in the enclosure, and with approximately the same enclosure reverberation times.

From studies made in classrooms, Bistafa and Bradley (2000) write, "The Sabine/Eyring formula is concluded to be a reasonable choice among the analytical expressions compared here." They found computer simulations of that year to give no better results. They also conclude, as does the above premise, that the absorption coefficients used in predicting reverberation times in classrooms should be determined in rooms that simulate the actual classroom sizes and shapes and locations of acoustical materials.

VII. CHOICE OF UNIT-AREA OR PER-PERSON AUDIENCE ABSORPTION IN CONCERT HALLS

Beranek (1960, 1962, 1969) has demonstrated that in concert halls the audience absorption is proportional to the area (with edge corrections—see S_T in Sec. II) over which the audience sits and not to the number of seats. The impetus for that series of studies followed the failure of preconstruction computations of reverberation times to predict those actually measured. Notable were the differences in Boston Symphony Hall (calculated 2.3 s, measured 1.9 s), Royal Festival Hall (1.8 s vs 1.5 s), Mann Auditorium (1.9 s vs 1.5 s); and Edmonton Jubilee Auditorium (1.8 s vs 1.4 s). In those four cases, the audience absorption used in the design stage was proportional to the seat count and the value chosen was nearly the same as Sabine's original value of 0.44 m^2 per person (Sabine, 1900), a number that is at least 20% too low for modern seat and row-to-row spacing (Nishihara *et al.*, 2001). Sabine published audience absorptions at 512 Hz both on "per person" (0.44) and "per unit area" (0.96) bases. If Sabine had used the unit-area number instead of 0.44 per person in calculating the reverberation time of Boston Symphony Hall, his calculated RTs would have been much closer to those measured. [The number of seats per 100 square meters varies from 244 (Amsterdam, Concertgebouw) to 172 (Munich, Philharmonie).]

Barron and Coleman (2001) have reinforced the "audience area" concept, reporting on measured data in seven British concert halls, saying "there is a very good correlation coefficient ($r=0.98$) with the [occupied] seating area, and a non-significant [very low] correlation with the number of seats… . This…supports Beranek's proposition that seat absorption should be treated by area." It must be noted, that in a reverberation chamber where the average room sound absorption coefficient is very low and the sample size is small, the absorption coefficient for an audience usually measures somewhere in-between the per person and the unit area values.

Clearly, which method is used would make no difference if all halls had the same row-to-row and seat-to-seat spacing, i.e., the same area per person. But, another consideration forces the choice of the unit-area method. The sound absorption is greater for an audience seating area that is steeply sloped (raked) because more of the human body is directly exposed to the sound (Nishihara *et al.*, 2001). If the per person assumption is made, the predicted total absorption in a room will not change with audience sloping. The sloped areas in a hall may be as much as 15% greater than the projected areas.

In the basic data for concert halls that follow, the per-area method is used. It must be noted that the acoustic audience areas S_T given here may differ significantly from the S_T's in Beranek (2004) for the same halls because the areas here are measured on the slopes, that is, they are not projected areas.

VIII. REMARKS ON THE PREMISE OF SECTION VI

Some believe that if the sound absorption by the seats in a hall is very large, the reverberation times will be independent of whether or not they are occupied. But, as is investigated here, even when the seats are occupied, the audience sound absorption is dependent on the seat design and on different degrees of upholstering. Furthermore, the sound absorption of an audience depends on whether a hall has upper areas where reverberation can develop, which is usual in older rectangular halls, or on whether the seats extend upward in front of one or more walls thus largely preventing the development of overhead reverberation.

In summary, the audience absorption coefficients shown in this paper for use in the Sabine or Eyring equations are probably only meaningful for halls with reverberation times in the range of 1.6–2.0 s, cubic volumes in the range of 10 000–30 000 m^3, and one of two shapes, i.e., (1) rectangular with large open wall areas above the highest balconies and (2) nonrectangular or near-rectangular with audience areas so steep that they mask or cover one or more side or end walls.

Following the belief that sound tends to accumulate on paths of longer life, investigation of the 85 concert halls presented in Beranek (2004), revealed that of the many rectangular halls in the book, a sample of nine would be an adequate representation. They evidenced upper interiors where the sound could reverberate without fully involving the audience below. Of the nonrectangular halls, a group of 11 halls was selected in all of which the seating sloped upward so steep that the rear or one or more side walls or both were masked [In Beranek (2004), more were not possible because all of the data needed were not available]. For all 20 halls, available were (1) the architectural plans, (2) a descrip-

TABLE I. Basic data for the halls of this study. The residual absorption coefficients used throughout this study are shown for use in the Sabine equation. Eyring residual absorption coefficients=Sabine coefficients times (α_{ey}/α_{tot}).

Nonrectangular halls	Volume (m^3)	Occupied areas S_T (m^2)	Residual areas S_R (m^2)	Combined areas S_{tot} (m^2)	RT occupied halls (sec) frequency (Hz) 125	250	500	1000	2000	4000
Sapporo, Kitara Concert Hall	28800	1786	6582	8368	2.05	1.95	1.90	1.90	1.75	1.50
Munich, Philharmonie am Gasteig	29700	2000	6134	8134	1.95	2.00	1.90	1.95	1.75	1.50
Rotterdam, DeDoelen Concert Hall	24070	1730	4098	5828	2.30	2.00	1.90	1.90	1.85	1.65
Berlin, Philharmonie	21000	1620	3784	5404	2.10	1.85	1.85	1.95	1.80	1.60
New York, Avery Fisher Hall	20400	1660	3936	5596	1.60	1.76	1.78	1.74	1.55	1.46
Cleveland, Severance Hall	16290	1486	3502	4988	1.75	1.70	1.65	1.55	1.45	1.30
Baltimore, Meyerhoff Hall	21530	1700	4154	5854	2.30	2.10	2.00	2.00	1.65	1.35
Manchester, Bridgewater Hall	25000	1850	5126	6976	2.30	2.12	2.00	2.00	1.80	1.65
New York, Carnegie	24270	1874	4553	6427	2.12	2.00	1.83	1.75	1.57	1.40
Tokyo, Suntory	21000	1578	4548	6126	2.14	2.08	1.95	2.00	1.90	1.75
Buffalo, Kleinhans	18280	2200	3626	5826	2.15	1.65	1.60	1.40	1.27	1.23
Rectangular halls										
Boston Symphony Hall	18750	1522	4150	5672	1.95	1.90	1.90	1.90	1.59	1.43
Berlin, Konzerthaus	15000	1101	3717	4818	2.20	2.10	2.00	2.00	1.80	1.60
Vienna, Musikvereinssaal	15000	1118	3147	4265	2.25	2.18	2.04	1.96	1.80	1.62
Lenox, Ozawa Hall	11610	919	2978	3897	2.00	1.90	1.80	1.75	1.70	1.40
Seattle, Benaroya Hall	19263	1668	4335	6003	2.15	1.90	1.80	1.75	1.65	1.55
Kyoto, Concert Hall	17800	1342	4571	5913	2.16	2.06	1.99	1.98	1.82	1.59
Amsterdam, Concertgebouw	18780	1285	3706	4991	2.20	2.15	2.05	1.95	1.80	1.55
Lucerne, Concert Hall	17823	1465	4018	5483	2.10	2.00	1.90	1.80	1.65	1.50
Tokyo, Tokyo Opera City Concert Hall	15300	1220	4791	6011	2.07	2.03	1.99	1.93	1.84	1.66
Sabine equation: Residual absorption coefficients used in calculations, all halls					0.14	0.12	0.10	0.09	0.08	0.07
Except: Boston					0.17	0.14	0.11	0.09	0.09	0.08
Tokyo, TOC					0.14	0.115	0.089	0.079	0.075	0.059
Berlin, Konzerthaus					0.15	0.12	0.10	0.09	0.08	0.07
New York, Avery Fisher					0.28	0.18	0.13	0.11	0.10	0.10
Berlin, Philharmonie					0.19	0.20	0.15	0.10	0.09	0.08
Eyring equation: Residual absorption coefficients must be determined from Sabine coefficients times (α_{ey}/α_{tot}).										

tion of the interior surfaces, (3) the type of audience chair, and (4) measured reverberation times of sufficient accuracy.

No halls were selected that have steeply, inwardly sloped upper side walls in which an unusually high percentage of the early sound energy is directed to the audience areas (e.g., Christchurch Town Hall). This style should be studied separately, but for the three in Beranek (2004), the necessary accurate basic data are available on only one.

The chosen nine rectangular halls and eleven nonrectangular halls are listed in Table I along with basic information on each. They have midfrequency reverberation times in the range of 1.6–2.0 s. Comparative data on all 20 halls, with seats fully occupied, using the Sabine equation are presented in Table II.

IX. RESIDUAL ABSORPTION COEFFICIENTS

Necessary to this study are residual absorption coefficients α_R's, that is, the average of the sound absorption coefficients of the surfaces in a hall other than those covered by the audience, orchestra, and heavily absorbing materials. Measurements have been reported of α_R's in ten fully completed halls with no seats present (Beranek and Hidaka, 1998). Of those, Boston Symphony Hall and Tokyo Opera City (TOC) Concert Hall are included here. For 15 of the halls the values for α_R has been estimated from the data for similar halls in the 1998 paper. Even if the assumed residual coefficients for the 16 halls are different by small amounts (as indicated by comparison with the residual coefficients for the Boston, Tokyo, and Berlin Konzerthaus Halls, at the bottom of Table I) the effect on the computed reverberation times will not be great because 60%–80% of the total room absorption comes from (occupied) audience areas. All residual absorptions are shown at the bottom of Table I.

For two of the halls, where the walls have unusual construction details, New York Avery Fisher and Berlin Philharmonie, the values of α_R were determined by assuming that the *audience absorption coefficients were the same* as the average of those in the Baltimore, Manchester, and Tokyo-Suntory Halls and calculating backwards. The residual values derived this way are plotted in Fig. 2. In the New York Avery Fisher Hall the walls are wooden and the seats are mounted on a wooden floor with airspace beneath. As expected, the wooden surfaces absorb more sound in the lowest two frequency bands than that for "most halls." In the Berlin

TABLE II. Comparative reverberation, physical, and audience absorption data for all halls in this study. Sabine equation used in calculating absorption coefficients.

Nonrectangular halls	Reverberation times (sec)			Total hall absorption coefficient $(0.161V/T_{60}S_{\text{tot}})$-$4mV/S_{\text{tot}}$			V/S_T (m)	Audience absorption coefficient			% of absorption in occupied areas			Resid. over other S_R/S_T	Area per Seat S_a/N
	125	500	2000	125	500	2000		125	500	2000	125	500	2000	–	(m^2)
Sapporo, Kitara	2.05	1.90	1.75	0.27	0.29	0.29	16.1	0.70	0.93	0.97	57	72	77	3.7	0.54
Munich, Philharmonie	1.95	1.90	1.75	0.30	0.31	0.33	14.9	0.80	0.95	0.99	65	76	80	3.1	0.58
Rotterdam, DeDoelen	2.30	1.90	1.85	0.29	0.35	0.35	13.9	0.64	0.94	0.90	66	80	83	2.4	0.55
Berlin, Philharmonie	2.10	1.85	1.80	0.30	0.34	0.34	13.0	[a]	[a]	[a]	56	69	80	2.3	0.50
New York, Avery Fisher	1.60	1.78	1.55	0.37	0.33	0.37	12.3	[a]	[a]	[a]	46	72	79	2.4	0.43
Cleveland, Severance	1.75	1.65	1.45	0.30	0.32	0.36	11.0	0.68	0.83	0.93	67	78	83	2.4	0.47
Baltimore, Meyerhoff	2.30	2.00	1.65	0.26	0.30	0.35	12.7	0.54	0.78	0.93	61	76	83	2.4	0.49
Manchester, Bridgewater	2.30	2.00	1.80	0.25	0.29	0.32	13.5	0.56	0.81	0.87	59	75	80	2.8	0.60
New York, Carnegie	2.12	1.83	1.57	0.29	0.33	0.38	13.0	0.63	0.86	0.93	65	78	83	2.4	0.43
Tokyo, Suntory	2.14	1.95	1.90	0.26	0.28	0.29	13.3	0.60	0.81	0.78	65	77	80	2.9	0.54
Buffalo, Kleinhans	2.15	1.60	1.27	0.24	32	0.39	8.3	0.39	0.67	0.85	63	80	87	1.6	0.58
										Avg	61	77	82	2.6	0.52
Rectangular halls															
Boston Symphony Hall	1.95	1.90	1.59	0.27	0.28	0.31	12.3	0.55	0.74	0.89	56	72	79	2.7	0.40
Berlin, Konzerthaus	2.20	2.00	1.80	0.23	0.25	0.25	13.6	0.49	0.76	0.83	50	70	76	3.4	0.50
Vienna, Musikvereinssaal	2.25	2.04	1.80	0.25	0.28	0.28	13.4	0.59	0.78	0.86	57	72	78	2.8	0.41
Lenox, Ozawa Hall	2.00	1.80	1.70	0.24	0.27	0.26	12.6	0.56	0.81	0.82	56	72	77	3.2	0.42
Seattle, Benaroya Hall	2.15	1.80	1.65	0.24	0.29	0.28	11.5	0.50	0.77	0.82	58	75	80	2.6	0.47
Kyoto, Concert Hall	2.20	2.00	1.80	0.25	0.27	0.27	14.9	0.63	0.86	0.91	59	74	79	3.4	0.48
Amsterdam, Concert Hall	2.20	2.05	1.80	0.28	0.30	0.30	14.6	0.64	0.82	0.90	62	74	80	2.9	0.41
Lucerne, Concert Hall	2.10	1.90	1.65	0.25	0.28	0.29	12.2	0.55	0.76	0.86	59	73	80	2.7	0.46
Tokyo, Tokyo Opera City	2.07	1.99	1.84	0.20	0.21	0.20	12.5	0.43	0.67	0.69	44	66	70	3.9	0.48
										Avg	56	72	78	3.1	0.45

[a]Assumed about same as average Baltimore & Manchester.

Philharmonie Hall the ceiling is partly covered with, "136 pyramidal-shaped, combination sound-diffusing, low-frequency Helmholtz-resonator-type absorbing boxes" (Beranek, 2004). Because the absorption of the Helmholtz resonators is mainly at low frequencies, the residual absorption coefficients in the 125–1000 Hz bands are also high, as expected.

X. AUDIENCE ABSORPTION COEFFICIENTS IN NINE RECTANGULAR HALLS

Using the Sabine equation and the basic data from Table I, the audience absorption coefficients for nine rectangular (shoebox) halls are given in Table III and are plotted in Fig. 3. Curve A is the average for the Kyoto and Amsterdam Halls which have heavily upholstered chairs. Curve B shows

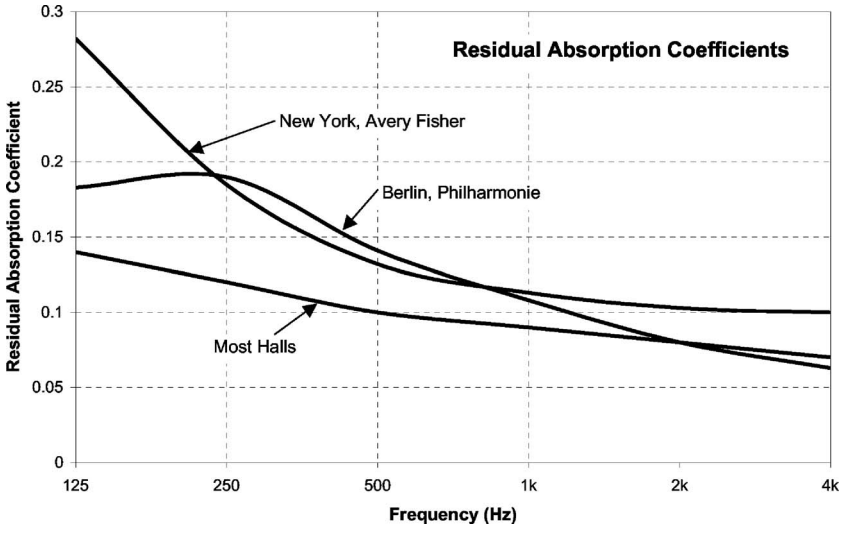


FIG. 2. Residual absorption coefficients using Sabine equation for Berlin Philharmonie and New York Avery Fisher Halls, shown along with the residual coefficients assigned to most of the halls.

TABLE III. Audience absorption coefficients for the rectangular halls in this study. The Sabine equation was used to determine the coefficients in the upper table. Eyring coefficients can be determined by multiplying the Sabine coefficients by the ratios in the lower table.

	Octave-band midfrequencies (Hz)					
Sabine equation	125	250	500	1000	2000	4000
Kyoto, Concert Hall	0.63	0.76	0.87	0.91	0.91	0.88
Amsterdam, Concertgebouw	0.64	0.71	0.82	0.90	0.90	0.89
Boston Symphony Hall	0.55	0.66	0.74	0.80	0.89	0.85
Berlin, Konzerthaus	0.49	0.64	0.76	0.79	0.83	0.78
Vienna, Musikvereinssaal	0.57	0.65	0.78	0.85	0.86	0.78
Lenox, Ozawa Hall	0.56	0.68	0.81	0.87	0.82	0.90
Seattle, Benaroya Hall	0.50	0.67	0.77	0.83	0.82	0.72
Lucerne, Concert Hall	0.55	0.65	0.76	0.84	0.86	0.80
Tokyo, Tokyo Opera City Hall	0.43	0.54	0.67	0.74	0.69	0.66
Average Sabine Audience Abs. Coef.	0.56	0.68	0.79	0.85	0.86	0.82
Eyring equation	The Erying audience absorption coefficients are equal to the above coefficients multiplied by the ratio "α_{ey}/α_{tot}" below					
	Ratio "α_{ey}/α_{tot}"					
Kyoto, Concert Hall	0.88	0.88	0.88	0.88	0.88	0.88
Amsterdam, Concertgebouw	0.88	0.87	0.87	0.86	0.86	0.87
Boston Symphony Hall	0.88	0.80	0.87	0.87	0.86	0.87
Berlin, Konzerthaus	0.90	0.89	0.89	0.89	0.89	0.89
Vienna, Musikvereinssaal	0.89	0.88	0.88	0.87	0.87	0.88
Lenox, Ozawa Hall	0.89	0.88	0.88	0.88	0.88	0.88
Seattle, Benaroya Hall	0.89	0.88	0.87	0.87	0.87	0.89
Lucerne, Concert Hall	0.89	0.88	0.87	0.87	0.87	0.88
Tokyo, Tokyo Opera City Concert Hall	0.91	0.90	0.90	0.89	0.90	0.92
Ratio Average	0.89	0.87	0.88	0.88	0.88	0.89
Average Eyring Audience Abs. Coef.	0.50	0.59	0.69	0.74	0.75	0.72

the average of the coefficients for the Boston, Berlin Konzerthaus, Vienna, Lenox, Seattle, and Lucerne Halls, which have light to medium upholstered chairs (Beranek, 2004). Curve C is for the TOC Hall. The average Sabine coefficients for all rectangular halls are given in the middle of Table III.

The audience absorption coefficients for TOC (curve C) are lower at all frequencies than those for the six halls of curve B. Three reasons probably account for this difference. First, the sound waves will involve the larger area of the residual surfaces more of the time than in other rectangular halls, because, from Table II, it is seen that for TOC S_R/S_T is the highest, 3.9, compared to 3.0 for the others. Second, it is possible that the unique pyramidal ceiling creates a somewhat different sound field at the surface of the audience area. Third, as will be shown later, the unoccupied chair absorption for TOC is lower.

Using the Eyring equation, the ratios α_{ey}/α_{tot} are plotted in the lower half of Table III, to show the range. This ratio only varies in these halls from 0.87 to 0.90—averaging 0.875. The Eyring coefficients can be found from the prod-

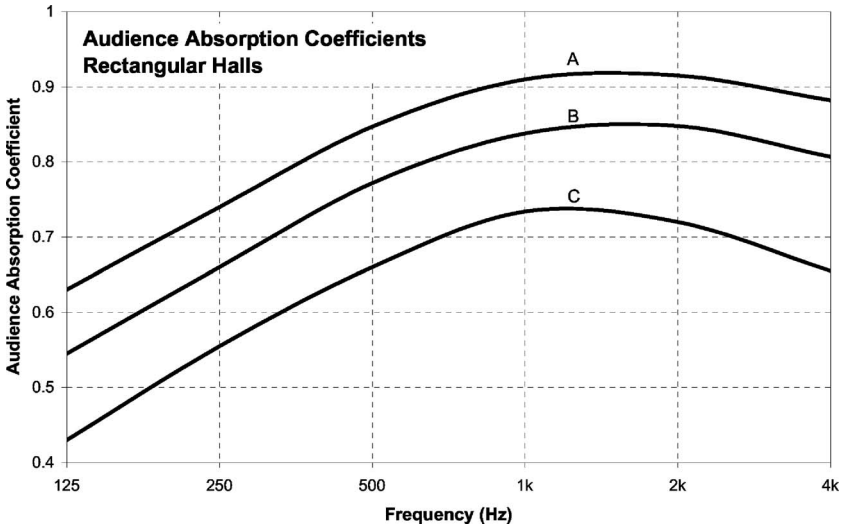


FIG. 3. Audience absorption coefficients for rectangular halls using Sabine equation. (A) Average of Kyoto and Amsterdam Halls. (B) Average of Boston, Berlin Konzerthaus, Vienna, Lenox, Seattle, and Lucerne Halls. (C) Tokyo Opera City Hall.

TABLE IV. Audience absorption coefficients for the nonrectangular halls in this study. Sabine coefficients are shown in the upper table and Eyring coefficients in the lower.

	Octave-band midfrequencies (Hz)					
Sabine equation	125	250	500	1000	2000	4000
Sapporo, Kitara Concert Hall	0.696	0.825	0.926	0.958	0.966	0.972
Munich, Philharmonie am Gasteig	0.797	0.827	0.952	0.950	0.989	0.990
Rotterdam, DeDoelen Concert Hall	0.642	0.836	0.942	0.966	0.898	0.827
New York, Carnegie	0.634	0.729	0.857	0.910	0.932	0.893
Cleveland, Severance Hall	0.679	0.755	0.834	0.927	0.931	0.905
Baltimore, Meyerhoff Hall	0.544	0.678	0.775	0.800	0.928	1.008
Manchester, Bridgewater Hall	0.558	0.694	0.811	0.838	0.867	0.771
Tokyo, Suntory	0.598	0.684	0.811	0.812	0.779	0.674
Buffalo, Kleinhans	0.391	0.613	0.671	0.807	0.848	0.755
Berlin, Philharmonie (see note)	0.557	0.673	0.775	0.827	0.801	0.778
New York, Avery Fisher Hall (note)	0.573	0.700	0.801	0.869	0.923	0.796
Av. Sabine Audience Absorp. Coef.	0.628	0.740	0.848	0.886	0.901	0.861
	Octave-band midfrequencies (Hz)					
Eyring equation	125	250	500	1000	2000	4000
Sapporo, Kitara Concert Hall	0.610	0.720	0.806	0.834	0.841	0.849
Munich, Philharmonie am Gasteig	0.687	0.718	0.818	0.819	0.854	0.856
Rotterdam, DeDoelen Concert Hall	0.558	0.711	0.797	0.817	0.768	0.718
New York, Carnegie	0.553	0.628	0.730	0.770	0.785	0.759
Cleveland, Severance Hall	0.587	0.650	0.717	0.787	0.810	0.792
Baltimore, Meyerhoff Hall	0.480	0.591	0.670	0.691	0.791	0.854
Manchester, Bridgewater Hall	0.494	0.607	0.703	0.727	0.754	0.681
Tokyo, Suntory	0.529	0.601	0.708	0.712	0.686	0.604
Buffalo, Kleinhans	0.350	0.529	0.576	0.678	0.710	0.643
Berlin, Philharmonie	0.482	0.571	0.658	0.709	0.689	0.675
New York, Avery Fisher Hall	0.481	0.596	0.683	0.737	0.781	0.686
Av. Eyring Audience Absorp. Coef.	0.546	0.639	0.729	0.760	0.776	0.747

ucts of the Sabine coefficients by these ratios. The average Eyring coefficients for the rectangular halls are given in the bottom line of Table III.

XI. AUDIENCE ABSORPTION COEFFICIENTS IN NONRECTANGULAR HALLS

The audience absorption coefficients for nonrectangular halls were calculated using the parameters in Table I. The results, for both formulas, are presented in Table IV and are plotted in Fig. 4. Curve A is for three halls, Sapporo, Munich, and Rotterdam, which have chairs with thicker upholstery than for the other halls. Curve B is for Carnegie and Cleveland Halls. Curve C is for the Baltimore, Manchester, and Tokyo-Suntory Halls. Curve D is for the Buffalo Hall which has chairs with minimum upholstering. In a section following, it will be seen that the unoccupied chairs in Carnegie Hall absorb more sound than the chairs in Baltimore and Manchester Halls.

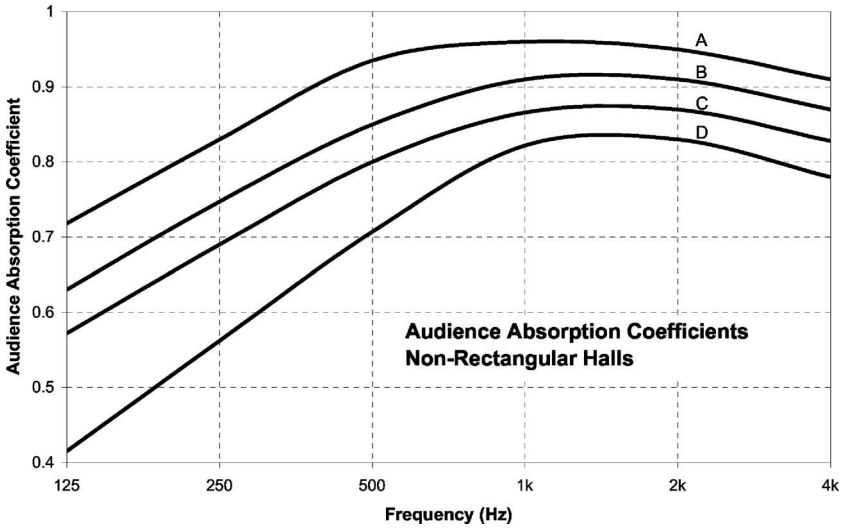


FIG. 4. Audience absorption coefficients for nonrectangular halls using Sabine equation. (A) Average of Sapporo, Munich, and Rotterdam halls. (B) New York, Carnegie, and Cleveland Halls. (C) Average of Baltimore, Manchester, and Tokyo Suntory Halls. (D) Buffalo Hall.

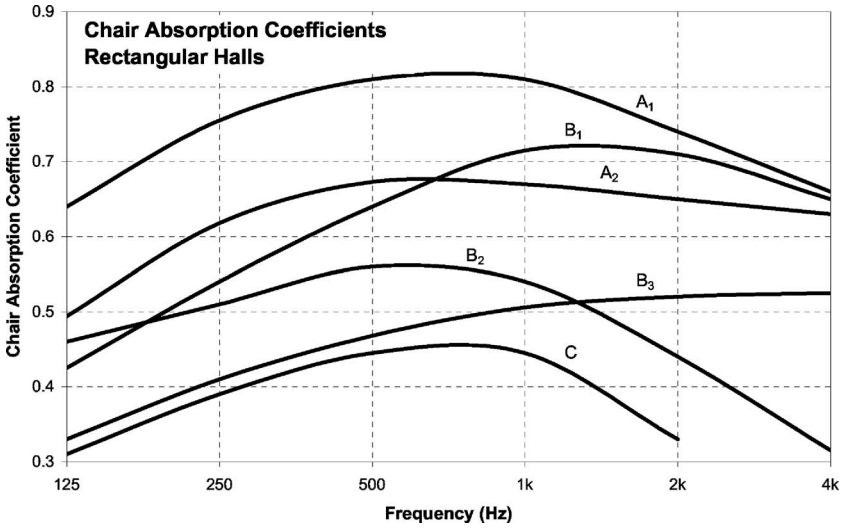


FIG. 5. Chair absorption coefficients for rectangular halls using Sabine equation. (A_1) Kyoto Hall. (A_2) Amsterdam Hall. (B_1) Average of Berlin Konzerthaus, Lenox, Seattle, and Lucerne Halls. (B_2) Boston Hall. (B_3) Vienna Hall. (C) Tokyo Opera City Hall.

XII. CHAIR ABSORPTION COEFFICIENTS IN RECTANGULAR HALLS

The absorption coefficients for unoccupied chairs in rectangular halls are given in Fig. 5. Six curves appear here as compared to three for Fig. 3 because A_1-Kyoto and A_2-Amsterdam have nearly the same absorptions when occupied, and B_1-BerlinK, Lenox, Seattle, and Lucerne, B_2-Boston, and B_3-Vienna are nearly alike when occupied. The only explanation as to why A_2 here is much lower than A_1, and B_2 and B_3 are much lower than B_1 is that those halls have the lowest areas per seat, 0.41 compared to 0.48 for A_1-Kyoto and 0.47 for B_1-Berlin, Lenox, Seattle and Lucerne (see Table II). If valid, this means that crowded people better shield the chairs, making the audience absorption less dependent on the kind of chairs on which they are seated.

For B_2-Boston the high absorptions at low frequencies are explained as follows: The seats are mounted on a layer of plywood with large airspace beneath. The high unoccupied chair absorption at 125 Hz is caused by the undamped plywood, but when occupied, it is significantly damped and assumes a greater mass due to peoples weight. Also, the seats are upholstered with leather over a 2 cm layer of felt so that there is no porosity, which means low absorption at higher frequencies. In Vienna, the rear balcony seats are bare wood, and all seats are very lightly upholstered.

The lowest curve of Fig. 5, C, is the chair absorption for the TOC. The rapid decrease in sound absorption at high frequencies for TOC and several of the other halls needs explanation. In an upholstered seat, the sound waves must penetrate the upholstery covering to get to the layer of foam or other material beneath. Several constructions can inhibit this penetration: (1) A thin plastic sheet that exists between the upholstery covering and the cushion, thus preventing the free flow of air. Such a sheet has a mass, which usually is light enough to vibrate freely for the lowest two bands, but that vibrates less and less freely as the frequency gets higher. (2) The upholstery covering has high flow resistance. This means that the sound wave is reflected off of the upholstery covering. But upholstering covering is a mass, and may vibrate freely at low frequencies. (3) The upholstery covering may have been back sprayed, which means it is nonporous—equivalent to very high flow resistance. (4) Upholstery covering is nonporous and heavy.

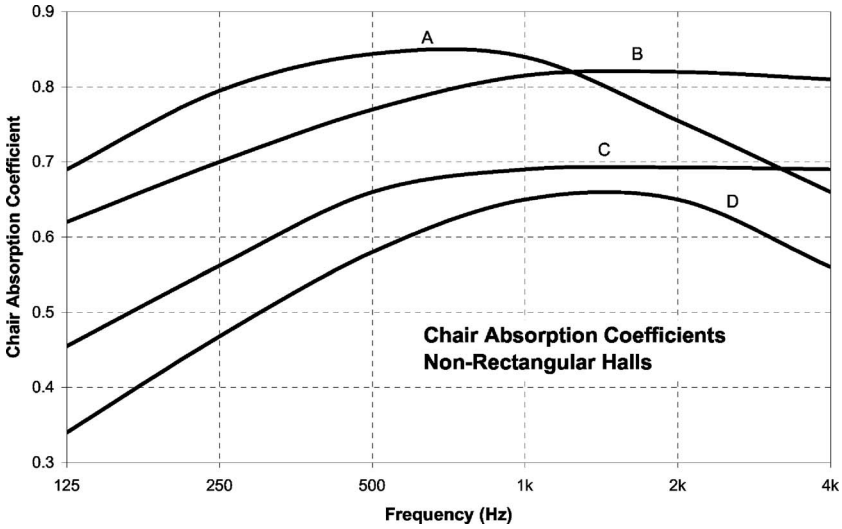


FIG. 6. Chair absorption coefficients for nonrectangular halls using Sabine equation. (A) Average of Sapporo, Munich, and Rotterdam Halls. (B) New York Carnegie Hall. (C) Average of Baltimore and Manchester Halls. (D) Buffalo Hall.

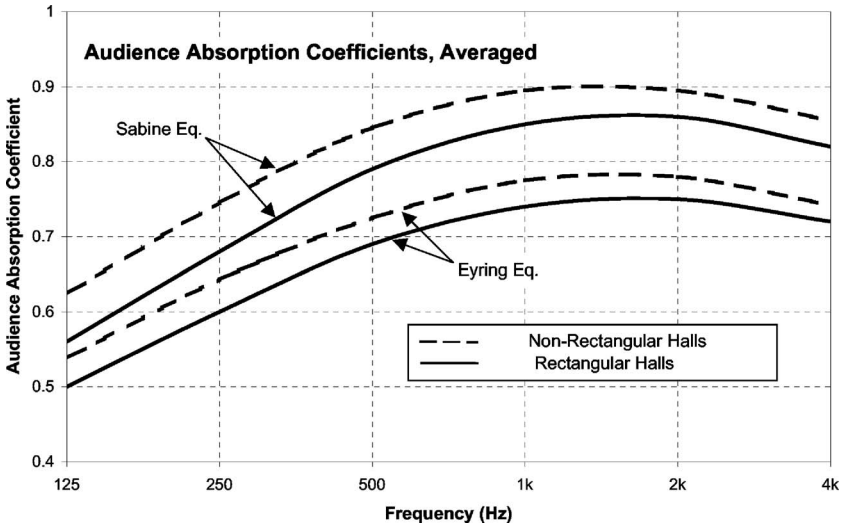


FIG. 7. Audience absorption coefficients: Average of all halls. Upper curves: Sabine equation; Lower curves; Eyring equation. Dashed lines: Nonrectangular halls. Solid lines: Rectangular halls. Data are from Tables III and IV.

Curve B_2 of Fig. 5 is for Boston Symphony Hall where the covering is leather, i.e., nonporous and heavy. For the TOC Hall, my notes as acoustical consultant say that there is a vinyl sheet underneath the upholstery covering (not planned). The chairs of Vienna and TOC were tested in a reverberation chamber and their unoccupied absorptions were alike at 500–4000 Hz, but the TOC chairs in that test did not have a vinyl sheet. The difference at low frequencies between the chair absorptions shown by curve B_3-Vienna and curve C-TOC is caused by the larger ratio S_R/S_T (See Table II). The chairs in the Kyoto Hall are upholstered on the rear of the seat back, which is not true for any of the other nineteen halls in this study and from curve A_1 of Fig. 5 the chair absorption for Kyoto is the highest of all halls.

XIII. CHAIR ABSORPTION COEFFICIENTS IN NONRECTANGULAR HALLS

The absorption coefficients for unoccupied chairs in nonrectangular halls are given in Fig. 6. Curve A is the average of coefficients for the Sapporo, Munich, and Rotterdam Halls, all of which have heavy upholstering (but not on rear of seatback). Curve B is for New York Carnegie Hall. Curve C is the average for Baltimore and Manchester Halls. Curve D is for Buffalo Hall where the chair upholstering is minimum. Except for variability at the high frequencies, the shapes of the curves resemble those of Fig. 4. The letterings for the curves are consistent in Figs. 4 and 6.

XIV. THEORETICAL RELATION OF HALL VOLUME TO AUDIENCE ABSORPTION

From Table II for nonrectangular halls it is seen that the percentage of absorption in the seating areas S_T to that in the residual areas S_R, $\alpha_T S_T/(\alpha_T S_T+\alpha_R S_R)$, is 77%. Assuming this percentage, on average, Eq. (1) becomes

$$V \approx 8.1(T_{60}S_T)\alpha_T \quad \text{(nonrectangular halls, 500 Hz)}. \tag{15}$$

For rectangular halls where the percentage average is 72%

$$V \approx 8.5(T_{60}S_T)\alpha_T \quad \text{(rectangular halls, 500 Hz)}. \tag{16}$$

Thus, for both types, the volume is directly related to the audience absorption coefficient.

XV. HALL VOLUME RELATED TO ROOM SHAPE AND SEAT UPHOLSTERING

A. Room shape

Audience absorption coefficients for the nonrectangular-shaped halls are greater than those for the other halls as seen in the plots of Fig. 7, taken from Tables III and IV. At 500 Hz, the ratio of the Sabine coefficients is 0.84/0.79 = 1.06, and at 1000 Hz 0.9/0.85 = 1.06. From this ratio and Eqs. (15) and (16), the cubic volume of a nonrectangular hall must be about 6% greater than that for a rectangular one, assuming the same audience size, reverberation time, and type of upholstering.

TABLE V. Audience absorption coefficients, Sabine equation, related to the degree of chair upholstering.

	Audience absorption coefficients					
Rectangular halls	125	250	500	1000	2000	4000
Heavily upholstered seats	0.63	0.75	0.85	0.91	0.91	0.88
Medium upholstered seats	0.55	0.67	0.79	0.84	0.85	0.83
Nonrectangular halls						
Heavily upholstered seats	0.73	0.82	0.90	0.95	0.95	0.92
Medium upholstered seats	0.63	0.75	0.84	0.90	0.90	0.88

TABLE VI. Upholstery details on seats in 20 concert halls.

Amount of upholstering	Front side of seat back	Rear side of seat back	Top of seat bottom	Arm rests
Heavily upholstered	7.5 cm	Sometimes[a]	10	2 cm
Medium upholstered	2.5 cm	0	5	Solid
Lightly upholstered	1.5 cm	0	2.5	Solid
Tokyo TOC[b]	65%, 2 cm	0	5	Solid

[a]Only in Kyoto in this paper.
[b]See text for details of upholstering.

B. Seat upholstering

A primary consideration in design of a concert hall is how much will the cubic volume be affected by the amount of seat upholstering if the reverberation time T_{60} and the audience size S_T are chosen. For rectangular halls, Fig. 3, the ratio between the α_T's at 500 Hz, curves A and B, is 1.1, and at 1000 Hz, 1.09 indicating a need for 10% greater cubic volume when the seats are heavily upholstered. For nonrectangular halls, Fig. 4, from curves A and B/C, the ratio of the α_T's at 500 Hz, is 1.13 and at 100 Hz, 1.05, indicating a 9% greater cubic volume. Thus, a 9%–10% increase in cubic volume is needed for a hall in which the seats are planned to be heavily upholstered as compared to light to medium upholstering.

In both rectangular and nonrectangular halls the percent increase of audience absorption with the amount of upholstering is highest in the lowest frequencies, but the increase even occurs at the highest frequencies where one might expect that the listeners' bodies would cover the absorbing surfaces (Table V). Higher audience absorption means lower sound levels in the room. At 63 Hz, the *loudness* halves for 5 dB drop in sound level. At 125 Hz, the loudness halves for 8 dB drop in sound level, while at higher frequencies the loudness halves for 10 dB drop. This means that the *loudness* in sones at low frequencies will decrease more rapidly with a change in absorption than at high frequencies, i.e., at 63 Hz a 10 dB drop will decrease the loudness by a factor of 4, while at 500 Hz and above a drop of 10 dB decreases the loudness by a factor of 2. Sounds in the 67–125 Hz bands frequency range are produced by the lower-pitched instruments in an orchestra.

Several years ago a questionnaire survey was sent to managers of concert halls asking them to describe the seats in their halls. The responses revealed that seats today are generally of shaped plywood, with no upholstering on the rear of the seat back or on the bottom of the seat. But, different degrees of upholstering were reported on other surfaces as given in Table VI.

XVI. CONCLUSIONS

The Sabine equation can be used for calculating reverberation times for a room if the Sabine sound absorption coefficients that are employed were previously determined in a site similar to the room in consideration. Sabine absorption coefficients must be allowed to take on all values from zero to infinity—false is the concept that a Sabine coefficient of 1.0 equates to complete sound absorption by a surface (as stated by Sabine, but contradicted by his own data). The Sabine absorption coefficients are easily derived and the Eyring absorption coefficients can be directly derived from Sabine coefficients because they are rigidly linked together by the logarithm. In concert halls the ratio of Eyring to Sabine coefficients is about 0.85, but in rooms with short reverberation times, the ratio may even go lower than 0.5.

Sound absorption by a seated audience is found to be higher at all frequencies if the sound absorption of the chairs on which they are seated is higher. The volume of a concert hall must be increased by approximately 10% if heavily upholstered seats are employed instead of light to medium upholstered seats. The sound absorption by an audience in halls where clear upper spaces for reverberation do not exist (usually in nonrectangular halls) is higher than that in halls were unimpeded surfaces above the top balcony are present (usually in rectangular halls). The volume of a concert hall must be about six percent greater if a design is selected that does not embody a large reverberant space above the top balcony.

ACKNOWLEDGMENTS

The author wishes to acknowledge the criticisms and recommendations of those who reviewed the first draft of this paper: W. B. Joyce, H. Mueller, M. Nagata, K. Oguchi, T. Hidaka, N. Nishihara, M. Barron, C. Jaffe, B. Markham, H. Tachibana, J. Bradley, R. Pirn, and D. Doria. The author is deeply indebted to Ben Markham at Acentech, Inc., who kindly prepared the figures.

[1]The Eyring equation was derived by Pat Norris [presented as an oral footnote to a paper by Norris and Andree at an Acoustical Society of America meeting in Spring (1929)], several months before Eyring's presentation. It was first published in an appendix to Knudsen's book (1932), obviously submitted to Knudsen earlier. Knudsen, wrote to Beranek "I recognized that he…was the first to derive this equation among American acousticians, and therefore I included his derivation in my *Architectural Acoustics*." The exact date of derivation of this equation by Schuster and Waetzmann, is not known, but it was published in 1929. Cremer and Mueller (1982) wrote, "This substitution $[-\ln(1-\alpha)]$ was first introduced by Fokker" (Fokker, 1924).

Andree, C. A. (**1932**). "The effect of position on the absorption of materials for the case of a cubical room," J. Acoust. Soc. Am. **3**, 535–551.
Barron, M., and Coleman, S. (**2001**). "Measurements of the absorption by auditorium seating—A model study," J. Sound Vib. **239**, 573–587.
Beranek, L. L. (**1960**). "Audience and seat absorption in large halls," J. Acoust. Soc. Am. **32**, 661–670.
Beranek, L. L. (**1962**). *Music, Acoustics, and Architecture* (Wiley, New York).
Beranek, L. L. (**1969**). "Audience and seat absorption in large halls. II," J. Acoust. Soc. Am. **45**, 13–19.
Beranek, L. L. (**1996**). *Acoustics* (Acoustical Society of America, Melville, NY), pp. 299–306.
Beranek, L. L. (**2004**). *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*, 2nd ed. (Springer, New York).
Beranek, L. L., and Hidaka, T. (**1998**). "Sound absorption in concert halls by seats, occupied and unoccupied, and by the hall's interior surfaces," J. Acoust. Soc. Am. **104**, 3169–3177.
Bistafa, S. R., and Bradley, J. S. (**2000**). "Predicting reverberation times in a simulated classroom," J. Acoust. Soc. Am. **108**, 1721–1731.
Bradley, J. S. (**1992**). "Predicting theater chair absorption from reverberation room measurements," J. Acoust. Soc. Am. **91**, 1514–1524.
Bradley, J. S. (**1996**). "The sound absorption of occupied auditorium seating," J. Acoust. Soc. Am. **99**, 990–995.

Carrol, M. M., and Chen, C. F. (**1977**). "Decay of reverberant sound in a spherical enclosure," J. Acoust. Soc. Am. **62**, 1442–1446.
Cremer, L., and Mueller, H. A. (**1982**). *Principles and Applications of Room Acoustics*, Vol. **1**, translated by T. J. Schultz (Applied Science Publishers, London), pp. 214–260.
Davis, W. J., Orlowski, R. J., and Lam, Y. W. (**1994**). "Measuring auditorium seat absorption," J. Acoust. Soc. Am. **96**, 879–888.
Eyring, C. F. (**1930**). "Reverberation time in 'dead' rooms," J. Acoust. Soc. Am. **1**, 168.
Fokker, A. D. (**1924**). Physica (The Hague) **4**, 262.
Hidaka, T. (**2006**), "Reverberation time, mean-free-path, and sound absorption in concert halls—Numerical examination by computer simulation," J. Acoust. Soc. Am. **119**, 3430.
Hidaka, T., Nishihara, N., and Beranek, L. L. (**2001**). "Relation of acoustical parameters with and without audiences in concert halls and a simple method for simulating the occupied state," J. Acoust. Soc. Am. **109**, 1028–1042.
Hodgson, M. (**1993**). "Experimental evaluation of the accuracy of the Sabine and Eyring theories in the case of nonlow surface absorption," J. Acoust. Soc. Am. **94**, 835–840.
Joyce, W. B. (**1978**). "Exact effect of surface roughness on the reverberation time of a uniformly absorbing spherical enclosure," J. Acoust. Soc. Am. **65**, 1429–1436.
Joyce, W. B. (**1980**). "Power series for the reverberation time," J. Acoust. Soc. Am. **67**, 564–571.
Kirdegaard, D. (**1996**). "Sound absorption of occupied chairs as a function of chair design and audience clothing," J. Acoust. Soc. Am. **99**, 2459(A).
Knudsen, V. O. (**1932**). *Architectural Acoustics* (Wiley, New York), Appendix.
Kuttruff, H. (**1973**). *Room Acoustics* (Wiley, New York), p. 113.
Mange, G. (**2005**). "Measurement of absorption,' Noise Control Eng. J., **53**, 268–270.
Nishihara, N., Hidaka, T., and Beranek, L. (**2001**). "Mechanism of sound absorption by seated audience in halls," J. Acoust. Soc. Am. **110**, 2398–2411.
Norris, R. F., and Andree, C. A. (**1929**). "An instrumental method of reverberation measurement," J. Acoust. Soc. Am. **1**, 32.
Sabine, W. C. (**1900–1915**). *Collected Papers on Acoustics* (Peninsula, Los Altos, CA).
Schroeder, M. R. (**1973**). "Computer models for concert hall acoustics," Am. J. Phys. **41**, 461–471.
Schuster, K., and Waetzmann, E. (**1929**). "Reverberation in closed rooms," Ann. D. Phys. **1**, 671.

An experimental evaluation of regular polyhedron loudspeakers as omnidirectional sources of sound

Timothy W. Leishman,[a] Sarah Rollins, and Heather M. Smith
Acoustics Research Group, Department of Physics and Astronomy, Eyring Science Center, Brigham Young University, Provo, Utah 84602

(Received 12 December 2005; revised 15 June 2006; accepted 15 June 2006)

Multiple-driver dodecahedron loudspeakers are commonly used in room acoustics measurements as omnidirectional sources of sound. Yet they and other regular polyhedron loudspeakers become "multidirectional" above their omnidirectional cutoff frequencies (often near 1 kHz). Because these cutoff frequencies normally fall within common measurement bandwidths, one might question whether anything is really extraordinary about the dodecahedron loudspeaker or whether another regular polyhedron geometry would actually produce better average omnidirectionality over these bandwidths. This paper explores these questions through measured data, analysis, and comparison of frequency-dependent directivities of several regular polyhedron loudspeaker prototypes. It provides insights into their radiation properties and introduces an alternative method of quantifying omnidirectionality: the area-weighted spatial standard deviation of radiated levels over a free-field measurement sphere. It compares this method to the ISO 3382:1997(E) standard method, revealing certain discrepancies between the two approaches. A dodecahedron loudspeaker is shown to produce a relatively high cutoff frequency and reasonable radiation uniformity over measurement bandwidths. However, it does not necessarily excel as a better omnidirectional source than other regular polyhedron loudspeakers. A tetrahedron loudspeaker with an equal midradius provides the best average radiation uniformity over a 4 kHz bandwidth, even though it exhibits the lowest cutoff frequency. [DOI: 10.1121/1.2221552]

PACS number(s): 43.55.Mc, 43.38.Ja, 43.38.Hz, 43.20.Rz [NX] Pages: 1411–1422

I. INTRODUCTION

In an effort to employ consistent omnidirectional sources in room acoustics measurements, acousticians have long resorted to multiple-driver loudspeakers based on regular polyhedra or platonic solid enclosure geometries. Regular polyhedron loudspeakers (RPLs) typically incorporate small in-phase direct-radiator drivers of consistent manufacture, mounted centrally in the faces of their sealed enclosures. One international standard suggests that their acceptance, particularly that of the dodecahedron loudspeaker, has become nearly universal.[1]

Despite the predominance of the dodecahedron loudspeaker, several important questions should be addressed. First, is anything really extraordinary about the dodecahedron loudspeaker—especially when its radiation is considered over *entire* measurement bandwidths? Second, how do directional characteristics of loudspeakers based on other platonic solid geometries compare to those based on the dodecahedron geometry over these same bandwidths? Third, is it possible that other platonic solid geometries actually provide better omnidirectional radiation when averaged over the bandwidths? This paper will respond to these questions to enhance understanding and implementation of the sources as measurement tools.

A regular polyhedron is a convex polyhedron composed of identical regular polygonal faces (equal side lengths and vertex angles). The five regular polyhedra are the tetrahedron (4 faces), hexahedron or cube (6 faces), octahedron (8 faces), dodecahedron (12 faces), and icosahedron (20 faces).[2] If an RPL is geometrically centered at the origin of the spherical coordinate system, its normal facial axes are distributed uniformly throughout the combined angular coordinates. However, despite their geometric uniformities and symmetries, RPLs are known to exhibit undesirable directional behaviors (i.e., nonspherical radiation patterns) above frequencies that may be termed omnidirectional "cutoff frequencies." Several factors contribute to the frequency-dependent departure from the omnidirectional ideal. They include interference between radiation from discrete drivers, inherent directivities of individual drivers, and diffraction effects.

An omnidirectional acoustic source is ideal for many applications because it radiates sound *equally* in *all* directions. A *unidirectional* source radiates sound predominantly in only *one* direction. A *multidirectional* source may be defined as one radiating sound predominantly in *several* directions, between the extremes of an omnidirectional source and a unidirectional source. Over large portions of common measurement bandwidths, RPLs are typically multidirectional sources.

In a previous study of RPLs, Tarnow used spherical harmonic expansions, group representation theory, and several source idealizations to computationally characterize low- to mid-frequency radiation and omnidirectional cutoff frequencies.[3,4] In his work, he found that cutoff frequencies should progressively increase for higher-order polyhedra with fixed effective enclosure radii. Close inspection of his work also suggests another important point: if an enclosure is

[a]Electronic mail: tim_leishman@byu.edu

reasonably sized to accommodate drivers with adequate low-frequency response and sound power output, the omnidirectional cutoff frequencies regularly fall within bandwidths of common room acoustics measurements. For example, his results suggest that a dodecahedron loudspeaker with an effective radius a begins to exhibit significant directivity above $ka=3$, where k is the acoustic wave number. For a common dodecahedron loudspeaker designed to balance omnidirectionality with desired sound power output, the effective radius typically places the omnidirectional cutoff frequency in the vicinity of 1 kHz. While this cutoff frequency can be increased by using smaller loudspeaker drivers and enclosures, low-frequency response and sound power output will generally suffer as a result. In a recent development, Witew and Behler employed multiple sources in a multiway configuration to maintain sufficient broadband omnidirectionality and sound power output for room acoustic parameter measurements.[5]

Recognition of undesirable directional behavior of RPLs is also apparent in established standards for room acoustics measurements.[1,6] They allow qualification of presumably omnidirectional sources through spatial averaging of free-field radiation over measurement arcs in a single plane and spectral averaging over proportional frequency bands. They also relax omnidirectional requirements at higher frequencies and ignore directional behavior above 5.6 kHz. One standard requires source rotations through three angular positions and averaging of subsequent measurement results when source directivity is found to significantly affect measured room acoustic parameters.[6] Nevertheless, none of these concessions changes the fact that presumably omnidirectional sources exhibit unprescribed and undesirable directivities at many frequencies of interest.

The authors have undertaken a research effort to further characterize RPL radiation through analysis and comparison of the frequency-dependent directivities of several RPLs over common measurement bandwidths. This paper focuses on key experimental findings of that effort. Two sets of RPLs were constructed with specific geometric properties to enable equitable comparisons. The first consisted of RPLs with equal midradii (EM). [The midradius of a regular polyhedron is the distance from its geometrical center to the midpoint of any edge (facial side).] The second consisted of RPLs with equal volume (EV) allocation per loudspeaker driver, resulting in progressively larger total enclosure volumes for higher-order polyhedra. While sources with other dimensions or geometries might also have been studied (including those of various RPLs with consistent spacing between adjacent loudspeaker drivers), the results presented here reveal several important trends. Special measurement and data processing tools were developed specifically for the effort. An alternative method of quantifying source omnidirectionlity was developed and compared to the current standard method found in ISO 3382:1997(E).[6] It involved an area-weighted spatial standard deviation of radiated levels evaluated over a free-field measurement sphere.

The following section discusses the experimental procedures. The subsequent sections present results and analysis, then compare the two methods of quantifying source omnidirectionality. The final section presents conclusions from the study and several suggestions for further work.

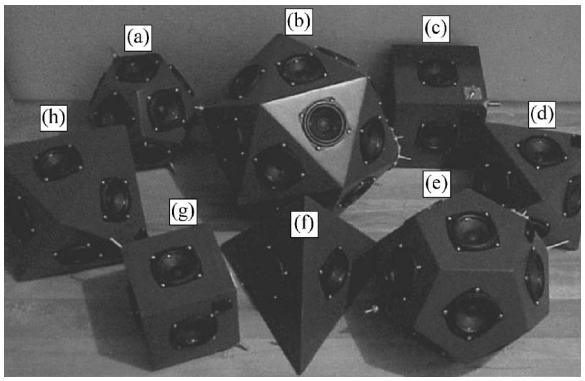


FIG. 1. Equal midradii (EM) and equal volume per driver (EV) RPLs. (a) EM dodecahedron. (b) EV icosahedron. (c) EV hexahedron. (d) EM octahedron. (e) EV dodecahedron. (f) EM/EV tetrahedron. (g) EM hexahedron. (h) EV octahedron.

II. EXPERIMENTAL PROCEDURES

Geometric properties of regular polyhedra, including their midradii and volumes, are often expressed in terms of edge or facial side lengths.[7,8] Established relationships were used to design and construct the eight RPLs shown in Fig. 1 with the properties listed in Table I. The loudspeakers consisted of one EM and one EV RPL for each polyhedron type, except the tetrahedron and icosahedron. A single tetrahedron loudspeaker was constructed to simultaneously satisfy both the EM and EV criteria. Only one EV icosahedron loudspeaker was constructed because the Aura NS3-194-8D drivers selected for the RPLs (with nominal 80 mm diameters) would not fit within a compact icosahedron satisfying the EM criterion. As indicated in Table I, the EM value was chosen to be 11.0 cm while the EV value (external enclosure volume per driver) was chosen to be 887.4 cm^3. The sealed enclosures were constructed of 1.9-cm-thick birch plywood. Approximately half of their sealed air volumes were filled with acoustically absorptive fibrous material. All RPLs were wired for equal in-phase signal to each driver and an overall nominal impedance between 4 and 8 Ω.

As suggested in Fig. 2, the radiated fields were measured by mounting each RPL on a narrow stand in an anechoic

TABLE I. External geometric properties for the eight experimental RPLs. The equal midradius (EM) value was chosen to be 11.0 cm while the equal volume per driver (EV) value was chosen to be 887.4 cm^3.

RPL	Edge length (cm)	Midradius (cm)	Volume per driver (cm^3)	Total volume (cm^3)
EM/EV tetrahedron	31.1	11.0	887.4	3 549
EM hexahedron	15.6	11.0	627.4	3 764
EM octahedron	22.0	11.0	627.4	5 020
EM dodecahedron	8.4	11.0	378.9	4 547
EV hexahedron	17.5	12.3	887.4	5 325
EV octahedron	24.7	12.3	887.4	7 099
EV dodecahedron	11.2	14.6	887.4	10 648
EV icosahedron	20.1	16.3	887.4	17 748

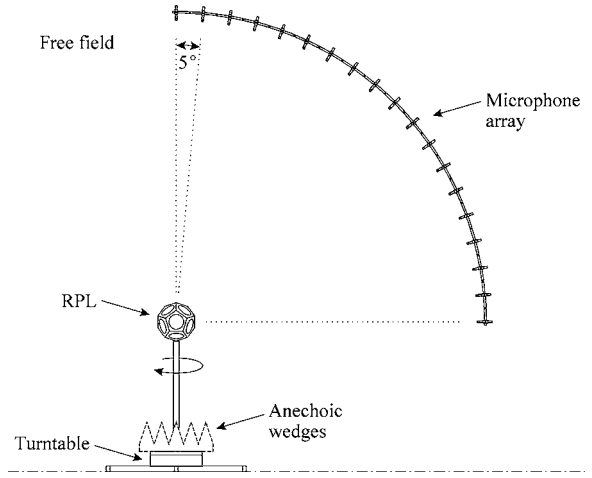


FIG. 2. Diagram of the experimental setup in an anechoic chamber. A dodecahedron is shown mounted on the turntable stand. The turntable is covered with wedges to maintain anechoic conditions. A 90° arc array of microphones senses the radiated field.

chamber, then rotating it under a quarter-circle microphone array with a computer-controlled turntable. Thick absorptive wedges were positioned on top of the turntable to reduce its reflection and scattering. The turntable was controlled to rotate in $\Delta\phi=5°$steps, producing measurement dependence in the azimuthal angle ϕ. Nineteen Larson Davis 2551 microphones (type 1 free-field electrets) were connected to The Modal Shop TMS426C01 preamplifiers and radially oriented in the 90° measurement arc. The arc had a 2.1 m radius and was positioned in the first quadrant of the vertical plane above the RPL. The axis of the top microphone $(\theta=0°)$ was directed along a vertical line running through the RPL center. The axis of the bottom microphone $(\theta=90°)$ was directed along a horizontal line running through the center. The curved array incorporated uniform $\Delta\theta=5°$ angular increments between adjacent microphones, enabling the same sampling in the polar angle as in the azimuthal angle. To obtain a full sphere of measured data, it was necessary to measure the top radiation hemisphere, turn the RPL upside down, align it, then measure the bottom hemisphere.

The microphones were individually calibrated at 1 kHz and connected to a multichannel dynamic signal analyzer based on Hewlett-Packard E1432A VXI cards. Their sensitivities were entered into the Data Physics DP620 control software and periodically checked for drift. The analyzer also provided a broadband pseudorandom excitation signal to drive the RPLs through a power amplifier. Frequency response and coherence functions between the excitation signal and the microphone outputs were generated between 0 and 20 kHz with 1600 uniform frequency bins (12.5 Hz bandwidths). The analyzer was configured to take 30 stable averages for each measurement. Frequency response functions were considered a good choice for the directivity measurements because they provided smooth frequency dependence in the measured data and many postprocessing options.

For each source rotation position in ϕ, the 19 frequency response functions in θ were automatically exported for postprocessing. The measurement set was repeated with each 5° turntable increment until the RPL had been rotated through a full 360°. Once both hemispheres were measured, a composite 2664-point set of complex frequency response functions was compiled to enable thorough characterization of the RPL radiation.

Observation of the frequency response and coherence data suggested that the low-frequency roll off of the RPLs was approximately 100 Hz. The anechoic cutoff frequency of the chamber was somewhat lower than 100 Hz. Nevertheless, those low-frequency limitations were inconsequential for the directivity measurements because all RPLs displayed consistent omnidirectional radiation near this spectral region.

III. MEASUREMENT RESULTS

To produce graphical representation of the RPL directivity patterns, normalized frequency response function levels were first determined using the formula

$$L_{m,n}(f) = 20 \log\left[\frac{|H_{m,n}(f)|}{|H_{m,n}(f)|_{\max}}\right], \tag{1}$$

where $|H_{m,n}(f)|$ is the modulus of the complex frequency response function at the point m,n on the measurement sphere and $|H_{m,n}(f)|_{\max}$ is the maximum of all $|H_{m,n}(f)|$ for the given frequency f. The indices m and n are integer multipliers of the 5° increments in θ and ϕ, respectively. Thus, $\theta_m=m\Delta\theta$, where $m=0,1,2,\ldots,M-1$, and $M=37$ is the number of measurement positions in θ. Similarly, $\phi_n=n\Delta\phi$, where $n=0,1,2,\ldots,N-1$, and $N=72$ is the number of measurement positions in ϕ. The levels were plotted for all RPLs at several frequencies to produce narrowband (12.5 Hz bandwidth) directivity balloons.

Animations of the balloons over frequency provided interesting insights into their frequency-dependent evolutions. As anticipated, the RPLs all produced nearly omnidirectional fields (spherical balloon plots) below about 1 kHz. However, their directivity patterns differed dramatically above this frequency. Figures 3 and 4 show the narrowband plots for the eight RPLs at 2 and 4 kHz, respectively, revealing several distinct lobe patterns and symmetries. Several factors contributed to the directivity patterns. For any given RPL and frequency, an individual loudspeaker driver has a radiation pattern that depends upon its cone diameter, other geometric features, and specific vibrational characteristics. It also depends upon the surrounding enclosure and diffraction effects. When all drivers are in simultaneous operation, the complex pressure produced by each and every driver superposes to generate the composite field and associated directivity pattern. Constructive and destructive interference is affected by the individual driver directivities, spacings, and angular orientations. Since the drivers are not perfectly matched, the pattern symmetries may not be ideally related to the associated polyhedron symmetries.

The presence of the lobes, which do not necessarily correspond to driver axes at these frequencies, demonstrate significant departure from the omnidirectional ideal. At 2 kHz, the larger EV RPLs appear to be more advanced in their multidirectional behaviors than their smaller EM RPL coun-

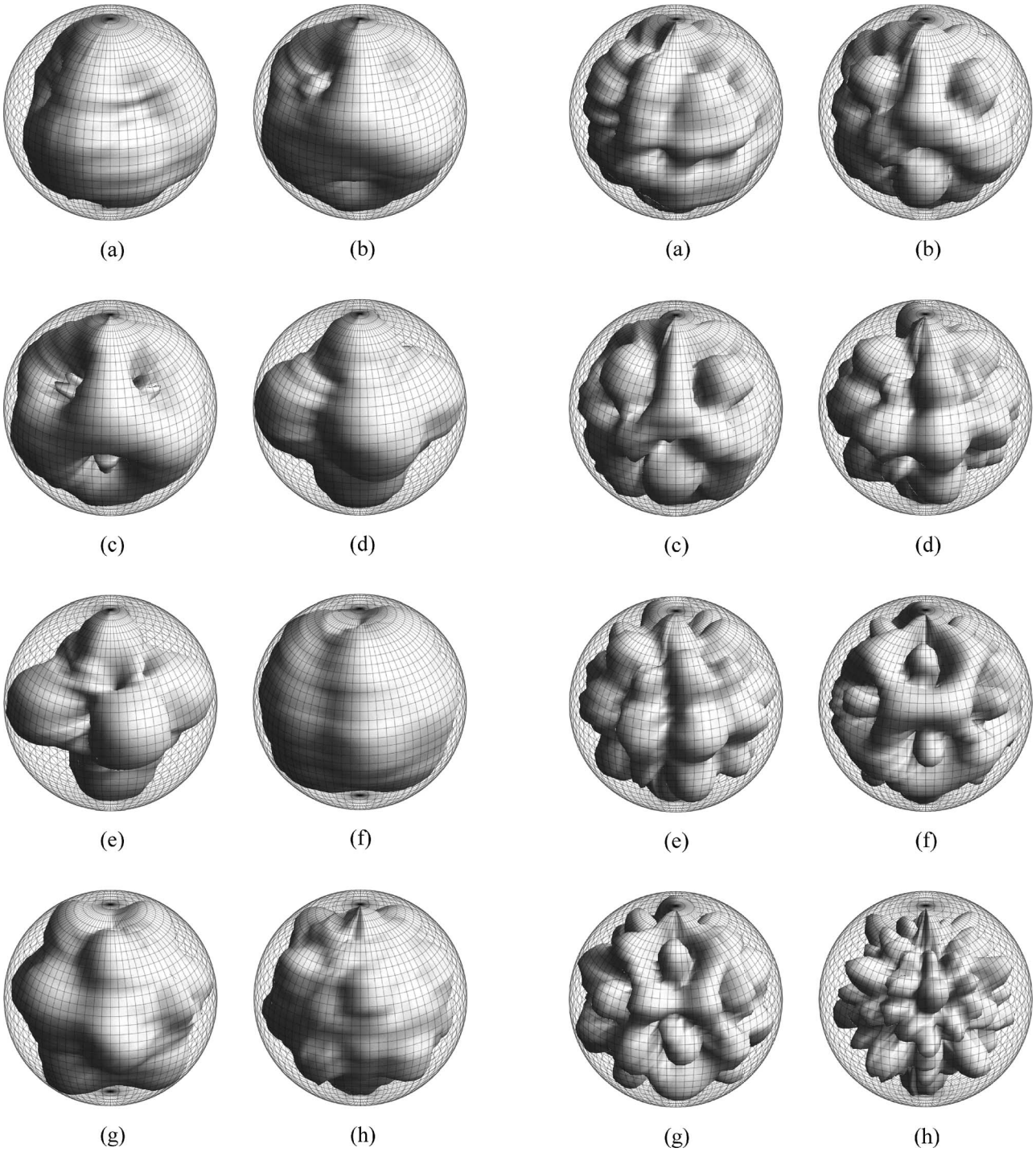


FIG. 3. Narrowband directivity balloon plots for all eight RPLs operating at 2 kHz. (a) EM/EV tetrahedron. (b) EM hexahedron. (c) EV hexahedron. (d) EM octahedron. (e) EV octahedron. (f) EM dodecahedron. (g) EV dodecahedron. (h) EV icosahedron. As shown in the plots, lobes become apparent above 1 kHz, providing clear evidence of departure from omnidirectional radiation. The surfaces of the superposed mesh spheres represent uniform normalized values of 0 dB. The origin represents a value of −20 dB, meaning that the mesh sphere radii are 20 dB.

FIG. 4. Narrowband directivity balloon plots for all eight RPLs operating at 4 kHz. (a) EM/EV tetrahedron. (b) EM hexahedron. (c) EV hexahedron. (d) EM octahedron. (e) EV octahedron. (f) EM dodecahedron. (g) EV dodecahedron. (h) EV icosahedron. The distinct lobes at this frequency generally demonstrate further departure from omnidirectional radiation. The radii of the superposed mesh spheres again represent 20 dB.

terparts. Interestingly, the tetrahedron loudspeaker appears to show greater omnidirectionality at both frequencies than several of the higher-order RPLs.

When considered over full-octave and third-octave proportional bands, the directivity balloons had patterns similar to the narrowband balloons, but with expected smoothing that increased substantially for larger bandwidths. From this observation alone, it is clear that one should exercise caution in using broad proportional-band measurements to characterize source directivity at higher frequencies.

IV. ANALYSIS

To consolidate the measurement results and improve understanding of the frequency-dependent RPL directivities,

the data was further analyzed using two methods: the ISO 3382 source qualification method, and an area-weighted spatial standard deviation method. Additional statistical analyses also provided insights for establishing RPL omnidirectionality.

A. ISO 3382

To regulate the impact of source directivity on room acoustics measurements, standards often state that excitation sources should be as omnidirectional as possible—with deviations from perfect omnidirectionality arising only within prescribed limits.[1,6] The source qualification procedure outlined in ISO 3382 is representative of those found in other standards. It requires that a source be driven with octave bands of pink noise to produce radiated free-field sound pressures. The pressure for each band is measured at a minimum radial distance of 1.5 m and averaged over 30° measurement arcs. These 30° "gliding" averages are then referenced to a 360° energetic average in the measurement plane to compute a "directional deviation" in decibels. The directional deviation must remain within maximum allowable limits to satisfy the qualification criteria. Because the standard specifically allows 5° field measurement increments to compute the averages, it is well suited to the data measured in this work.

Two types of filters were applied to the measured frequency response functions as postprocessing options. One was designed to equalize the RPLs for flat sound power response. The other was designed to weight either the equalized or unequalized responses as though the RPLs were driven by a pink noise signal. Since the standard suggests pink noise excitation and octave-band sound pressure level measurements without mention of equalization, the squared moduli of the weighted narrowband frequency response functions were first summed into base-2 octave bands[9] without sound power equalization. The directional deviation in each octave band was then calculated as the worst-case difference between various energetic gliding averages (six neighboring values in 30° arcs) of the equatorial measurement plane ($\theta_{18}=90°$) and the energetic average of all values in a full 360° circle of the same plane. Sound power equalization and pink-noise weighting were subsequently found to produce only small changes to the frequency-dependent curves.

The directional deviations for the EM and EV RPLs are plotted in Figs. 5 and 6, along with the maximum ISO 3382 deviation limits. The two lines shown for each RPL represent the largest positive and negative 30° arc deviations from the 360° energetic average. Such figures are commonly published by omnidirectional source manufacturers to demonstrate compliance with ISO standards and allow potential users to draw conclusions about their source directivities.

According to Fig. 5, all EM RPLs demonstrated nearly ideal omnidirectional radiation characteristics through the 500 Hz octave band. The tetrahedron loudspeaker was the first to depart from minimal deviation values beginning in the 1 kHz octave band. All other sources departed in the 2 kHz octave band. In that band, the hexahedron appears to be the most omnidirectional source, followed by the dodecahedron, the tetrahedron, then the octahedron. In the 4 kHz octave band, the octahedron appears to be the most omnidirectional source, followed by the tetrahedron, the hexahedron, then the dodecahedron. In the 8 kHz octave band, the hexahedron and octahedron appear to perform best, while the tetrahedron and dodecahedron appear to perform worst. These rankings are summarized in Table II, along with comparative rankings to be discussed later. Based on these results, the hexahedron appears to be the most omnidirectional EM RPL over the measurement bandwidth.

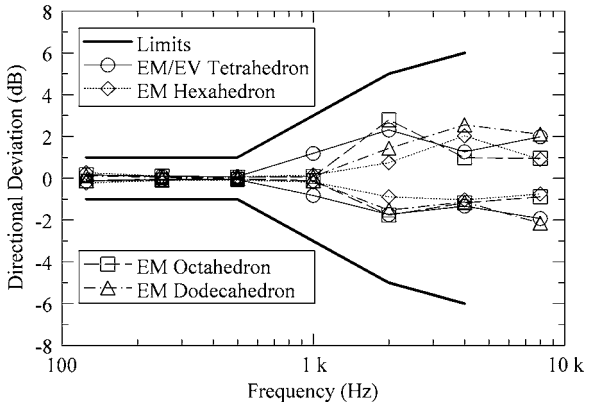


FIG. 5. Octave-band directional deviations for the EM RPLs based on an implementation of the ISO 3382 source qualification procedure. The extreme positive and negative deviation curves are plotted along with the maximum ISO 3382 deviation limits.

For the EV RPL results presented in Fig. 6, we again find the tetrahedron to be the first to depart from minimal deviation values in the 1 kHz octave band. Rankings for higher octave bands are summarized in Table II. From these results, the icosahedron appears to be the most omnidirectional EV RPL over the measurement bandwidth.

Given such comparative assertions, one may be justified in asking whether the data represented in the graphs is truly sufficient to establish omnidirectionality or to form a solid basis for rankings. A better approach would involve a more comprehensive understanding of the radiated fields and a pertinent summary of their characteristics.

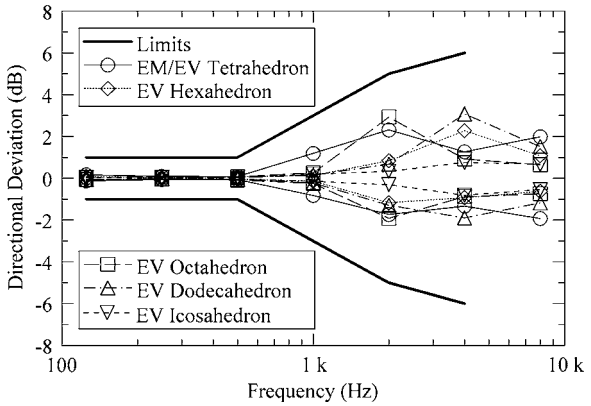


FIG. 6. Octave-band directional deviations for the EV RPLs based on an implementation of the ISO 3382 source qualification procedure. The extreme positive and negative deviation curves are plotted along with the maximum ISO 3382 deviation limits.

TABLE II. Approximate EM and EV RPL performance rankings for the 2, 4, and 8 kHz octave bands, according to results from the ISO 3382 and spatial standard deviation methods. Overall rankings are also given as composite rankings from the three octave bands.

	ISO 3382 method				Standard deviation method			
RPL	2 kHz	4 kHz	8 kHz	Overall	2 kHz	4 kHz	8 kHz	Overall
EM/EV tetrahedron	3	2	3–4	3	2	1	4	2
EM hexahedron	1	3	1–2	1	3	3	3	3
EM octahedron	4	1	1–2	2	4	4	2	4
EM dodecahedron	2	4	3–4	4	1	2	1	1
EM/EV tetrahedron	4	3	5	5	3	1	5	3–4
EM hexahedron	3	4	1–3	2	4	2	3	3–4
EV octahedron	5	1–2	1–3	3	5	4	4	5
EV dodecahedron	2	5	4	4	1	5	1–2	2
EV icosahedron	1	1–2	1–3	1	2	3	1–2	1

B. Area-weighted spatial standard deviation of levels

In the search for other means to characterize RPL directivity, an area-weighted spatial standard deviation was calculated from each complete set of 2664 narrowband measurements. Several standard deviation formulas (e.g., energetic, arithmetic level, unweighted, area-weighted, etc.) were explored to determine which provided results corresponding best to observed directional behaviors. Because the standard deviation should be zero for perfect omnidirectional radiation, it should be consistently small for all measured RPLs below 1 kHz to agree with observations. At any frequency above 1 kHz, it should agree (from a qualitative standpoint) with visualizations of the RPL balloon plots.

While an energetic standard deviation formulation was found to be useful and physically appropriate, another formulation was also found to favorably characterize the observations. It was the area-weighted spatial standard deviation of either normalized or unnormalized frequency response function levels,

$$\sigma_{\mathrm{AWL}}(f)=\sqrt{\frac{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}S_{m,n}[L_{m,n}(f)-\langle L_{m,n}(f)\rangle_S]^2}{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}S_{m,n}}}, \quad (2)$$

where the arithmetic area-weighted spatial average of levels is given by

$$\langle L_{m,n}(f)\rangle_S=\frac{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}S_{m,n}L_{m,n}(f)}{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}S_{m,n}}. \quad (3)$$

The factor $S_{m,n}$ in these equations is the area weighting factor (i.e., the effective sampling area per microphone on the measurement sphere of radius $r=2.1$ m), determined by surface integration over appropriate sections of the measurement sphere:

$$S_{m,n}=\int_{\phi_n-\Delta\phi/2}^{\phi_n+\Delta\phi/2}\int_0^{\Delta\theta/2}r^2\sin(\theta)d\theta d\phi$$
$$=2r^2\Delta\phi\sin^2\left(\frac{\Delta\theta}{4}\right)=\frac{4\pi r^2}{N}\sin^2\left(\frac{\Delta\theta}{4}\right),\quad m=0, \quad (4)$$

$$S_{m,n}\int_{\phi_n-\Delta\phi/2}^{\phi_n+\Delta\phi/2}\int_{\theta_m-\Delta\theta/2}^{\theta_m+\Delta\theta/2}r^2\sin(\theta)d\theta d\phi$$
$$=2r^2\Delta\phi\sin(\theta_m)\sin\left(\frac{\Delta\theta}{2}\right),\quad 1\leqslant m\leqslant 35, \quad (5)$$

$$S_{m,n}\int_{\phi_n-\Delta\phi/2}^{\phi_n+\Delta\phi/2}\int_{\pi-\Delta\theta/2}^{\pi}r^2\sin(\theta)d\theta d\phi$$
$$=2r^2\Delta\phi\sin^2\left(\frac{\Delta\theta}{4}\right)=\frac{4\pi r^2}{N}\sin^2\left(\frac{\Delta\theta}{4}\right),\quad m=36. \quad (6)$$

These definitions are consistent for all values of n (see Fig. 7). The sampling areas of the two poles have been segmented into N equal pie slices for convenience in the summations. As formulated, the standard deviation is an asymptotically unbiased estimator, but for a finite number of samples, it has a small downward bias. No assumption is made about a normal distribution of the levels.

Interestingly, this approach is similar to approaches mentioned by Schroeder,[10] D'Antonio,[11] Cox,[12] and Hargreaves *et al.*[13,14] for evaluating the effectiveness of diffusing surfaces. The primary difference here is that radiated pressures are being measured rather than scattered pressures. Cox and Hargreaves *et al.* argued in favor of using levels in such formulas because they form a better "linear perceptual scale" for listeners and those visually evaluating decibel plots. They also suggested that an energetic spatially averaged level could be used in place of Eq. (3). However, this modification only produced a maximum positive correction of around 0.2 dB in the narrowband standard deviations well above the omnidirectional cutoff frequencies. Since the RPLs radiated fields with "moderate to good" uniformity at all frequencies of interest, the arithmetic spatial average in Eq. (3) was considered adequate to rank their performances.

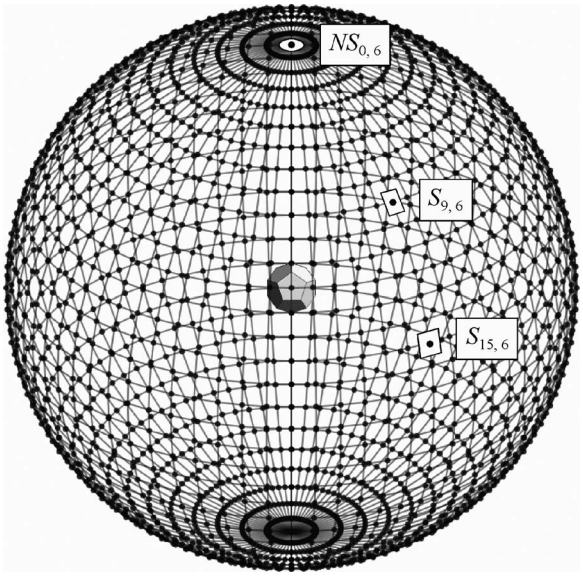


FIG. 7. A representation of sampling points (microphone positions) on the measurement sphere. The effective sampling area for each point differs, as suggested by the sampling areas $NS_{0,6}$, $S_{9,6}$, and $S_{15,6}$. The sampling areas of the two poles are segmented into N equal pie slices of areas $S_{0,n}$ and $S_{36,n}$ for convenience in the summations. The total polar sampling areas are then $NS_{0,n}$ and $NS_{36,n}$.

The resulting narrowband standard deviation plots are given in Figs. 8 and 9. They confirm that the RPLs did depart from omnidirectional behavior at cutoff frequencies in the vicinity of 1 kHz, then exhibited varying degrees of directivity above those frequencies. (Rising standard deviation values below 150 Hz are caused by poor measurement coherence and may be ignored.) Table III shows the cutoff frequencies for the RPLs, determined as the frequencies above which the standard deviations consistently exceeded an arbitrary threshold of 0.5 dB (shown by the dotted horizontal lines in Figs. 8 and 9). These frequencies are consistently but not dramatically higher than the conservative estimates given by Tarnow.[3,4] To compensate for the prediction offset and to better address expected cutoff orders, the next two columns of the table show ratios of measured cutoff frequencies for each RPL to the EM/EV tetrahedron cutoff

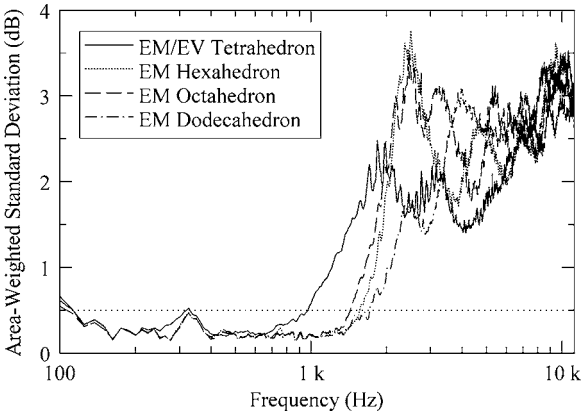


FIG. 8. Narrowband area-weighed spatial standard deviations for the EM RPLs. The omnidirectional cutoff frequency for each RPL is determined as the frequency above which the standard deviation consistently exceeds an arbitrary threshold of 0.5 dB (shown by the horizontal dotted line).

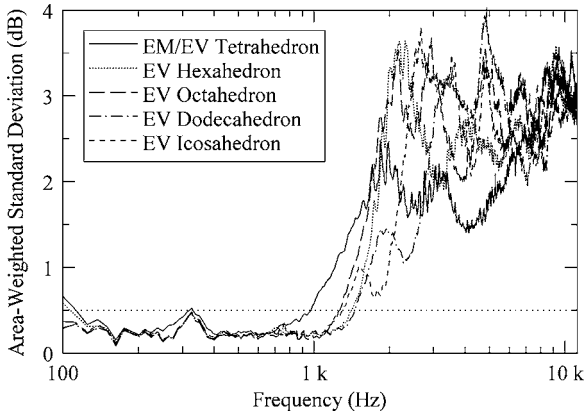


FIG. 9. Narrowband area-weighted spatial standard deviations for the EV RPLs. The omnidirectional cutoff frequency for each RPL is determined as the frequency above which the standard deviation consistently exceeds an arbitrary threshold of 0.5 dB (shown by the horizontal dotted line).

frequency, and similar approximate cutoff frequency ratios based on Tarnow's work. (The latter assume that a constructed RPL midradius can be substituted directly for Tarnow's "mean radius" a.) In the next two columns, the table provides the applicable EM/EV cutoff frequency ratios for each RPL shape and a comparative radius ratio with respect to the EM/EV tetrahedron midradius (which is equivalent to all EM RPL midradii). This radius ratio is used in the last column to adjust the measured cutoff frequency ratios from the second column and thus enable rough evaluation of relative EV RPL cutoff frequencies caused more by their shapes than their midradii. The EV and EM values for similar shapes in this column are generally closer than those in the second column. Of course, since the loudspeaker drivers were not scaled for the larger EV RPLs, one would not expect ideal agreement. Other properties may also come into play.

From low frequency to high frequency, the cutoff order for the four EM RPLs was (1) tetrahedron, (2) octahedron, (3) hexahedron, and (4) dodecahedron, almost agreeing with Tarnow's order. Because the cutoff frequency of the octahedron comes before its polyhedron order would suggest, one might initially surmise that its geometry is somehow less desirable than expected. However, comparison of the measured and Tarnow cutoff frequency ratios in Table III suggests that the hexahedron geometry is actually more desirable than expected. Of course, other factors must also be considered. The dodecahedron loudspeaker produced a relative cutoff frequency that was lower than expected. While the tetrahedron loudspeaker produced the lowest of all cutoff frequencies (as expected), many of its higher-frequency standard deviation values were notably lower than those of the other RPLs. In fact, as shown in Fig. 8, it clearly produced the most omnidirectional characteristics in a range extending from approximately 3 to 6 kHz, agreeing with the balloon plot observations mentioned in Sec. III.

The cutoff order for the five EV RPLs was (1) tetrahedron, (2) octahedron, (3) icosahedron, (4) hexahedron, and (5) dodecahedron. In this case, the cutoff orders of both the octahedron and icosahedron came before their polyhedron orders would suggest. However, for this set, the higher-order

TABLE III. Measured cutoff frequencies and related values for the eight RPLs. The cutoff frequencies f_c were defined as those above which the narrowband spatial standard deviations consistently exceeded a threshold of 0.5 dB. Ratios of the various frequencies to that of the EM/EV tetrahedron ($f_{c,t}$) are given, along with similar approximate ratios derived from Tarnow's work (Refs. 3 and 4). Ratios of cutoff frequencies for similar EM and EV RPL shapes are also given as $f_{c,\mathrm{EM}}/f_{c,\mathrm{EV}}$, followed by comparative radius ratios a/a_t and radius-adjusted cutoff frequency ratios $f_c/f_{c,t}\times a/a_t$.

RPL	Measured cutoff frequency f_c (Hz)	Measured cutoff ratio $f_c/f_{c,t}$	Tarnow cutoff ratio $f_c/f_{c,t}$	EM/EV cutoff ratio $f_{c,\mathrm{EM}}/f_{c,\mathrm{EV}}$	Radius Ratio a/a_t	Adjusted measured cutoff ratio $f_c/f_{c,t}\times a/a_t$
EM/EV tetrahedron	975	1.00	1.0	1.00	1.00	1.00
EM hexahedron	1562.5	1.60	1.2	1.11	1.00	1.60
EM octahedron	1437.5	1.47	1.4	1.14	1.00	1.47
EM dodecahedron	1712.5	1.76	2.2	1.17	1.00	1.76
EV hexahedron	1412.5	1.49	1.0	1.11	1.12	1.67
EV octahedron	1262.5	1.29	1.3	1.14	1.12	1.44
EV dodecahedron	1462.5	1.50	1.6	1.17	1.33	2.00
EV icosahedron	1300	1.33	1.5	N/A	1.48	1.97

polyhedra had progressively larger midradii that could have impacted the sequence. Figures 8 and 9 clearly show that the larger EV hexahedron, octahedron, and dodecahedron loudspeakers had consistently lower cutoff frequencies than their smaller EM counterparts. Nevertheless, even after the cutoff frequency ratios were adjusted by radius ratios in the last column of Table III, other geometric features seemed to persist as significant determinants of the order. In fact, with the exception of the newly introduced icosahedron, the adjusted cutoff order was identical to that found for the EM RPLs. Comparison of the measured and Tarnow cutoff frequency ratios in Table III again reveals that the EV hexahedron ratio is higher than expected, while those of the EV dodecahedron and EV icosahedron are lower than expected.

The standard deviation formulation was also applied to the octave-band levels calculated from the pink-noise-weighted frequency response functions discussed in Sec. IV A (without prior sound power equalization). This enabled observation of broad standard deviation trends and a more equitable comparison to the results of the ISO 3382 qualification method. Progressively increasing octave bandwidths again produced sufficient spectral averaging to effectively decrease the spatial variation of the radiated fields at higher frequencies and therefore decrease the standard deviation values. Sound power equalization and pink-noise weighting were again found to produce only small changes to standard deviation curves derived from unfiltered octave-band data.

A careful comparison of the octave-band standard deviation curves revealed both agreement and disagreement with the results of the ISO 3382 method given in Figs. 5 and 6. For the EM RPLs, the tetrahedron was the first to depart from nearly ideal omnidirectional behavior in the 1 kHz octave band. In the 2 kHz octave band, the dodecahedron produced the most uniform radiation, followed by the tetrahedron and hexahedron, then finally the octahedron. In the 4 kHz octave band, the tetrahedron performed best, followed by the dodecahedron, the hexahedron, then the octahedron. In the 8 kHz octave band, the dodecahedron performed best, followed by the octahedron, the hexahedron, then the tetrahedron. These rankings are summarized in Table II for comparison with the rankings from the ISO 3382 method. Overall, the octave-band standard deviation curves suggested that the dodecahedron maintained the most uniform radiation while the octahedron maintained the worst. This observation agrees with the mean standard deviation results discussed in Sec. IV C 1 and given in Table IV.

For the EV RPLs, the tetrahedron was again the first to depart from its omnidirectional behavior. Rankings for higher octave bands are summarized in Table II. Overall, it appeared that the icosahedron maintained the best radiation uniformity over the octave bands while the octahedron maintained the worst. However, the performance ranking for the icosahedron does not agree well with the mean standard deviation ranking given in Table IV.

Some of the most prominent points of disagreement between the ISO curves in Figs. 5 and 6, and the octave-band standard deviation curves can be found through inspection of

TABLE IV. Frequency-averaged standard deviation for the eight RPLs. This figure of merit was calculated by taking the arithmetic average of the narrowband spatial standard deviation values through the 4 and 8 kHz octave bands. Rankings for the lowest average values are given separately for the EM and EV RPLs in each range.

	100 Hz to 5.6 kHz (through 4 kHz octave)		100 Hz to 11.2 kHz (through 8 kHz octave)	
RPL	$\langle\sigma(f)\rangle_f$ (dB)	Rank	$\langle\sigma(f)\rangle_f$ (dB)	Rank
EM/EV tetrahedron	1.48	1	2.17	2
EM hexahedron	1.75	3	2.27	3
EM octahedron	1.89	4	2.37	4
EM dodecahedron	1.59	2	2.16	1
EM average	1.68	N/A	2.24	N/A
EM/EV tetrahedron	1.48	1	2.17	1
EM hexahedron	1.88	2	2.34	3
EV octahedron	2.12	5	2.53	5
EV dodecahedron	1.95	3	2.32	2
EV icosahedron	1.98	4	2.35	4
EV average	1.88	N/A	2.34	N/A

Table II. They include dramatically different orders of omnidirectional performance rankings for the two octahedrons in the 4 kHz octave band and the two dodecahedrons in the 8 kHz octave band. They also include opposite overall performance rankings for the EM dodecahedron. Furthermore, only the standard deviation method confirms the visual observation that the tetrahedron displays the greatest uniformity in the 4 kHz octave band. Other less significant discrepancies become apparent through closer inspection of the table and figures.

C. Additional statistical properties

Even with the useful data reduction provided by frequency-dependent line graphs and octave-band analysis, additional reduction was desirable to more readily compare the omnidirectionality of the sources. While no assumption was made about normal distributions, basic statistical properties of the frequency-dependent spatial standard deviation functions could be useful for this purpose. These properties might even be used to produce single-number figures of merit.

1. Mean value

The frequency-averaged standard deviation $\langle \sigma(f) \rangle_f$ was the first property considered as a figure of merit. As shown in Table IV, it was calculated for each RPL as an average of the frequency-dependent standard deviation functions given by Eq. (2) for all narrowband frequencies within two averaging bandwidths. The first bandwidth was from 100 Hz to 5.6 kHz (i.e., through the 4 kHz octave band) and the second was from 100 Hz to 11.2 kHz (i.e., through the 8 kHz octave band). Lower values were expected to correspond more closely (on average) to ideal omnidirectional radiation patterns over the bandwidths. According to this approach, the EM/EV tetrahedron was found to be the most omnidirectional source over the 4 kHz bandwidth—even though it had the lowest cutoff frequency. The EM dodecahedron loudspeaker ran a close second in this bandwidth and performed best over the 8 kHz bandwidth. The EM/EV tetrahedron performed second best over the 8 kHz bandwidth and was consistently better than any EV RPL over either bandwidth. The EM hexahedron, octahedron, and dodecahedron loudspeakers performed consistently better than their larger EV counterparts. Unfortunately, it was found (as suggested by Cox in relationship to diffusing surfaces[15]) that a simple average of the standard deviations over a measurement bandwidth allowed good omnidirectional spectral regions to unduly compensate for poor spectral regions. A simple average also failed to represent other undesirable swings in directional behavior.

2. Notched box-whisker diagrams

A more thorough approach to the problem involves the creation and observation of notched box-whisker diagrams. These provide an excellent overview of directional behavior over frequency, without necessarily producing a single-number figure of merit. Such diagrams are typically used to summarize several statistical properties of data sets, enabling an observer to assess whether their distributions are skewed, or if outliers or irregular data values are present. In this case, they are used to summarize the distribution and grouping of the frequency-dependent spatial standard deviation values. Figures 10 and 11 show the diagrams for each EM RPL through the 4 and 8 kHz octave bands, respectively.

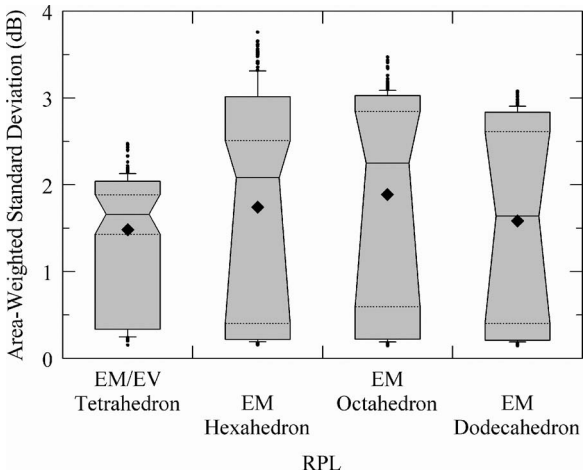


FIG. 10. Notched box-whisker diagrams showing the distribution and grouping of frequency-dependent spatial standard deviation values for the EM RPLs from 100 Hz to 5.6 kHz (through the 4 kHz octave band).

For a given RPL and measurement bandwidth, an individual notched box-whisker diagram is interpreted as follows. The median, or 50th percentile, is represented by a solid horizontal line at the narrowest point of the notched region. This line divides the diagram into upper and lower portions. The 25th and 75th percentiles are marked by dotted horizontal lines at the bottom and top extremes of the notched regions, respectively, thus completing trapezoidal areas. The rectangular areas below and above these lines represent the data between the 10th and 25th percentile, and the 75th and 90th percentile, respectively. The lower "whisker," below the lower rectangular area, represents the data between the 5th percentile and 10th percentile. Data samples shown by dots below the whisker represent the outlying 5%, including the minimum. The upper whisker, above the upper

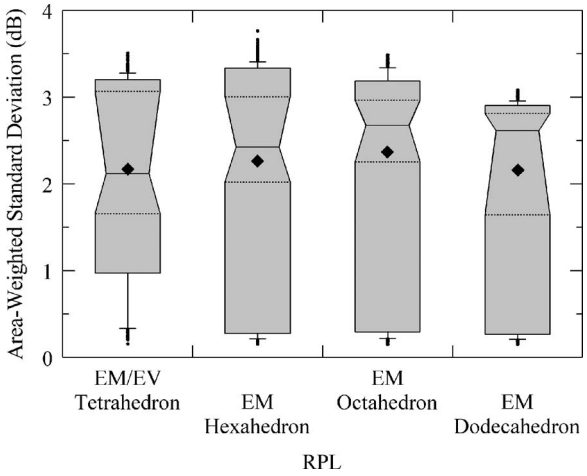


FIG. 11. Notched box-whisker diagrams showing the distribution and grouping of frequency-dependent spatial standard deviation values for the EM RPLs from 100 Hz to 11.2 kHz (through the 8 kHz octave band).

rectangular area, represents the data between the 90th percentile and 95th percentile. Data samples beyond this whisker represent the outlying 5%, including the maximum. The mean value from Table IV is indicated in the diagram by a closed black diamond.

As a general rule, lower values of the diagram indicators suggest superior omnidirectional behavior. For example, lower maximum, minimum, first quartile, third quartile, median, and mean values all suggest better performance, but in different ways. A lower maximum value indicates lower maximum directivity at any frequency. As discussed earlier, a lower mean value may suggest that a RPL is more omnidirectional, but one must still observe the spread in data values and other statistical properties to form an adequate judgment of its actual omnidirectional performance. A lower mean with a large box or whisker spread could clearly be less desirable than a slightly higher mean with a very small spread. A smaller notched box indicates greater uniformity in directivity over much of the bandwidth, but does not necessarily suggest acceptable omnidirectionality. A source that is perfectly omnidirectional at all frequencies would consistently show zero box and whisker spread, no outliers, and zero median and mean values.

By comparing the box-whisker diagrams in Fig. 10, one would likely conclude that the EM/EV tetrahedron is the most omnidirectional EM RPL over the 4 kHz bandwidth, since it produces the lowest maximum, median, and mean values, as well as the smallest box spread. However, one also finds that its first quartile is higher than that of any other EM RPL. The EM dodecahedron would likely be chosen as the second-best performing RPL. Depending upon the selection criteria, either the EM hexahedron or EM octahedron would probably be ranked as the worst performing over the bandwidth. The octahedron produces the highest median and mean values, but the hexahedron produces the highest maximum value and largest upper whisker spread.

Figure 11 demonstrates that the EM dodecahedron produces the lowest mean, upper quartile, and maximum values over the 8 kHz bandwidth, but the EM/EV tetrahedron again produces the lowest median and smallest box spread. Nevertheless, because the 10th and 75th percentile values of the tetrahedron are relatively high, one would likely choose the dodecahedron as the best source over this bandwidth and the tetrahedron as the second best. Once again, either the EM hexahedron or octahedron would likely be chosen as the worst-performing RPL, depending upon the selection criteria.

Using a similar approach for the EV RPLs, the EM/EV tetrahedron would be chosen as the best-performing EV RPL over both the 4 and 8 kHz bandwidths. Over the 4 kHz bandwidth, one would likely choose the EV hexahedron as the second-best performer and either the EV octahedron or EV dodecahedron as the worst performer. The EV icosahedron would probably be chosen as the third-best performer. Over the 8 kHz bandwidth, either the hexahedron or icosahedron could be chosen as the second-best performer, most likely followed by the dodecahedron, then the octahedron.

3. Additional metrics

Other approaches to the data reduction problem might involve the computation of "omnidirectionality coefficients." While such work is beyond the scope of the present paper, a few possibilities are worth mentioning.

A frequency-dependent spatial autocorrelation of radiated energy is one intriguing example that stems from recent analysis of diffusing surfaces[13,14,16,17] and the analogy already mentioned between these surfaces and omnidirectional sources. A single-number figure of merit might also be derived for RPLs in a fashion similar to that suggested by D'Antonio and Cox[15,18] (also see Ref. 13). This would involve computation of the spectral standard deviation of the frequency-dependent spatial standard deviation to show the degree of variation about the spatial standard deviation mean. This single-number standard deviation would then be added to the mean to form the figure of merit.

As in other areas of acoustics, one must exercise caution in producing or using single-number or multiple-number figures of merit if the data reduction robs the data of its interpretive value. A compromise between excessive detail and excessive reduction must be made to ensure that necessary detail is not sacrificed for the sake of simplicity alone.

V. SUMMARY AND CONCLUSIONS

While dodecahedron loudspeakers are currently the most widely used "omnidirectional" sources for room acoustics measurements, they should not necessarily be viewed as more omnidirectional than other RPLs. This is especially true when their directivities are considered over complete measurement bandwidths. The omnidirectional cutoff frequency of a dodecahedron loudspeaker may be higher than that of another RPL with an equal midradius, but all RPLs should be recognized as "multidirectional" sources with varying degrees of directivity above their cutoff frequencies. Since these frequencies often occur near 1 kHz (well within common measurement bandwidths), understanding and comparison of higher-frequency directional behaviors becomes crucial.

Area-weighted spatial standard deviations of 2664 frequency response function levels over a free-field measurement sphere provided more information about omnidirectional performance than gliding-average directional deviations that are implemented using single-plane measurement arcs. The standard deviation method is supported by visual balloon plot observations. It provides an effective global view of source radiation using either narrowband or proportional-band analysis.

This method revealed that a constructed tetrahedron loudspeaker produced better average radiation uniformity than a dodecahedron loudspeaker (or any other constructed RPL) over specified measurement bandwidths. This was true even though its omnidirectional cutoff frequency was the lowest. While no RPL stood out as being consistently exceptional above its cutoff frequency, the tetrahedron was found to produce the most uniform radiation in the 4 kHz octave band. The dodecahedron was also a good choice over the measurement bandwidths, but not simply because it pro-

duced the highest omnidirectional cutoff frequency. The hexahedron produced a cutoff frequency above that of the octahedron, suggesting that its cutoff frequency was better than expected (based on its polyhedron order). The standard deviation method also confirmed that RPLs with smaller midradii consistently produced higher cutoff frequencies than similarly shaped RPLs with larger midradii and volumes. However, attempts to decrease cutoff frequencies through smaller enclosure volumes should be carefully weighed so as to avoid degradation of sound power output, low-frequency output, and spectral uniformity.

Because it was difficult to rank overall omnidirectional performance from the fluctuating frequency-dependent standard deviation curves, notched box-whisker diagrams were introduced as a means of summarizing the directivity information they contained. Octave-band standard deviation curves were also assessed for data reduction and compared to curves from the ISO 3382 source qualification method. The two approaches agreed on some omnidirectional characteristics and rankings, but significantly disagreed on others.

The findings of this paper are based on a limited number of source embodiments. One should therefore use caution in generalizing them to other source geometries, dimensions, loudspeaker driver diameters, etc. Additional theoretical, numerical, and experimental work is needed to further generalize the results and develop needed predictive tools. In particular, greater clarification is required to establish the interconnected roles of individual driver directivities, driver spacings, driver orientations, enclosure diffraction effects, and multiple-driver interference effects in the production of composite RPL directivity patterns. Nevertheless, the findings of this work do provide an experimental basis for future work, while enhancing understanding of RPLs as measurement tools for room acoustics and other applications. They emphasize that the omnidirectional quality of a source should not be judged exclusively by its omnidirectional cutoff frequency or by the satisfaction of relaxed omnidirectional requirements at higher frequencies of interest. They motivate the possibility of using lower-order polyhedra than the dodecahedron with fewer drivers and lower cost as equally viable sources for current standards. While the sound power output of these sources would be smaller for a given type of loudspeaker driver, alternate drivers could be selected as a means of compensation.

The work has demonstrated a viable alternative for assessing source omnidirectionality that is complementary or preferable to current methods. As a result, standardized omnidirectional source qualification methods may benefit from pertinent review and modification. In the future, we recommend that they be approached in a fashion that encourages omnidirectional source designers to significantly improve source omnidirectionality at higher frequencies of interest, rather than merely satisfy the status quo of relaxed requirements. One aim for designers would be to decrease the spatial standard deviation of radiated fields above cutoff frequencies. However, a further-reaching aim would be to significantly increase omnidirectional cutoff frequencies so that sources maintain nearly ideal omnidirectional radiation (as opposed to multidirectional radiation) throughout common measurement bandwidths.

An improvement to the measurement methods described in this paper would involve an increase in the spatial resolution of frequency response measurements over an entire measurement sphere, so that lobing structures at frequencies within the 8 kHz octave band (or higher bands) can be better resolved. While manufacturers should be willing and able to provide such comprehensive high-resolution testing of omnidirectional sources, they and others may also be interested in simplified measurement schemes for occasional rapid testing. To this end, we recommend that the effectiveness of the standard deviation method be investigated for single-plane measurement arcs in addition to complete measurement spheres.

An intriguing aspect of the standard deviation method is its close parallel to established methods of assessing scattering uniformity from diffusing surfaces. The analogies should be further investigated. Related metrics such as "omnidirectionality coefficients" could be developed to better characterize omnidirectional sources for comparative purposes. We encourage research in these areas to improve the evaluation and development of better omnidirectional sources for use in acoustical applications.

ACKNOWLEDGMENTS

This work was funded in part by the Research Experiences for Undergraduates (REU) program of the National Science Foundation (NSF). The authors also acknowledge Gordon Dix, Kent Gee, Todd Kitchen, Wesley Lifferth, and Jacob Robinson for their assistance in the construction of the experimental RPLs and measurement apparatuses, and in the development of numerical analysis tools. They further acknowledge William Christensen for his helpful comments on statistical matters and Viggo Tarnow for encouragement and information regarding his early efforts in omnidirectional source characterization.

[1]"Acoustics—measurement of sound insulation in buildings and of building elements—part IV: field measurements of airborne sound insulation between rooms," ISO 140-4:1998(E), International Organization for Standardization, Geneva, Switzerland, 1998.
[2]H. S. M. Coxeter, *Regular Polytopes*, 2nd ed. (Macmillan, New York, 1963).
[3]V. Tarnow, "Computation of sound radiation from loudspeaker systems with the symmetry of the platonic solids," Proceedings of the Eighth International Congress on Acoustics, London, 1974, p. 617.
[4]V. Tarnow, "Sound radiation from loudspeaker system with the symmetry of the platonic solids," Brüel & Kjær Technical Review **4–1974**, 23–31 (1974).
[5]I. B. Witew and G. K. Behler, "Low-mid frequency measurement of single number parameters in room acoustics with multi-way dodecahedron speakers," Proceedings of Internoise 2005, the 2005 Congress and Exposition on Noise Control Engineering, Rio de Janeiro, Brazil, 2005.
[6]"Acoustics—measurement of the reverberation time of rooms with reference to other acoustical parameters," ISO 3382:1997(E), International Organization for Standardization, Geneva, Switzerland, 1997.
[7]*CRC Standard Mathematical Tables and Formulae*, 29th ed., edited by W. H. Beyer (CRC Press, Boca Raton, FL, 1991), p. 110.
[8]H. M. Cundy and A. P. Rollett, *Mathematical Models*, 3rd ed. (Tarquin, Norfolk, UK, 1961), pp. 84–88.
[9]"Specification for octave-band and fractional-octave-band analog and digital filters," ANSI S1.11-1986, American National Standards Institute, Washington, DC, 1986.

[10]M. R. Schroeder, "Binaural dissimilarity and optimum ceilings for concert halls: More lateral sound diffusion," J. Acoust. Soc. Am. **65**, 958–963 (1979).
[11]P. D'Antonio, "Performance evaluation of optimized diffusors," J. Acoust. Soc. Am. **97**, 2937–2941 (1995).
[12]T. J. Cox, "Designing curved diffusers for performance spaces," J. Audio Eng. Soc. **44**, 354–364 (1996).
[13]T. J. Hargreaves, "Acoustic diffusion and scattering coefficients for room surfaces," Ph.D. thesis, University of Salford, Salford, UK, 2000.
[14]T. J. Hargreaves, T. J. Cox, Y. W. Lam, and P. D'Antonio, "Surface diffusion coefficients for room acoustics: Free-field measures," J. Acoust. Soc. Am. **108**, 1710–1720 (2000).
[15]T. J. Cox, "The optimization of profiled diffusers," J. Acoust. Soc. Am. **97**, 2928–2936 (1995).
[16]T. J. Cox and P. D'Antonio, *Acoustic Absorbers and Diffusers: Theory, Design, and Application* (Spon, London, 2004).
[17]"AES information document for room acoustics and sound reinforcement systems—Characterization and measurement of surface scattering uniformity," AES-4id-2001, Audio Engineering Society, New York, 2001.
[18]P. D'Antonio and T. Cox, "Two decades of sound diffusor design and development. 2. Prediction, measurement, and characterization," J. Audio Eng. Soc. **46**, 1075–1091 (1998).

Evaluation of an autoregressive spectral estimator for modal wave number estimation in range-dependent shallow water waveguides

Kyle M. Becker[a)]
Applied Research Laboratory, The Pennsylvania State University, State College, Pennsylvania 16804-0030

George V. Frisk
Department of Ocean Engineering, Florida Atlantic University, Dania Beach, Florida 33004

(Received 25 October 2005; revised 3 May 2006; accepted 14 June 2006)

In shallow water, geoacoustic properties of the seabed can be inferred from knowledge of acoustic normal modes propagating in the waveguide. For range-varying waveguides, modal content varies locally in response to changes in the environment. The problem becomes estimating the local modal content of a propagating field. For acoustic fields measured on horizontal arrays, resolution of closely spaced modal eigenvalues depends on the data aperture length. Therefore, range variability in a waveguide can only be detected for range scales of order of the resolvability of the individual modes. When only short data apertures are available for modal estimation, high-resolution methods must be used. In this paper, a high-resolution autoregressive (AR) spectral estimation method for extracting the modal information from measurements of a continuous wave acoustic field made on a horizontal array is evaluated. Performance is discussed for estimation accuracy and resolution in the presence of noise. Results are compared to previous work characterizing high-resolution wave number estimators. The AR estimator performance is comparable to other high-resolution methods and does not require prior information about the number of propagating modes. In addition, range/wave number plots obtained using a sliding window technique yield a robust method for identifying propagating modes and any changes with range.
 [DOI: 10.1121/1.2221540]

PACS number(s): 43.60.Gk, 43.30.Bp, 43.20.Mv [WMC] Pages: 1423–1434

I. INTRODUCTION

Acoustic propagation in shallow water is greatly influenced by the interaction of the sound field with the bottom. At ranges greater than a few wavelengths, the field generated by a point source can be modeled as a discrete sum of normal modes propagating in the waveguide. For a horizontally stratified medium, the modes are the eigenvalues and eigenfunctions that satisfy the boundary conditions for a depth-dependent Helmholtz equation. Assuming the air-sea interface is pressure release, and that properties of the water column are sufficiently well known, the modes are primarily a function of the bottom boundary conditions. Thus, estimates of the modal wave number content in shallow water can be used as input data to inversion schemes for inferring geoacoustic properties of the seabed.[1,2] Wave numbers can be estimated from continuous wave point-source acoustic pressure field data measured on a synthetic aperture horizontal array created by relative motion between the source and receiver.[3,4] For a single hydrophone at a fixed depth in the water column, this configuration yields measurements of complex pressure as a function of range and is equivalent to using a large horizontal array of receivers. The problem addressed in this paper is that of estimating individual modal eigenvalues from acoustic pressure field data as required for the geoacoustic inverse problem.

For range-independent, axisymmetric shallow water waveguides, a Hankel transform pair governs the forward and inverse acoustic problems[5]

$$p(r;z_s,z) = \int_0^\infty g(k_r;z_s,z) J_0(k_r r) k_r dk_r,$$

$$g(k_r;z_s,z) = \int_0^\infty p(r;z_s,z) J_0(k_r r) r dr, \tag{1}$$

where $p(r;z_s,z)$ is the complex pressure field measured over range, r, and $g(k_r;z_s,z)$ is the depth-dependent Green's function for horizontal wave numbers k_r. In Eq. (1), the source and receiver depths, z_s and z, are treated as parameters in the problem and J_0 is the zeroth-order Bessel function. In shallow water, the Green's function is characterized by a finite number of sharp peaks occurring at horizontal wave number values corresponding to the eigenvalues of the propagating modes. Thus, to obtain estimates of modal eigenvalues for the geoacoustic inverse problem, the Green's function is estimated by a Hankel transform of the measured acoustic field. The locations of the peaks in the Green's function yield the modal eigenvalue estimates.

In practice, for increased computational efficiency, the Hankel transform is often evaluated by converting to a two-sided integral representation and using a large-argument

[a)]Electronic mail: kmbecker@psu.edu

approximation.[5,6] The Green's function is then approximated by the horizontal wave-number spectrum resulting from an inverse Fourier transform

$$g(k_r) \sim \frac{e^{i\pi/4}}{\sqrt{2\pi k_r}} \int_{-\infty}^{\infty} p(r)\sqrt{r}e^{-ik_r r}dr, \quad k_r r \gg 1, \quad k_r > 0, \tag{2}$$

where the depth parameters have been suppressed. Using this form, the horizontal wave-number spectrum, which approximates the Green's function, can be obtained using standard spectral estimation techniques applied to the product $p(r) \times \sqrt{r}$. As done in previous work,[6] the terms "Green's function" and "horizontal wave-number spectrum" are considered to be synonymous in the remainder of this paper.

Based on Eq. (2), computationally efficient spectral estimates can be obtained using classical methods employing the fast Fourier transform (FFT). However, the use of FFT-based methods to determine wave number content is often limited due to windowing and/or finite aperture length considerations. The ability to resolve individual wave numbers is inversely proportional to aperture size, i.e., resolution of closely spaced wave numbers requires larger apertures. Also, the use of window functions to truncate data sequences degrades wave number estimates through contamination of weak spectral components by sidelobes of strong spectral peaks. These limitations become more restrictive as frequency and/or waveguide depth increase which cause the lower order modal eigenvalues to group closer together. These problems are compounded when waveguide properties are range dependent, or only a limited data aperture is available for analysis.

Over the years, a number of methods have been developed to extract modal content for range-dependent environments. Kessel[7] used bandpass filtering in the wave number domain to separate out individual modes and applied a phase demodulation technique to examine changes in wave number values as a function of range. Ohta and Frisk[2] developed a technique analogous to short-time Fourier transform methods for extracting modal content with range. Both methods were developed as a means to extract modal eigenvalues at intervals along a waveguide in order to perform local geoacoustic inversions. The methods both require that there be sufficient separation between modes in the wave-number domain or, equivalently, that large range-independent regions of the waveguide exist, to resolve individual modes with range. In this work, the interest was to apply the sliding-window Hankel transform approach employed by Ohta and Frisk,[2] based on Eq. (2), that assumes the waveguide can be segmented into locally range-independent regions. Therefore, in order to obtain the best possible localized wave number estimates, it is desirable to minimize the required data aperture length. One way to achieve this is through the application of high-resolution frequency estimation methods.

The evaluation of high-resolution frequency estimators for wave number estimation has been considered in the past. An approach based on the energy spectral density function computed from a deterministic exponential model fit to measurement data, or Prony's method, was evaluated by Diemer and found to perform poorly on real data.[8] The signal subspace algorithms, multiple signal classification (MUSIC) and estimation of signal parameters via rotational invariance (ESPRIT) were evaluated by Rajan and Bhatta.[9] In their work, *a priori* knowledge of the number of wave numbers in the signal was assumed in the evaluation. However, application of these methods to real data is limited in that the number of propagating modes may not be known.

In this paper a high-resolution autoregressive (AR) spectral estimator is evaluated for use in wave number estimation. The AR estimator does not require *a priori* information and is applied to both time series and range data for extracting frequency and wave number content. Resolution of closely spaced frequencies/wave numbers is investigated for various combinations of frequency separations, signal-to-noise ratios (SNR), and aperture lengths. For the various combinations of parameters and apertures considered, results are presented as a function of the dimensionless product of parameter separation and data aperture length referred to here as *estimation length* (L_{es}). The use of estimation length allows a more direct comparison of estimator performance whether applied to frequency or wave number estimation. For frequency estimation $L_{es}=\Delta f T$, where Δf is the minimum frequency separation between tones of a multitone signal, and $T=N/f_s$ is the total time aperture determined by the number of data samples, N, and the sampling frequency f_s. For wave number estimation $L_{es}=\Delta k_r R$, where Δk_r is the separation between the first two horizontal wave numbers corresponding to modes 1 and 2, and $R=N\Delta r$ is the total range aperture of the data determined by the number of samples and the range sampling interval Δr.

The remainder of the paper is organized as follows. A description of wave number estimation based on an autoregressive spectral estimator is provided in Sec. II. In Sec. III, statistical properties of the estimator are determined numerically for the estimation of individual frequencies and wave numbers. Performance of the estimator is determined by its ability to both resolve and accurately determine individual values of closely spaced frequencies/wave numbers. Parameter values are extracted from signals having various frequency/wave number separations, increasing amounts of additive noise, and for varying estimation lengths. Comparisons are made with the Cramér-Rao Bound (CRB) as well as estimates from the signal subspace algorithms MUSIC and ESPRIT applied to the same input signals. In Sec. IV, the AR estimator is applied to range-dependent acoustic field data. A summary of estimator performance is given along with some criteria for application in Sec. V.

II. FREQUENCY AND WAVE NUMBER ESTIMATION

The spectral representation of a point source acoustic field in shallow water has a finite number of sharp peaks. These peaks occur at horizontal wave numbers corresponding to the trapped modes in the waveguide. As such, the wave number estimation problem described is intimately related to the problem of frequency estimation from a finite number of noisy discrete-time signal measurements. Conse-

quently, for this work, a spectral estimator approach, typically associated with estimating functions of frequency, is adopted for use as a parameter estimator, where the parameters are the horizontal wave numbers. In this context, an estimator is desired for identifying and localizing peaks in the depth-dependent Green's function using short-aperture data. Based on these criteria, an AR spectral estimator was chosen for its high-resolution frequency estimation property.[10]

In this paper, the work of Rajan and Bhatta[9] is expanded to examine AR estimator performance. Estimates obtained using the AR method are compared with those obtained using both MUSIC and ESPRIT. The MUSIC and ESPRIT approaches are well documented in the literature. In addition to the overview in the paper of Rajan and Bhatta,[9] ESPRIT is described by Roy *et al.*,[11] and the algorithms adapted for use in this work are discussed in Stoica,[12] where the *root*MUSIC form was implemented.

For consistency with the previous work, in the numerical studies that follow, estimator performance is evaluated by using the same signal models as Rajan and Bhatta.[9] However, results are presented with respect to the estimation length defined previously rather than by number of samples and/or frequency separation. This work examines estimator performance over a greater range of estimation lengths than the reference paper. The MUSIC and ESPRIT results presented are from simulations run by the authors together with the AR algorithm. Results using the MUSIC and ESPRIT algorithms implemented for this work are consistent with the reference paper over the range of common estimation lengths.

Finally, in contrast to the *root*MUSIC and ESPRIT estimators, which return discrete values for frequency or wave number estimates, the AR algorithm returns a spectral function from which frequencies/wave numbers are estimated by determining the peak locations of the spectrum. In the case of very short apertures and/or very noisy signals, picking the peak locations from the spectrum becomes more difficult due to spurious peaks and/or peak broadening and overlap. Consequently, determining the peak locations for the AR estimator can be highly dependent on the peak picking algorithm used. In this work, an automated procedure was implemented for peak picking based on finding values above a threshold 30 dB above the mean normalized spectrum level. Because of the very narrow peaks of the high-resolution spectral estimates, the threshold was effectively 30 dB above the spectrum noise floor. The resulting peak locations were then sorted lowest to highest in order to yield the frequency or wave number estimates. This approach was greater than 85% effective at determining the proper peak locations. Failure occurred at very short apertures when only one peak was detected, or for high SNR when more peaks than frequencies were detected and the correct frequencies were in the incorrect bin after being sorted. In the second case, incorrect frequency estimates were classified as spurious and eliminated from the statistical analysis of estimator performance. Justification for doing so is provided later in the paper.

Before proceeding, it should be noted that for this work emphasis was placed on extracting the locations of the modal peaks and not on estimating their amplitudes. This distinction was made in recognition that variances in peak levels for AR spectral estimates can be large.[13] Thus, in comparing spectral estimates made using the AR estimator with estimates made using other methods, such as classical methods, observed differences in spectral amplitudes for resolved modal eigenvalues were ignored. Further, this property of the AR estimator makes it unsuitable for estimating compressional wave attenuation from modal amplitude estimates.[14]

Autoregressive spectral estimators

Autoregressive spectral estimators are based on approximating a discrete-time process by a rational transfer function model for the data. They are based on an all-pole model for a linear system where the data are modeled as the output of the system driven by white noise.[10] The all-pole nature of the model is appropriate for describing spectra of signals having large peaks at particular frequencies, or at particular wave numbers, as in the case of the depth-dependent Green's function for shallow water.

For an autoregressive process, the output sequence $\{x[n]\}$ is a linear regression on itself along with a random input or driving process $\{w[n]\}$. Using the notation of Marple[10]

$$x[n] = -\sum_{m=1}^{p} a[m]x[n-m] + w[n], \tag{3}$$

where p is the model order of the AR process used to represent the data. Assuming the input driving process is a zero-mean white noise sequence with variance σ^2, the power spectral density P_{AR} is expressed in terms of the AR model coefficients $a[k]$ as

$$P_{\mathrm{AR}}(f) = \frac{\sigma^2}{|1 + \Sigma_{m=1}^{p} a[m]\exp[-i2\pi fm]|^2}, \tag{4}$$

where f is frequency and a unity sampling interval has been assumed. Equivalently, P_{AR} can be expressed by the discrete Fourier transform

$$P_{\mathrm{AR}}(f) = \sum_{m=-\infty}^{\infty} r_{xx}[m]\exp[-i2\pi fm], \tag{5}$$

where $r_{xx}[m]$ is the autocorrelation sequence

$$r_{xx}[m] = E\{x[n]x^*[n-m]\}. \tag{6}$$

P_{AR} is obtained by first estimating the coefficients $a[1], \ldots, a[p]$ and σ^2. The spectrum is then obtained by evaluating the denominator of (4) by applying a FFT to the estimated AR coefficients, where the frequency sampling interval is determined by the number of FFT points specified. Zero padding the signal by increasing the number of FFT points does not affect the resolution of the estimator, but can impact its performance. In order to get meaningful performance statistics, the number of points must be selected such that the resulting frequency bin widths are less than the inherent resolution of the estimator.

There are several ways of estimating the AR parameters presented in the literature, including the Yule-Walker method, the Burg method, the covariance method, and the

modified covariance method. Each of these is discussed in some detail in the review article by Kay and Marple,[15] as well as in the respective texts by the individual authors.[10,16] Of these methods, the modified covariance approach was selected based on its potential for yielding high-resolution estimates with the least amount of bias and no spectral line splitting.[10]

III. EVALUATION OF THE AR ESTIMATOR

In order to judge the performance characteristics of the AR algorithm to estimate modal wave numbers, several tests were first performed on time series data for a signal plus noise. These tests were adopted from the paper of Rajan and Bhatta[9] for comparison with their results. The time domain signals were chosen with frequency and time-sampling intervals yielding estimation lengths that are typical of wave number differences and range-sampling intervals for acoustic measurements made in shallow-water waveguides. The signal data for the numerical studies were generated using

$$y[n] = \sum_{i=1}^{L} A_i \exp[i2\pi f_i n/f_s] + w[n], \quad n = 0, \ldots, N-1, \tag{7}$$

where the A_i are amplitudes, f_i are frequencies, f_s is the sampling rate in samples/second, and $w[n]$ is complex white Gaussian noise. For all studies, the amplitudes were equal with a value of 1, and the sample rate was $f_s = 1$ sample/s.

Using this signal model, estimator performance was evaluated with respect to the resolution of closely space frequencies or wave numbers along with the variance of the estimates for signals with additive noise. Resolution of single frequencies was compared with the theoretical resolution of AR based frequency estimators given by[10]

$$\delta f = \frac{1.03}{Tp[\mathrm{SNR}_{\mathrm{lin}}(p+1)]^{0.31}}, \tag{8}$$

where δf is the frequency resolution, T is the data time aperture, p is AR model order, and $\mathrm{SNR}_{\mathrm{lin}}$ is the signal-to-noise ratio in linear units.

Algorithm performance was evaluated by determining the mean square error (mse) for individual frequency or wave number estimates[10]

$$\mathrm{mse}[\tilde{\alpha}] = E\{|\alpha - \tilde{\alpha}|^2\} = \mathrm{var}[\tilde{\alpha}] + |B(\tilde{\alpha})|^2, \tag{9}$$

where α is the true parameter value, $\tilde{\alpha}$ is the estimate, $E\{*\}$ is the expected value operator, and B is the bias.

A. Frequency estimation for multitone signals

1. Frequency estimation error

Estimation errors for estimating both single and multiple-tone components in noisy signals were treated by Rife and Boorstyn[17,18] where they derived the Cramér-Rao bound. For an unbiased estimator, the CRB is the best estimate that can be obtained for a particular parameter given the available data. In other words, the CRB gives an estimate of the minimum variance obtainable for estimating a parameter from noisy data. For their work, and the discussion that follows, the noise was assumed to be independent Gaussian white noise with zero mean and variance σ^2.

Using the signal model of Eq. (7) with $L=1$, the CRB for estimating a single frequency is[17]

$$\mathrm{mse}[\tilde{f}] \geqslant \mathrm{var}[\tilde{f}] = \frac{1.5\sigma^2 f_s^2}{A^2 N(N-1)(2N-1)}, \tag{10}$$

where N is the number of samples and the initial sample index starts at zero. The bounds for parameter estimates of multitone signals are more complicated. When multiple frequencies are present, in addition to SNR, the CRB depends on the difference frequencies. For difference frequencies less than $2f_s/N$, constructive and destructive interference occurs and the CRB increases relative to the single-tone case. For difference frequencies greater than $2f_s/N$, Eq. (10) holds.

In this work, the CRB was determined numerically for the signal model given by Eq. (7), which can be expressed as

$$\mathbf{Y} = \mathbf{X} + \mathbf{W}, \tag{11}$$

where $\mathbf{X}$ is the signal sample vector and $\mathbf{W}$ the noise sample vector, each of length N. Assuming independent noise samples with variance σ^2, the noise correlation matrix is given by $\mathbf{R}_w = \sigma^{-2}\mathbf{I}$, with $\mathbf{I}$ an L by L identity matrix, where L is the number of frequencies. The probability density function, $\wp$, of $\mathbf{Y}$ given $\mathbf{X}$ is

$$\wp(\mathbf{Y}/\mathbf{X}) = \frac{|\mathbf{R}_w|^{1/2}}{\sqrt{2\pi}} \exp\left[\frac{1}{2}(\mathbf{Y}-\mathbf{X})^T \mathbf{R}_w (\mathbf{Y}-\mathbf{X})\right]. \tag{12}$$

Referring to (7), $\mathbf{X}$, and, hence, $\mathbf{Y}$, is a function of the parameters A_i and f_i. The CRB for each parameter is determined from the diagonal elements of the inverse of the Fisher information matrix, $\mathbf{J}$. For the parameter pairs α_a and α_b, the elements of $\mathbf{J}$ are given by,

$$J_{ab} = -E\left\{\frac{\partial^2}{\partial\alpha_a \partial\alpha_b} \log \wp\right\}. \tag{13}$$

Inverting $\mathbf{J}$ gives the bounds as

$$\mathrm{var}[\tilde{\alpha}_a - \alpha_a] \geqslant J^{aa}, \tag{14}$$

where J^{aa} is the ath diagonal element of $\mathbf{J}^{-1}$.

It can be shown that the bounds are the same for individual frequency estimates obtained from signals comprised of equal amplitude complex tones with equally spaced samples and independent noise. For this work, using Eqs. (33)–(42) in the paper by Rife and Boorstyn,[18] $\mathbf{J}$ was determined numerically using (13) and inverted to get the CRB. Calculated this way, the CRB was used as a basis of comparison for estimating single frequencies using the AR estimator as well as the MUSIC and ESPRIT methods.

For the two-frequency case, the two term summation of (7) yields a spectrum with two peaks. Estimates were made for various noise levels as indicated by the signal-to-noise ratio. SNR for this study was defined as $10\log_{10}(\sigma_s^2/\sigma_n^2)$, where σ_s^2 and σ_n^2 are the mean-square signal power and noise power, respectively. For most test cases that follow, 500 realizations of the additive noise were generated to determine performance statistics. AR model order was determined by

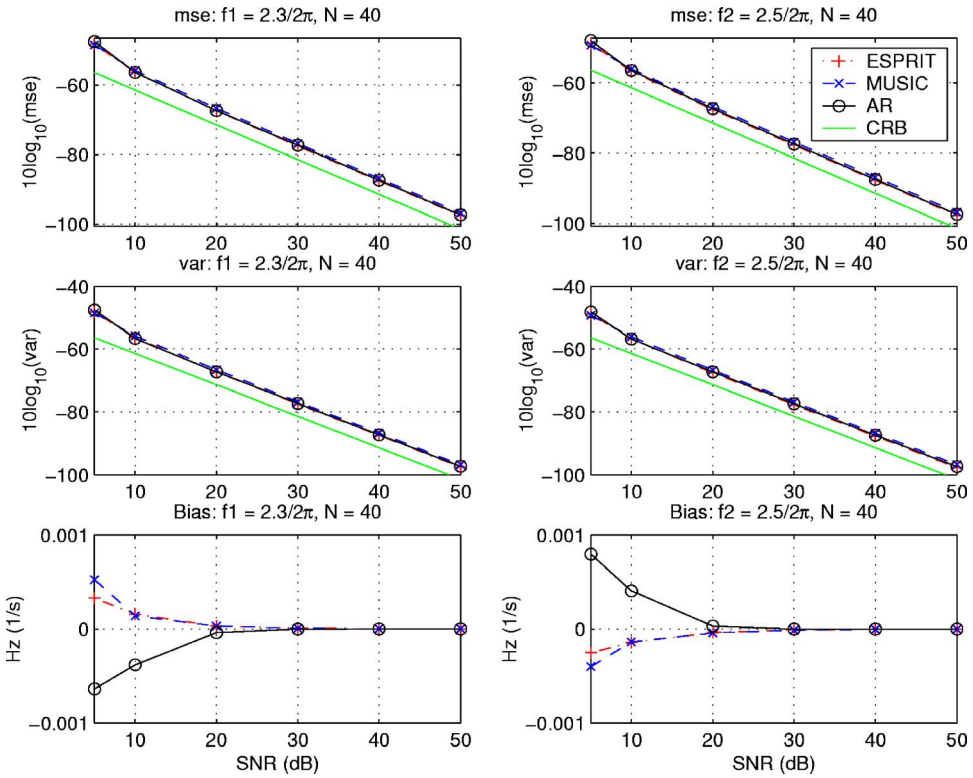


FIG. 1. (Color online) Frequency estimation error and bias vs SNR for $f_1=2.3/2\pi$ and $f_2=2.5/2\pi$ and $L_{es}=1.27$.

trial and error at each SNR as the maximum order that gave only two peaks for all realizations of the signal. Although there are several methods for determining the appropriate model order for the AR estimator, selecting a model order between $N/3$ and $N/2$ usually yields acceptable results.[10] This empirical result was observed in numerical trials reported by Becker.[19] For a 40 sample signal and SNR <30 dB, a model order of 13 was used which was approximately equal to $N/3$. The mean square error of Eq. (9) was used as a basis for performance and compared to the CRB as described earlier. Results are given in decibels (dB) calculated as $10\log_{10}(\text{mse})$. In addition to the CRB, the AR results are compared with estimates obtained using MUSIC and ESPRIT algorithms.

2. Frequency estimation error with increasing SNR

Frequency estimates were made for individual frequencies $f_1=2.3/2\pi$ Hz and $f_2=2.5/2\pi$ Hz for increasing SNR from 5 to 50 dB. The number of data samples was 40 for a constant estimation length $L_{es}=1.27$. 500 realizations of the noise were made at each SNR and the AR, MUSIC, and ESPRIT estimation algorithms applied to each. The mse for estimating the two frequencies are plotted in Fig. 1 with numerical values given in Table I. For this case, where $\delta f=f_2-f_1=0.1/\pi\leqslant 2f_s/N=2/40$, and for most other cases in this study, the CRB was determined numerically. In addition to mse, the variance and bias were determined separately in order to gauge their individual contributions to the error. Variance and bias estimates for each of the two frequencies at the different SNRs are given in Tables II and III, where variance results are presented as $10\log_{10}(\sigma^2)$ for easy comparison with the CRB. From the results, estimator performance can be summarized by the following:

(1) For SNR >5 dB and an estimation length of 1.27, ESPRIT obtained the lowest mse. MUSIC obtained the lowest mse at a SNR of 5 dB.
(2) For SNR $\geqslant 10$ dB and an estimation length of 1.27, the mse of the AR estimates were lower than that of MUSIC and ranged within 0.01 to 0.38 dB of ESPRIT.

TABLE I. SNR (dB) vs $10\log_{10}(\text{mse})$ for estimation of frequencies $f1=2.3/2\pi$ Hz and $f2=2.5/2\pi$ Hz for estimation length $L_{es}=1.27$.

	MUSIC		ESPRIT		AR		
SNR	$f1$	$f2$	$f1$	$f2$	$f1$	$f2$	C-R
5	−48.47	−49.26	−48.20	−48.81	−47.51	−48.04	−56.43
10	−55.82	−56.10	−56.35	−56.74	−56.39	−56.51	−61.43
20	−66.68	−66.75	−67.39	−67.63	−67.31	−67.34	−71.43
30	−76.78	−76.81	−77.51	−77.72	−77.31	−77.44	−81.43
40	−86.80	−86.82	−87.53	−87.73	−87.35	−87.46	−91.43
50	−96.80	−96.81	−97.53	−97.73	−97.35	−97.43	−101.43

TABLE II. SNR (dB) vs $10\log_{10}(\sigma^2)$ of estimates for frequencies $f1=2.3/2\pi$ Hz and $f2=2.5/2\pi$ Hz with $L_{es}=1.27$.

	MUSIC		ESPRIT		AR	
SNR	$f1$	$f2$	$f1$	$f2$	$f1$	$f2$
5	−48.55	−49.31	−48.22	−48.82	−47.61	−48.21
10	−55.84	−56.12	−56.39	−56.78	−56.68	−56.83
20	−66.69	−66.78	−67.40	−67.65	−67.34	−67.35
30	−76.79	−76.84	−77.52	−77.74	−77.30	−77.44
40	−86.80	−86.84	−87.53	−87.75	−87.34	−87.46
50	−96.80	−96.84	−97.54	−97.75	−97.35	−97.43

(3) The higher mse of the AR estimates at very low SNR can be attributed to a large variance resulting from an inability to resolve individual peaks in the spectra.
(4) The AR bias error is the same order of magnitude as that of ESPRIT and MUSIC except at very low SNR.

The results presented are consistent with the results for MUSIC and ESPRIT reported previously.[9] Although the estimation results had a slight bias, violating an assumption of the CRB, the bias was small enough to justify comparison with the bound. Further, the bias for SNR > 10 dB was sufficiently small that mse effectively gives a measure of the variance of the estimates. The observed AR estimator variance of around 2×10^{-8} at 30 dB SNR also compared favorably to the variance estimate given by Sakai[20] for an AR frequency estimator which is proportional to $\mathrm{SNR}_{\mathrm{lin}}^{-2}/N$ or 2.5×10^{-8} for $N=40$.

3. Resolution of two close frequencies

Frequency resolution performance of the AR estimator was determined by identifying two closely spaced frequencies in the presence of noise. This scenario is analogous to the case of closely spaced wave numbers for a shallow water waveguide. The signal model of Eq. (7) was used for two frequencies separated by a δf that ranged from 0.009 to 0.1 Hz. The signal length was held fixed at 40 samples giving estimation lengths of $0.36\leqslant L_{es}\leqslant 4$. Using Eq. (8), at 20 dB SNR, a minimum model order of 13 was required to resolve a $\delta f=0.009$ Hz. The frequencies used to generate the complex signal were $f_2=f_0=0.4$ Hz and $f_1=f_0-\delta f$ Hz. 500 realizations of the noise were generated for each δf. Model order was fixed at 13 for SNR < 40 dB. At higher SNR model order was chosen as the maximum order which only yielded two peaks in the spectrum. Results from this analysis are presented for a SNR of 20 dB in Fig. 2. The figures show that as L_{es} increases, the mse approaches the CRB for all estimators. The CRB is constant for an estimation length greater than 1.6 where the frequency difference is sufficiently large that the two frequencies do not influence each other. These results suggest that for a given SNR there is an optimal value for L_{es} where the minimum estimation error is achieved. Finally, the performance of the AR estimator for $L_{es}>0.7$ is better than MUSIC and on par with ESPRIT.

4. Size of data vector and model order selection

For the previous results, AR model order was selected in an *ad hoc* manner based on the known number of frequency components in the signal. However, in addition to being dependent on SNR and aperture, resolution and variance of the AR estimator are also a function of model order.[10,20] Although the all pole model of the AR estimator suggests that model order be selected equal to the number of complex sinusoids in the signal, the presence of noise requires an increase in the number of model parameters used to best fit the data. AR estimator performance was evaluated for the resolution of closely spaced frequencies for fixed frequency content and varying estimation lengths by increasing the data aperture, i.e., increasing the number of samples. Resolution and variance were also examined at a fixed estimation length for increasing model order.

The frequencies, $f_1=0.39$ Hz and $f_2=0.40$ Hz, with SNR = 10 dB were to be resolved for increasing data apertures. Estimation length was increased from 0.4 to 2, and the probability of success in identifying $f_1\pm0.001$ Hz and $f_2\pm0.001$ Hz were determined for 500 realizations of the noisy signal. For SNR = 10 dB and $N=40$, an AR model or-

TABLE III. SNR (dB) vs bias of estimates for frequencies $f1=2.3/2\pi$ Hz and $f2=2.5/2\pi$ Hz for $L_{es}=1.27$.

	MUSIC		ESPRIT		AR	
SNR	$f1$	$f2$	$f1$	$f2$	$f1$	$f2$
5	5.25×10^{-4}	-4.01×10^{-4}	3.29×10^{-4}	-2.52×10^{-4}	-6.37×10^{-4}	7.95×10^{-4}
10	1.39×10^{-4}	-1.37×10^{-4}	1.58×10^{-4}	-1.43×10^{-4}	-3.81×10^{-4}	4.06×10^{-4}
20	2.96×10^{-5}	-4.01×10^{-5}	2.99×10^{-5}	-3.46×10^{-5}	-3.78×10^{-5}	2.99×10^{-5}
30	8.25×10^{-6}	-1.25×10^{-5}	7.46×10^{-6}	-9.87×10^{-6}	-1.52×10^{-6}	-1.55×10^{-6}
40	2.51×10^{-6}	-3.94×10^{-6}	2.16×10^{-6}	-3.01×10^{-6}	8.11×10^{-7}	-1.61×10^{-6}
50	7.82×10^{-7}	-1.24×10^{-6}	6.64×10^{-7}	-9.43×10^{-7}	3.48×10^{-7}	-5.78×10^{-7}

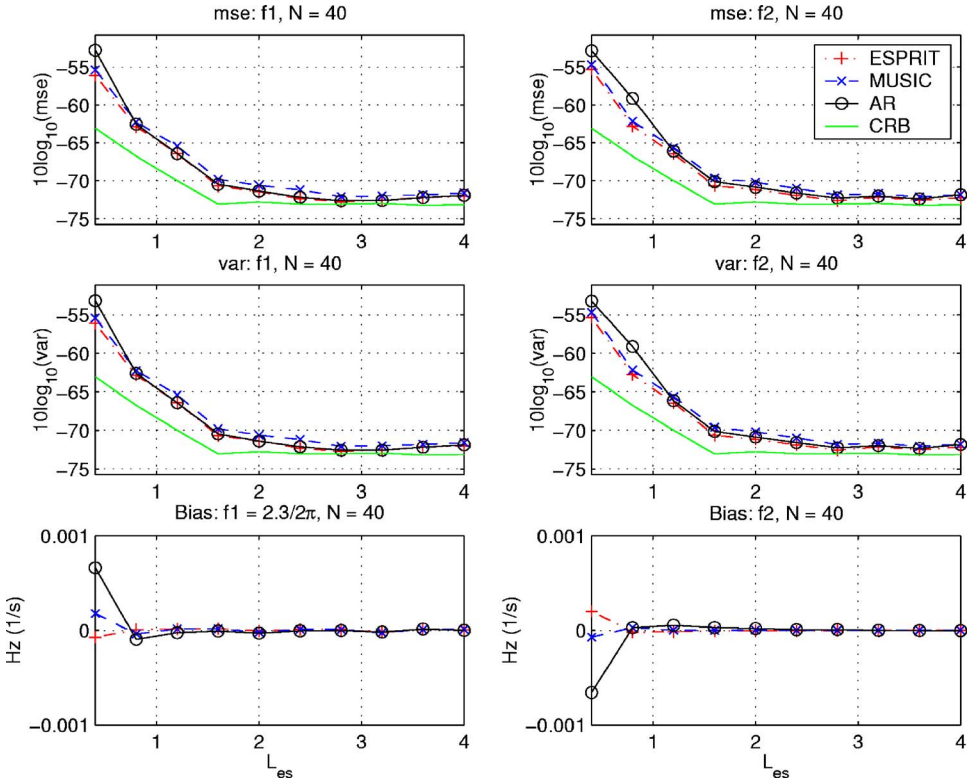


FIG. 2. (Color online) Frequency estimation error and bias vs L_{es} for estimating $f_1=2.3/2\pi$ and $f_2=2.5/2\pi$. SNR=20 dB and $0.36 \leqslant L_{es} \leqslant 4$.

der of 20 was necessary to theoretically resolve the two frequencies. For the shortest estimation length, the AR algorithm was not always able to resolve two individual peaks in the spectrum. In these instances, the AR estimator was considered to have failed. Nonetheless, estimator performance was determined as a percentage of the number of successful runs. Justification for this methodology is given in the next paragraph. Figure 3 shows the probability of success for AR, MUSIC, and ESPRIT as for estimating f_1 as L_{es} was increased. Overall, the three estimators showed about equal performance over the different estimation lengths. In particular, for $L_{es} \geqslant 1.2$, a 100% success rate was achieved. These results are consistent with the results of the previous section where all estimators performed equally well for $L_{es} > 1.5$ and also with the results of Rajan and Bhatta.[9]

The effect of model order selection on resolution and variance was considered for a signal comprised of 40 samples with two sinusoids at $f_1=0.4$ Hz and $f_2=0.35$ Hz and a SNR of 20 dB. In addition to shedding light on the effect of model order on resolution, this test illustrates the introduction of spurious peaks into the spectrum for high model orders. This effect is clearly shown in Figs. 4 and 5.

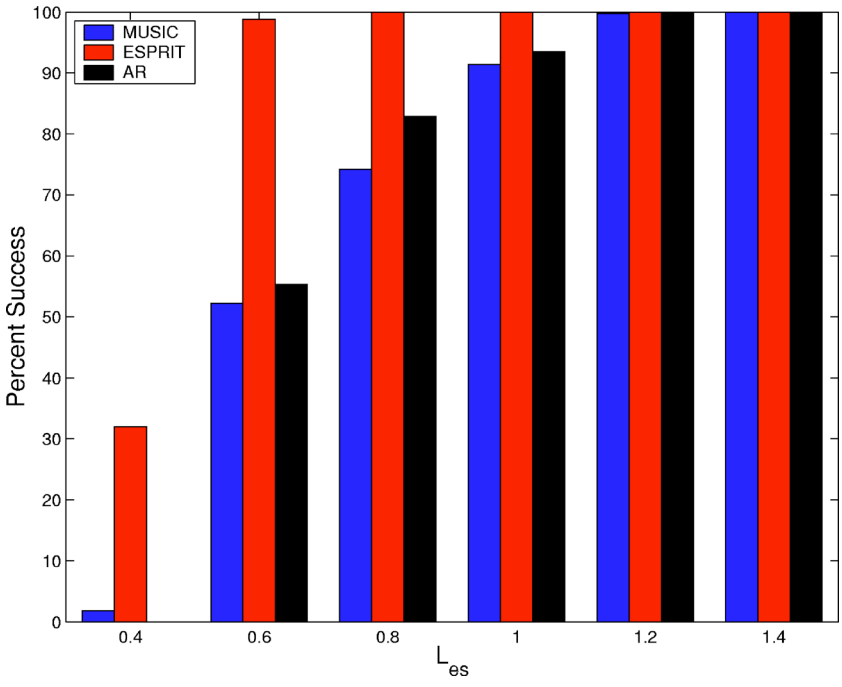


FIG. 3. (Color online) Probability of success for estimating $f_1 \pm 0.001$ for increasing L_{es}.

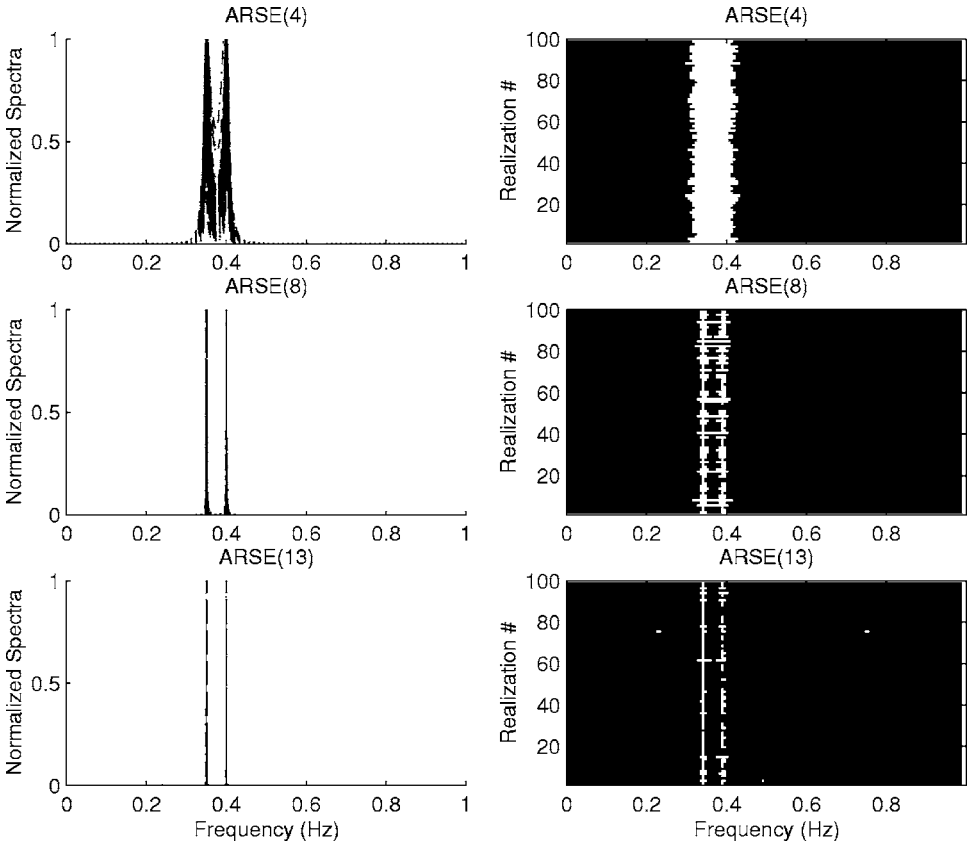


FIG. 4. Spectral estimates for 100 realizations of two frequency cisoid with additive noise. SNR=20 dB, f_1 =0.35 Hz, f_2=0.40 Hz. AR model order is indicated at top of each plot in parenthesis.

For these two the figures, in the left column are the estimates of each spectrum for the 100 signal realizations plotted on top of one another. The right column is a plan view of the individual spectral estimates plotted by realization number. In the plan view, spectral peaks below 60 dB down are not shown. This display gives an indication of both peak width and location and shows a decrease in peak width with increasing model order. It is interesting to note that for a signal

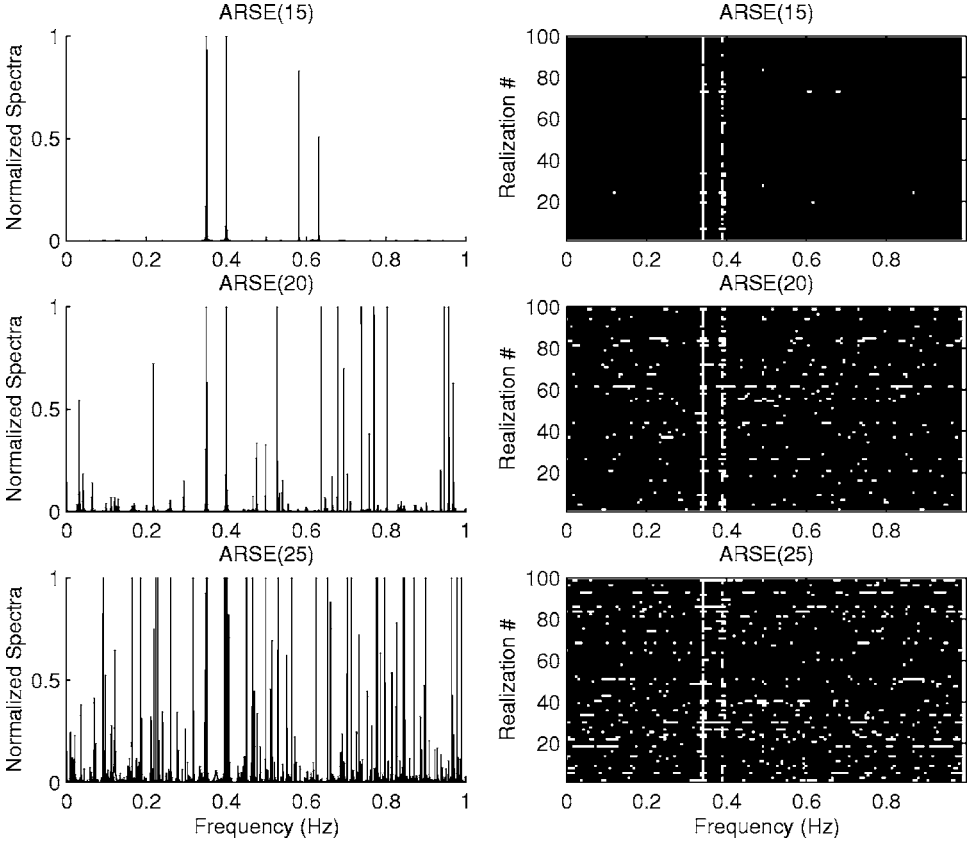


FIG. 5. Spectral estimates for 100 realizations of two frequency cisoid with additive noise. SNR=20 dB, f_1 =0.35 Hz, f_2=0.40 Hz. AR model order is indicated at top of each plot in parenthesis.

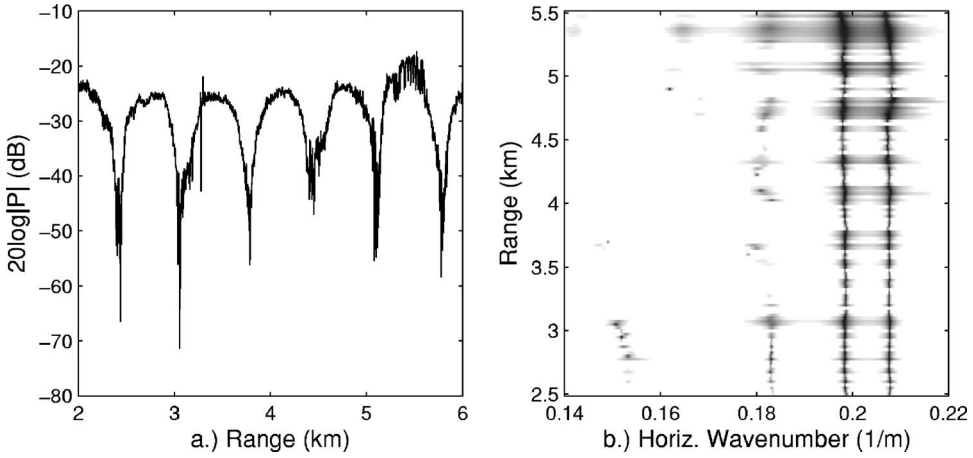


FIG. 6. Range dependent wave number estimates vs range using AR estimator with 1000 m aperture stepping every 25 m. Total data aperture was 4 km and frequency was 50 Hz.

comprised of a known number of sinusoids, the addition of noise requires that the AR model order be larger than the number of frequencies to better estimate the spectrum as given by Eq. (4). For analysis of acoustic data, this precludes preselection of model order based on the estimated number of propagating modes from *a priori* knowledge of the environment. In addition, as model order increases past $N/3$, it can be seen that spurious peaks are introduced into the spectra. Extra peaks cause ambiguity in the selection of the peaks which represent the true signal. However, the plots also indicate that the peaks associated with the two known excited frequencies are stable over all realizations. This type of behavior is typical of AR wave number estimates for range-independent or slowly varying shallow-water waveguides, as illustrated by application to a real data set in Fig. 6, and has been used to identify the prominent propagating modes using the sliding window transform method.[19,21] Figure 6(b) indicates that modes 1 and 2 were strongly excited at all ranges, and mode 3 was only weakly excited and more variable as a function of range. These results provide justification for determining successful and unsuccessful AR runs as discussed previously. After identifying the strong spectral lines that are stable over realization numbers or ranges, spurious peaks appearing well outside the stable region for a single realization could be classified as outliers. The contributions from spurious peaks were removed when calculating the probability of success for small L_{es} shown in Fig. 3. The earlier analysis indicates that model order selection for AR spectral estimates with no *a priori* knowledge must strike a balance between resolution and the introduction of spurious peaks. This topic is covered in detail in Marple.[10] Further, the analysis indicates how it is possible to confidently estimate wave numbers corresponding to propagating modes in a waveguide by strong "lines" on the wave number/range plots obtained from a sliding window spectral estimator. For general analysis, it is recommended that model order be selected as $N/3$. If SNR is very high, a lower model order might also be used with equal success.

B. Application of AR estimator to acoustic data

The AR spectral estimator was applied to determine the horizontal wave numbers for the normal modes propagating in a shallow-water waveguide. For a range-independent shallow-water waveguide, the acoustic pressure field written in terms of normal modes is[5]

$$p(r;z_s,z) \sim \frac{\sqrt{2\pi}}{\rho(z_s)} e^{i\pi/4} \sum_{l=1}^{\infty} \Psi_l(0,z_s)\Psi_l(r,z) \frac{e^{ik_l r}}{\sqrt{k_l r}}, \tag{15}$$

where $\Psi_l(0,z_s)$ is the mode function at the source location, $\Psi_l(r,z)$ is the mode function in range and depth, l is the mode number, and $\rho(z_s)$ is density. Correcting for geometric spreading and adding a noise term, the pressure signal to be analyzed can be represented for a fixed depth and discrete range intervals as[9]

$$y[n] = \sqrt{n\Delta r}P[n\Delta r] + w[n\Delta r] = \sum_{l=1}^{L} A_l \Psi_l + w[n\Delta r], \tag{16}$$

where the range interval is sampled by Δr, and $\Psi_l = e^{ik_l n\Delta r}$, is the mode function with horizontal wave number k_l and amplitude A_l which depends on the mode excitation at the source location. The objective was to estimate the k_l from the data measured along r. To generate synthetic complex-pressure data the normal-mode acoustic propagation code KRAKEN (Ref. 22) was used. KRAKEN calculates normal modes at a given frequency for a given environmental model. The complex pressure field can then be constructed from the modes at all ranges for given source and receiver depths. A Pekeris waveguide environment was used with the parameters listed in Table IV. The source frequency was 100 Hz, with the source and receivers at depths of 20.5 and 15.5 m, respectively. The field was calculated on a 1 m range grid. Modal eigenvalues and amplitudes at the source and the receiver determined by KRAKEN are given in Table V. For this waveguide and source/receiver geometry modes 1 and 2 had almost equal excitation amplitudes, while mode 3 was only very weakly excited. At the receiver depth, the ampli-

TABLE IV. Pekeris ocean model. Frequency=100 Hz, source depth =20.5 m, and receiver depth=15.5 m.

Depth (m)	c_s (m/s)	ρ (g/cm^3)	α (Np/m)
0-40	1515.0	1.0	0.0
40+	1800.0	1.56	0.001

TABLE V. Modal eigenvalues and corresponding amplitudes at source and receiver for Pekeris ocean model.

Mode No.	k_l (1/m)	Ψ_s	Ψ_r
1	0.409260	0.179	0.204
2	0.392023	0.181	0.074
3	0.361364	−0.003	−0.178

tude of mode 2 was about 1/3 that of mode 1. Based on the individual mode excitation amplitudes and the relative amplitudes at the receiver depth, relative-wave number spectrum amplitudes were expected to range from high for mode 1 to very low for mode 3. Thus, mode 3 was expected to be difficult to detect even at high SNR. For the given field, wave number estimates were made and mse calculated for various aperture lengths and SNR levels. For a Δk_r between modes 1 and 2 of 0.0172 m^{-1} the estimation lengths ranged from 4.1 to 8.2, where N was equal to the aperture length for $\Delta r=1$. AR model order was set to be $N/3$ and determined by total aperture length. The mse results in Table VI are consistent with the expectations discussed above. Using either the AR estimator or ESPRIT, the mse for estimating the first mode was approximately 10 dB better than the estimates for mode 2. Errors for estimating mode 3, when it could be resolved, were considerably worse than for estimating modes 1 and 2. For most cases the AR estimator and ESPRIT estimator results were closely matched. One exception occurred when estimating the first mode for signals with a 40 dB SNR and using a 240 m aperture. For this case, the AR estimator result was considerably better than the ESPRIT result. This result was not expected and is not very well understood. However, it is beyond the scope of this paper to explore the ESPRIT results in any greater detail. Overall, and consistent with the previous study of time series data, the AR estimator performed as well as or slightly better than the other high-resolution methods.

IV. APPLICATION TO RANGE-DEPENDENT ENVIRONMENTS

This work was primarily motivated by the need to estimate wave number content for range-dependent environments. Some evidence of range dependence can be seen in the example shown in Fig. 6. There is a slight shift in the wave numbers for modes 1 and 2 starting around 4.7 km and there is much variability in the third mode. However, the data have not been fully assessed to draw any firm conclusions about range dependence. Previous work also considered the use of high-resolution estimators for environments known to be range dependent.[9] Wave number estimates were obtained using short aperture transforms for an environment with range-dependent bathymetry. Synthetic acoustic data were generated for a waveguide with homogeneous sediment properties and a bathymetric slope of 1.67 m/km, making an angle of 0.955° from the horizontal. For this slightly range-dependent case, the pressure field can be approximated using adiabatic mode theory as[23]

$$p(r;z_s,z) \sim \frac{\sqrt{2\pi}}{\rho(z_s)} e^{i\pi/4} \sum_{l=1}^{\infty} \Psi_l(0,z_s)\Psi_l(r,z) \frac{e^{i\int_0^r k_l(r')dr'}}{\sqrt{\int_0^r k_l(r')dr'}}. \tag{17}$$

Under this assumption, horizontal wave numbers are expressed as a function of range. Estimates for wave numbers from the range-dependent synthetic acoustic field fell between those expected for the minimum and maximum depths spanned by the aperture. It was concluded that errors associated with wave number estimates under the assumption of locally range-independent segments were small. Further, errors associated with this type of slight range dependence would be indistinguishable in the presence of noise. The behavior described above in the reference study[9] was observed for wave number estimates made on an independent data set provided as part of a recent geoacoustic inversion techniques (GAIT) workshop.[19,24]

A. Range dependent sediment properties

Examining Eq. (17) earlier, in addition to range-dependent bathymetry, wave numbers must satisfy local boundary conditions and are sensitive to changes in bottom type with range. Therefore, a more interesting environment from the GAIT workshop was a waveguide with range-dependent sediment properties. In contrast to the previous case where the sediment depth profile was constant with range and waveguide thickness varied, an environment was

TABLE VI. SNR (dB) vs $10\log_{10}$(mse) of eigenvalues for different data aperture lengths.

SNR (dB)	Aperture (m)	L_{es}	Mode No.	ESPRIT	AR
40	240	4.1	1	−70.94	−75.23
			2	−62.44	−63.94
			3	−8.80	−30.07
40	360	6.2	1	−87.38	−87.56
			2	−77.70	−78.35
			3	−51.28	−53.08
20	360	6.2	1	−70.57	−69.80
			2	−62.53	−62.29
20	480	8.2	1	−79.09	−79.18
			2	−72.41	−72.39

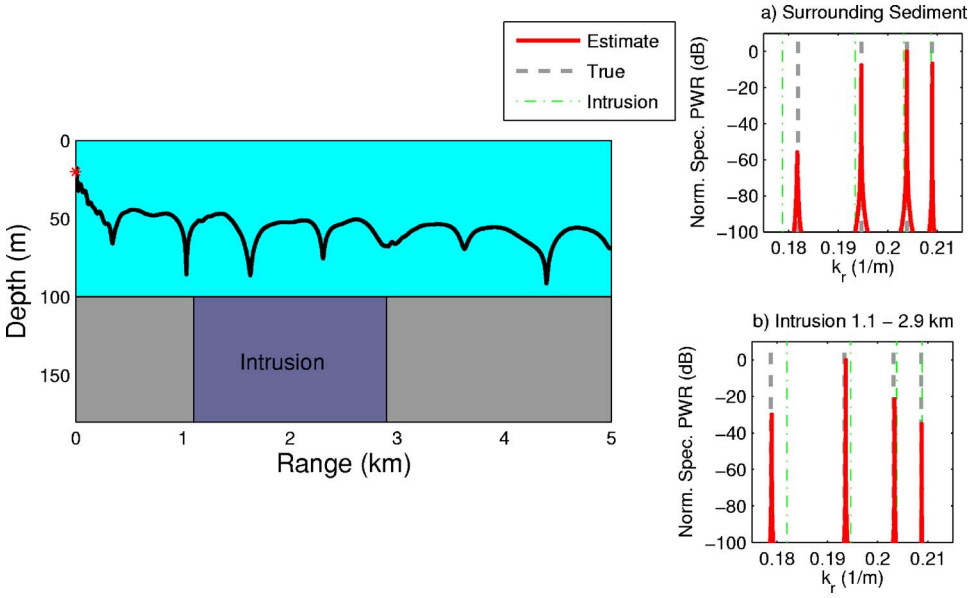


FIG. 7. (Color online) Schematic of waveguide and pressure magnitude at 50 Hz for GAIT test case 3. Resulting wave number spectral estimates for (a) surrounding sediment region and (b) intrusion for 50 Hz GAIT data. Dashed lines are true wave number values and dash-dot lines are wave numbers for respective adjacent sediment.

designed to simulate an uplifted fault by introducing an intrusion of basement material into the sediment layer.[25] The waveguide depth was constant and no *a priori* information regarding the nature or the location of the intrusion was provided. Range-dependent geoacoustic parameter estimates were to be obtained using the given acoustic field data and accompanying water column data, including depth, sound speed, and source/receiver locations.

When first attempting to analyze the given data set, a sliding window approach was used based on a classical FFT based estimator. The data considered were 50 Hz complex pressure data on a horizontal receiver array at a depth of 25 m. The total array aperture was 5 km, with elements spaced every 5 m. Using apertures of 1 and 2 km, the initial results indicated range dependence of the wave numbers. However, individual modal peaks could not be well estimated due to aperture length and windowing effects.[19] Experimenting with the data, it was determined that a high-resolution wave-number estimator would be required to resolve individual mode numbers and to track any changes with range.

An AR algorithm was implemented in a sliding window wave-number estimator and applied to the GAIT workshop data. Using a 1000 m window stepped every 5 m the location of the intrusion was determined by observing changes in estimated wave numbers as the window was moved in range. Using this approach, it was found that the intrusion occurred for ranges between 1100 and 2900 m from the source and was composed of material with a faster sound speed than the surrounding area. This result was verified by Chapman *et al.*[24] with full details of the methodology and results given by Becker *et al.*[21] It should be emphasized that the location of the leading edge of the intrusion at 1100 m ultimately dictated the size of the data aperture that could be used for this type of analysis. Having identified the location and extent of the intrusion, best estimates of wave number content were determined by using the full extent of the range-independent segments available. The nature of the waveguide environment and the resulting wave number estimates for the given pressure field data are shown in Fig. 7. Wave number spectral estimates are presented for both the intrusion and the surrounding material. The true wave number values for both regions are included in each subplot to illustrate the differences between the two regions.

V. CONCLUSIONS

An autoregressive spectral estimator was applied to estimate either the frequency content or wave number content of noisy signals. The high-resolution frequency estimation property of the AR estimator gives it an advantage over FFT-based methods for analyzing acoustic data with closely spaced wave numbers, and in particular for short aperture or range-dependent data. Here its performance was shown to be as good as ESPRIT and MUSIC for estimation lengths typical of shallow-water waveguide data. In addition, the AR approach is advantageous in that it does not require that the number of propagating modes be known *a priori*. Further, the number of and values for the individual modes can easily be established by strong lines in range/wave number plots. This approach has been successfully applied to a number of recent data sets, both synthetic[21] and real.[19] Its performance, evaluated here and from use on real data, has proven it to be a useful tool for extracting the modal content of acoustic field data measured on horizontal arrays.

ACKNOWLEDGMENTS

This work was supported by an Office of Naval Research Special Research Awards in Underwater Acoustics Entry-Level Faculty Award, Grant No. N00014-04-1-0248, and also ONR Grant No. N00014-04-1-0296.

[1]K. Ohta, K. Okabe, I. Morishita, S. Ozaki, and G. V. Frisk, "Inversion for seabed geoacoustic properties in shallow water experiments," Acoustical Science and Technology **26**, 326–337 (2005).
[2]K. Ohta and G. V. Frisk, "Modal evolution and inversion for seabed geoacoustic properties in weakly range-dependent shallow-water waveguides," IEEE J. Ocean. Eng. **22**, 501–521 (1997).
[3]G. V. Frisk and J. F. Lynch, "Shallow water waveguide characterization

using the Hankel transform," J. Acoust. Soc. Am. **76**(1), 205–216 (1984).
[4]G. V. Frisk, K. M. Becker, and J. A. Doutt, "Modal mapping in shallow water using synthetic aperture horizontal arrays," invited paper in *Proceedings of the Oceans 2000 MTS/IEEE Conference and Exhibition*, Providence, RI, 11–14 September 2000, Vol. **1**, 185–188.
[5]G. V. Frisk, *Ocean and Seabed Acoustics: A Theory of Wave Propagation* (PTR Prentice-Hall, Englewood Cliffs, NJ, 1994).
[6]G. V. Frisk, J. F. Lynch, and S. D. Rajan, "Determination of compressional wave speed profiles using modal inverse techniques in a range-dependent environment in Nantucket Sound," J. Acoust. Soc. Am. **86**(5), 1928–1939 (1989).
[7]R. T. Kessel, "Extracting modal wave numbers from data collected in range dependent environments," J. Acoust. Soc. Am. **104**(1), 156–163 (1998).
[8]F. J. Diemer, "A Prony algorithm for shallow water waveguide analysis," Ocean Engineer's thesis, MIT/WHOI Joint Program, Cambridge, MA and Woods Hole, MA, 1987.
[9]S. D. Rajan and S. D. Bhatta, "Evaluation of high-resolution frequency estimation methods for determining frequencies of eigenmodes in shallow water acoustic field," J. Acoust. Soc. Am. **93**(1), 378–389 (1993).
[10]S. L. Marple, Jr., *Digital Spectral Analysis with Applications* (Prentice-Hall, Englewood Cliffs, NJ, 1987).
[11]R. Roy, A. Paulraj, and T. Kailath, "ESPRIT—A subspace rotation approach to estimation of parameters of cisoids in noise," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-34**(5), 1340–1342 (1986).
[12]P. Stoica and R. Moses, *Introduction to Spectral Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1997).
[13]A. B. Baggeroer, "Confidence intervals for regression (MEM) spectral estimates," IEEE Trans. Inf. Theory **IT-22**(5), 534–545 (1976).
[14]S. D. Rajan, G. V. Frisk, and J. F. Lynch, "On the determination of modal attenuation coefficients and compressional wave attenuation profiles in a range-dependent environment in Nantucket Sound," IEEE J. Ocean. Eng. **17**(1), 118–128 (1992).
[15]S. M. Kay and S. L. Marple, "Spectrum analysis—A modern perspective," Proc. IEEE **69**(11), 1380–1419 (1981).
[16]S. M. Kay, *Modern Spectral Estimation: Theory and Application* (Prentice-Hall, Englewood Cliffs, NJ, 1988).
[17]D. C. Rife and R. R. Boorstyn, "Single tone parameter estimation from discrete-time observations," IEEE Trans. Inf. Theory **IT-20**(5), 591–598 (1974).
[18]D. C. Rife and R. R. Boorstyn, "Multiple-tone parameter estimation from discrete-time observations," Bell Syst. Tech. J. **55**(9), 1389–1410 (1976).
[19]K. M. Becker, "Geoacoustic inversion in laterally varying shallow-water environments using high-resolution wavenumber estimation," Ph.D. thesis, MIT/WHOI Joint Program, Cambridge, MA and Woods Hole, MA, 2002.
[20]H. Sakai, "Statistical properties of AR spectral analysis," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-27**(4), 402–409 (1979).
[21]K. M. Becker, S. D. Rajan, and G. V. Frisk, "Results from the geoacoustic inversion techniques workshop using modal inverse methods," IEEE J. Ocean. Eng. **28**, 331–341 (2003).
[22]M. B. Porter, "The KRAKEN normal mode program," Rep. SM-245 (SACLANT Undersea Research Centre, La Spezia, Italy, 1991).
[23]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, New York, 1994).
[24]N. R. Chapman, S. Chin-Bing, D. King, and R. B. Evans, "Benchmarking geoacoustic inversion methods for range-dependent waveguides," IEEE J. Ocean. Eng. **28**, 320–330 (2003).
[25]ONR/SPAWAR Inversion Techniques Workshop, 15–18 May 2001; http://itworkshop.nrlssc.navy.mil/

Predictions of formant-frequency discrimination in noise based on model auditory-nerve responses

Qing Tan
Boston University Hearing Research Center, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215

Laurel H. Carney[a)]
Boston University Hearing Research Center, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, and Departments of Biomedical & Chemical Engineering and Electrical Engineering & Computer Science, Institute for Sensory Research, 621 Skytop Road, Syracuse University, Syracuse, New York 13244

(Received 18 August 2005; revised 19 June 2006; accepted 21 June 2006)

To better understand how the auditory system extracts speech signals in the presence of noise, discrimination thresholds for the second formant frequency were predicted with simulations of auditory-nerve responses. These predictions employed either average-rate information or combined rate and timing information, and either populations of model fibers tuned across a wide range of frequencies or a subset of fibers tuned to a restricted frequency range. In general, combined temporal and rate information for a small population of model fibers tuned near the formant frequency was most successful in replicating the trends reported in behavioral data for formant-frequency discrimination. To explore the nature of the temporal information that contributed to these results, predictions based on model auditory-nerve responses were compared to predictions based on the average rates of a population of cross-frequency coincidence detectors. These comparisons suggested that average response rate (count) of cross-frequency coincidence detectors did not effectively extract important temporal information from the auditory-nerve population response. Thus, the relative timing of action potentials across auditory-nerve fibers tuned to different frequencies was not the aspect of the temporal information that produced the trends in formant-frequency discrimination thresholds. © *2006 Acoustical Society of America*
.[DOI: 10.1121/1.2225858]

PACS number(s): 43.64.Bt [WPS] Pages: 1435–1445

I. INTRODUCTION

The auditory system has a remarkable ability to extract speech information in the presence of noise. The study presented here was intended to improve our understanding of signal-processing and encoding mechanisms in the peripheral auditory system. Formant-frequency discrimination in background noise was studied using simulations of cat model auditory-nerve (AN) responses, and predictions of discrimination performance based on the AN model responses were compared with behavioral data in cat (Hienz *et al.*, 1998). Analyses of the AN model responses were based on either the information present in the discharge rate alone or on the combined information in both the rate and timing of AN discharges. In addition, one simple neural mechanism for decoding temporal information, coincidence detection, was investigated to determine whether cross-frequency coincidences of AN discharges contributed to the temporal encoding of formant frequencies.

Human thresholds for frequency discrimination of the second formant (F2) have been reported in the range of 1%–5% (Flanagan, 1955) to 6.8% (Mermelstein, 1978) in earlier studies, and as low as 1.9% (Hawks, 1994) and 1%–2% (Kewley-Port and Watson, 1994; Kewley-Port *et al.*, 1996) in more recent studies. The latter studies included highly trained listeners in paradigms with low stimulus uncertainty. These thresholds are roughly an order of magnitude higher than pure-tone frequency discrimination thresholds; the pure-tone frequency discrimination threshold for a tone in the F2 frequency region (e.g., 2000 Hz) is 3.2 Hz (0.16%) (Wier *et al.*, 1977). However, formant-frequency discrimination differs fundamentally from pure-tone frequency discrimination. Changes in the formant frequency are not associated with changes in the frequencies of the harmonics, but rather with a pattern of level changes distributed across the harmonic frequencies in the region of the formant peak (Fig. 1). Thus, understanding formant-frequency discrimination presumably requires an understanding of how the combined changes in the amplitudes of the harmonics associated with changes of the spectral envelope's center frequency affect the response patterns of the auditory periphery.

Quantitative tools are available for predicting behavioral thresholds based on the responses of AN fibers. Siebert (1965, 1968) developed a strategy for predicting limits of level discrimination for pure tones based on a very simple description of the auditory periphery. Heinz *et al.* (2001a) extended Siebert's approach to allow the use of recent computational models for AN responses. Since both approaches

[a)]Author to whom correspondence should be addressed: Institute for Sensory Research, 621 Skytop Road, Syracuse University, Syracuse, NY 13244. Electronic mail: Lacarney@syr.edu

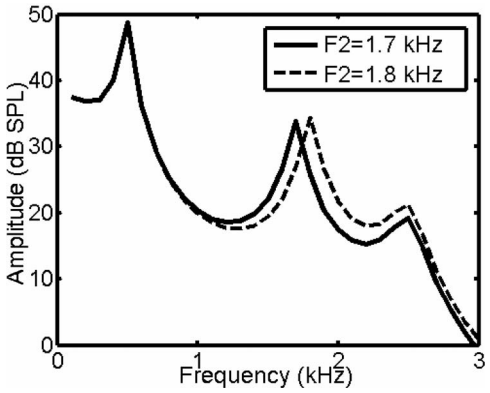


FIG. 1. Examples of spectral envelopes of synthesized vowels, illustrating the vowel with the standard second formant frequency (1700 Hz, solid) and with a frequency shift in the second formant frequency (1800 Hz, dashed).

include analyzing population responses and combining information across fibers tuned to different frequencies, they provide appropriate tools to study formant-frequency encoding. Furthermore, Heinz (2000) developed a computational strategy for predicting psychophysical performance on tasks that involve tones in noise maskers. We have applied and extended his strategies in this study of formant-frequency discrimination of vowels in quiet and in the presence of noise maskers.

Understanding speech coding in the presence of masking noise is a real challenge for both physiologists and psychophysicists. A central question in this puzzle is what aspects of AN responses convey information about complex sounds to the central nervous system. Most studies of speech sounds, including studies of speech recognition, focus on spectral energy, and the most straightforward coding strategy for energy is typically assumed to be average discharge rate. However, intense noise maskers and the relatively high levels of harmonics near formant peaks tend to dominate the rates of AN fibers, reducing the sensitivity of rate to small changes in the levels of neighboring harmonics. Hienz *et al.* (1996) concluded that behavioral results in cat for vowel discrimination in the presence of high-level maskers could not be explained by the rates of high-spontaneous-rate (HSR) AN fibers, which are severely degraded in this stimulus condition (Sachs and Young, 1979; Delgutte and Kiang, 1984b; Sachs *et al.*, 1983). Although low-spontaneous-rate (LSR) fibers may provide a wider dynamic range based on rate, it is clear from our results that there is ample information in the timing of small populations of HSR fibers to explain the behavioral results. The importance of the temporal responses of small populations of fibers tuned near the frequency of interest is consistent with previous studies (Young and Sachs, 1979; Sachs *et al.*, 1983; Delgutte and Kiang, 1984a, b; Miller *et al.*, 1987; Tan and Carney, 2005).

One concern is that quantitative predictions of psychophysical performance based on physiological responses often require comparisons across species. Most knowledge of AN responses is based on the cat, whereas most psychophysical knowledge of vowel discrimination is based on studies with human listeners. In the case of formant-frequency discrimination, however, behavioral results in the cat (Hienz *et al.*, 1998) can be used to directly compare predictions based on cat model AN responses to psychophysical results for F2 discrimination. Hienz *et al.* (1998) measured behavioral formant-frequency discrimination thresholds in cat, for both vowels in quiet and with noise maskers at various signal-to-noise ratios. Their results show that formant-frequency-discrimination thresholds of cat at low and medium noise levels [signal-to-noise ratios (SNRs) of 23 and 13 dB, respectively] are similar to the threshold measured in quiet (approximately 2.5% of the F2 frequency), whereas the threshold in the presence of a high noise level (SNR of 3 dB) is significantly elevated (approximately 5.8%). In human listeners, a similar trend in thresholds is seen as a function of SNR, with thresholds increasing very gradually for high SNRs and then increasing more significantly at low SNRs (Liu and Kewley-Port, 2004).

In the current study, the Hienz *et al.* (1998) formant-frequency discrimination task was simulated using a population of cat AN model fibers (Tan and Carney, 2003). This AN model was used in a previous study (Tan and Carney, 2005) of a related but simpler task involving center-frequency discrimination for harmonic complexes (Lyzenga and Horst, 1995). In the previous study, the prediction methods assumed that an ideal central processor used detailed knowledge of the AN population response pattern for each waveform. However, this assumption is not realistic when a random noise masker is included in the stimulus, because knowledge of the details of the noise response allows an ideal processor to perform at extremely low thresholds (this issue is discussed further below). The study presented here represents an application of the approach of Heinz (2000) and Heinz *et al.* (2002), in which it was assumed that the central processor only has knowledge of the *averaged* response patterns to the stimuli that include noise (i.e., the AN model response was averaged across an ensemble of stimuli that included different tokens of random noise maskers).

Given the result that temporal information is sufficient to explain behavioral performance for formant-frequency discrimination, it is interesting to explore what aspects of the temporal responses are most important, and what neural mechanisms might be used to extract them. In this study, predictions for formant-frequency discrimination thresholds were also made using a coincidence-detection mechanism, which has achieved some success in predicting psychophysical thresholds for tone detection in noise (Carney *et al.*, 2002). The best predictions in the current study utilized both average-rate and timing information of a relatively small population of model auditory-nerve fibers tuned near the formant frequency. Predictions based on the average rates of cross-frequency coincidence detectors were not as successful as predictions based directly on temporal responses of model AN fibers or as predictions based on the timing of the responses of coincidence detectors. This comparison suggests that the relative timing of AN responses tuned to different frequencies was not the critical temporal aspect in the peripheral responses for the formant-frequency discrimination task.

II. METHODS[1]

A. Stimuli

Vowel stimuli were based on those used in Hienz *et al.* (1998): the synthetic vowel /ɛ/ was produced by a cascade Klatt Speech Synthesizer (Klatt, 1980) with a sampling rate of 10 kHz. The fundamental frequency of the vowel was 100 Hz. The first, third, fourth, and fifth formant frequencies of the synthesized speech signals were fixed at 500, 2500, 3300, and 3750 Hz, respectively. The second formant frequency varied between 1700 and 1900 Hz. Figure 1 shows the spectral envelopes of two sample vowels with different second-formant frequencies: 1700 Hz (solid line) and 1800 Hz (dashed line). The discrimination threshold was computed for changes in the second formant frequency from the standard frequency of 1700 Hz based on the response pattern of the AN model fibers, as described below. The vowels were 250 ms in duration, with rise/fall times of 20 ms. Sound pressure levels of the synthetic vowels were computed on the basis of the total power in the first 30 harmonics (100 Hz to 3 kHz). Results are presented in two level ranges; vowel levels were randomly varied (roved) over a range of 10 dB (±5 dB, in 1-dB increments) around 50 dB and 70 dB SPL rms. These levels and rove ranges were chosen to match the stimulus conditions used in the Hienz *et al.* (1998) behavioral study.

Noise maskers were added to the vowels to create stimuli with three different signal-to-noise ratios: 23, 13, and 3 dB SNR, selected to match the Hienz *et al.* (1998) experiments. The masking noise had a flat spectrum; signal-to-noise ratios for vowels in background noise were computed based on the total noise power up to 3 kHz, and noise levels were varied with respect to the vowel level to adjust the SNR. The overall levels of the noise masker and vowel were roved while maintaining the SNR at the indicated value for each condition.

B. AN model

The Tan and Carney (2003) AN model was used for all simulations presented here. The time-varying rate function of the AN model response was used for all calculations; individual action potentials were not simulated. The time-varying rate function is proportional to the probability of AN discharges throughout the timecourse of the stimulus waveform. Its name derives from the fact that it represents the time-varying average arrival rate for a nonhomogenous Poisson process that describes the AN response. The units of the rate function are spikes/s, and it can be thought of as a prediction of the peri-stimulus-time (PST) histogram of an AN fiber.

Characteristic frequencies (CFs, the frequency to which a model AN fiber is most sensitive) were evenly distributed on a log frequency scale. To study second formant frequency encoding, CFs from 1200 to 2500 Hz were included, representing 3.1 mm along the cochlear frequency map in cat (Liberman, 1982). Based on Keithley and Schreiber's (1987) measurement of the frequency map of the spiral ganglion in cat, this population of CFs comprised approximately 2000 neurons/mm $\times$ 3.1 mm, for a total of 6200 AN fibers.

Approximately 61% of the total AN population are HSR fibers (Liberman, 1978). Because our AN model only applies to HSR fibers, only 61% of the AN population was included in our threshold predictions. Therefore, 3782 (0.61×6200) model AN fibers with CFs logarithmically distributed from 1200 to 2500 Hz were used in these simulations. Fifty CFs were simulated; therefore, 76 AN fibers represented each CF channel for a total of 3782 fibers.

C. Threshold predictions based on AN model responses

Predictions of formant-frequency-discrimination thresholds were computed from AN model responses using techniques introduced to our field by Siebert (1965, 1968, 1970). His approach was to determine the threshold of an optimal detector based on observations (i.e., AN discharge patterns) that included the randomness associated with a Poisson process. The randomness was assumed to describe the variability in AN responses. Siebert's studies focused on frequency and level discrimination of pure tones in quiet; however, Heinz (2000) extended this approach to study tone detection in the presence of a masker noise.

It is not realistic to assume that an ideal central processor takes advantage of every detail of the response to every waveform when the stimulus includes random noise. For example, a subtle aspect of a vowel-plus-noise waveform might change when the formant frequency is slightly shifted. An ideal processor can take advantage of such subtle changes because it can directly compare the response to the noise plus a vowel with the standard formant frequency with the response to the same noise waveform plus the vowel with the shifted formant frequency. Realistically, the randomness of the noise waveforms, and their variation from trial to trial, precludes the use of subtle details of these waveforms in behavioral experiments. To better model the experimental situation, a suboptimal processor can be considered, such that detailed knowledge of each masker waveform is not available to the processor. Heinz (2000, see also Heinz *et al.*, 2002) used this approach to explore psychophysical performance limits for detection of tones in the presence of random noise by assuming that the central processor uses knowledge of the *expected* response to a given type of masker, rather than knowledge of the responses to individual noise waveforms on each trial. In other words, the central processor uses the expected response pattern (the averaged response to an ensemble of waveforms that include random noises) as the *a priori* information for each trial.

In this study, we adopted Heinz's (2000) approach and extended it to include the stimulus variations associated with the roving-level paradigm. Thus, the suboptimal detector used *a priori* information that was based on the expected response to an ensemble of stimuli that included several noise masker waveforms and a range of vowel and masker levels. In particular, the suboptimal detector's decision was based on the comparison of the model AN response to a particular vowel-plus-noise stimulus with *a priori* information consisting of averaged model AN responses to 220 different stimuli. Responses to vowels were simulated for 20

noise masker waveforms at 11 different levels (±5 dB in 1 -dB increments).

A straightforward decision variable based on observations of the AN model *population* response is an equally weighted combination of the decision variables based on *single* AN model responses, i.e.,

$$Y(\tau) = \sum_{i=1}^{M} Y_i(\tau^i), \tag{1}$$

where $Y_i(\tau^i)$ is the decision variable based on the observation of the set of AN discharge times, τ, of the ith fiber's response pattern, and M is equal to 3782, the total number of fibers in the population. Thus, AN model fibers with relatively large changes in their responses as formant frequency is varied, and thus large changes in their decision variables [i.e., large changes in $Y_i(\tau^i)$], contribute more to the changes in the total value of the decision variable $Y(\tau)$.

In Heinz (2000, Eq. 5.1), the sensitivity index Q for the central processor is defined by

$$Q(f, f+\Delta f) = \frac{\{E_{w,\tau}[Y(\tau)|f+\Delta f] - E_{w,\tau}[Y(\tau)|f]\}^2}{\mathrm{Var}_{w,\tau}[Y(\tau)|f]}, \tag{2}$$

where $E_{w,\tau}[Y(\tau)|f+\Delta f]$ and $E_{w,\tau}[Y(\tau)|f]$ are the *expected* values of the decision variable used by the central processor across an ensemble of waveforms, w, and across the set of AN discharge times, τ, with and without the formant-frequency shift (Δf). The ensemble of waveforms for the current study includes the set of different vowel plus noise levels, including all of the different random masker tokens, that were included to simulate the roving-level paradigm. $\mathrm{Var}_{w,\tau}[Y(\tau)|f]$ is the variance of the decision variable due to both the random noise of the masker waveform (external noise) and the randomness of the AN discharges (internal noise), which was assumed to be described by Poisson statistics. Note that the metric Q is very similar to the commonly used measure of sensitivity, d'. (The two measures are equivalent in the case of Gaussian distributions with equal variance.) It is convenient to work with the normalized sensitivity squared, given by

$$[\delta'(f)]^2 = \frac{Q(f, f+\Delta f)}{(\Delta f)^2}. \tag{3}$$

Combining Eqs. (1)–(3) and using the description of sensitivity based on a single AN fiber rate function [e.g., Siebert (1970); see Heinz *et al.* (2001a) for derivation], the normalized sensitivity squared for the AN model population is [Eq. 5.34 in Heinz (2000)]

$$[\delta'(f)]^2 = \frac{\{\sum_{i=1}^{M} \int_0^T [1/\bar{r}_i(t|f)][\bar{\dot{r}}_i(t|f)]^2\, dt\}^2}{(\sum_{i=1}^{M} \int_0^T [1/\bar{r}_i(t|f)][\bar{\dot{r}}_i(t|f)]^2\, dt + \mathrm{Var}_w\{\sum_{i=1}^{M} \int_0^T [[\bar{\dot{r}}_i(t|f)]/\bar{r}_i(t|f)]r_i(t|f,w)dt\})}, \tag{4}$$

where M is the total number of fibers in the population; $r_i(t|f,w)$ is the response of the ith AN model fiber to the wth vowel-plus-masker with the second formant frequency, f. T is the duration of the stimulus waveform. The partial derivative of the rate function with respect to second formant frequency is indicated by a dot over the symbol, e.g., $\dot{r}_i(t|f)$. The overbar, e.g., $\bar{r}_i(t|f)$, indicates the averaged response of the ith model fiber across an ensemble of different masker waveforms and across a set of vowel and masker levels (for the roving-level paradigm), i.e., $\bar{r}_i(t|f) = E_w[r_i(t|f,w)]$, and $\bar{\dot{r}}_i(t/f) = E_w[\dot{r}_i(t|f,w)] = E_w[\partial\{r_i(t|f,w)\}/\partial f]$ is the averaged partial derivative of the response of the ith fiber. The partial derivative was estimated computationally as the difference between the rate functions in response to two slightly different second formant frequencies, divided by the difference between the formant frequencies. (A formant-frequency difference of 1 Hz was used in these calculations.) In the simulations presented here, an ensemble of 220 noise tokens (20 different noises at each of 11 SPLs within the rove range) was used to estimate the expected responses of the AN fibers to the vowel plus noise stimuli.

The complete derivation of Eq. (4) is rather extensive (see Chap. 5 of Heinz, 2000); however, it is possible to provide some intuitive explanation for the equation's form. The numerator of Eq. (4) is proportional to the change in the AN fiber's rate for small changes in formant frequency; thus, a larger partial derivative of the rate function with respect to formant frequency leads to higher sensitivity. The final sensitivity, though, depends not only on changes in the decision variable but also on its variance [see Eq. (2)]. The denominator in Eq. (4) thus represents the total variance of the decision variable; the first term in the denominator of Eq. (4) corresponds to the variance due to the Poisson randomness of the AN discharges (internal noise), and the second term corresponds to the variance introduced by the random noise maskers that vary across waveforms (external noise).

Finally, the relationship between the central processor's sensitivity and behaviorally measured thresholds must be established to compare model predictions to behavioral results. The form of the equation for Q [see Eq. (2)] is equivalent to

that for the square of the commonly used sensitivity index, d'. A conventional value of the sensitivity index associated with threshold (e.g., in a two-alternative forced-choice experiment) is $d'=1$, or equivalently $Q=1$. By setting Q equal to 1 in Eq. (3), the threshold, or just-noticeable difference (jnd), for the formant frequency can be identified. The model threshold corresponding to $Q(f, f+\Delta f_{\text{jnd}})=1$ is

$$\Delta f_{\text{AN}} = \frac{1}{\delta'(f)}. \tag{5}$$

This model threshold can then be directly compared to psychophysical thresholds for discrimination of formant frequency.

The above equations included both the rate and temporal information from observations of AN responses. Predictions can also be computed with only the average-rate information of the AN responses by assuming that the central processor knows only the number of discharges for each trial. In this case, the response of the ith fiber was assumed to be a stationary Poisson process with an average discharge rate of $R_i(f,w)$. The predicted performance based only on the information in the average response rate of the AN model fibers is given by [Eq. 5.36 in Heinz (2000)]:

$$[\delta'(f)]^2 = \frac{\{\sum_{i=1}^{M} ([1/\overline{R_i}(f)][\overline{\dot{R}_i}(f)]^2) \times T\}^2}{(\sum_{i=1}^{M} ([1/\overline{R_i}(f)][\overline{\dot{R}_i}(f)]^2) \times T + \text{Var}_w\{\sum_{i=1}^{M} [[\overline{\dot{R}_i}(f)]/\overline{R_i}(f)] \times R_i(f,w) \times T\})}. \tag{6}$$

The form of Eq. (6) is the same as that of Eq. (4), except that the time-varying rate functions have been replaced by average rates. The model threshold obtained with only average rate information was computed using Eqs. (5) and (6), and this model threshold was then directly compared to psychophysical thresholds.

D. Threshold predictions based on coincidence detection

Cross-frequency coincidence detection has been proposed as a mechanism to detect temporal cues encoded in sound stimuli (Heinz *et al.*, 2001b; Carney *et al.*, 2002; Colburn *et al.*, 2003). It has been shown that cross-fiber coincidence counts do not successfully predict trends in psychophysical thresholds for discriminating the center frequency of harmonic complexes (Tan, 2003). However, the coincidence-detection mechanism has been reported to successfully predict psychophysical thresholds for detection of tones in the presence of background noise (Carney *et al.*, 2002). Thus the ability of the coincidence-detection mechanism to discriminate formant frequencies in the presence of a noise masker was quantitatively evaluated here.

For simplicity, the outputs of the coincidence detectors were assumed to be independent nonstationary Poisson processes. The quantification methods were adapted from those used for the AN model fiber predictions [Eq. (4)], except that the total sensitivity was accumulated over the population of coincidence detectors instead of the population of AN model fibers:

$$[\delta'(f)]^2 = \frac{\{\sum_i \sum_j \int_0^T [1/\overline{C_{ij}}(t|f)][\overline{\dot{C}_{ij}}(t|f)]^2 dt\}^2}{(\sum_i \sum_j \int_0^T [1/\overline{C_{ij}}(t|f)][\overline{\dot{C}_{ij}}(t|f)]^2\, dt + \text{Var}_w\{\sum_i \sum_j \int_0^T [\overline{\dot{C}_{ij}}(t|f)/\overline{C_{ij}}(t|f)] C_{ij}(t|f,w) dt\})}. \tag{7}$$

$C_{ij}(t|f,w)=\sum_{l=1}^{K_i}\sum_{m=1}^{K_j} F(t_l^i - t_m^j, w)$ describes the coincidence detector that receives the responses to waveform w of the ith and the jth AN model fibers, where F is a brief (e.g., 20 μs) rectangular coincidence window with unity height. K_i and K_j are the total number of discharges on the i^{th} and j^{th} model fibers. As above, the partial derivatives with respect to the second formant frequency are indicated by dots in Eq. (7), and averages across different vowel-plus-masker waveforms and levels (for the roving-level paradigm) are indicated by overbars. The time-varying rate function of each coincidence detector $[C_{ij}(t)]$ was estimated based on the products of the time-varying rate functions of the convergent model AN fibers (Colburn, 1969, 1977; Heinz *et al.*, 2001b). The product of the rate function and the time step used in the simulations is equal to the probability of a discharge within a time step, and the

probability of a response in the coincidence detector is the product of the response probabilities of its two inputs, thus

$$C(t) = p_{ij}(t)/\Delta t = [p_i(t)p_j(t)]/\Delta t = [r_i(t)\Delta t][r_j(t)\Delta t]/\Delta t = r_i(t)r_j(t)/\Delta t, \quad (8)$$

where $p(t)$ is the time-varying probability of discharge, $C(t)$ and $r(t)$ are time-varying rate functions, and Δt is the time step used in the simulation. (Note that these time-varying functions are actually implemented in discrete time, but are shown as continuous-time functions for simplicity.)

The population of coincidence detectors consisted of model cells that received all pairwise combinations of the range of model AN CFs used in the simulation. In this study, each AN fiber innervated a single coincidence detector and the response of each model coincidence detector was simulated using one pair of model AN fibers. Thus the 3782 HSR AN fibers with CFs in the frequency range of interest innervated 1891 coincidence detectors. As stated above, the responses of these 3782 AN fibers were simulated by AN models of 50 different CFs, and these 50 AN models resulted in $(50^2+50)/2=1275$ distinct pairwise CF combinations as the inputs to the model coincidence detectors. This expression excludes the redundant pairwise combinations of AN CFs but retains coincidence detectors with same-CF inputs. Each of these simulated CF combinations represented about one or two coincidence detectors $(1891/1275=1.5)$.

Formant-frequency discrimination thresholds were predicted by using the rate-plus-timing information in the responses of coincidence detectors and by using only the rate information in these responses. For the rate-only predictions, coincidence-detector counts were analyzed in a manner similar to that used for the AN responses [see Eq. (6)]. The interpretation of the coincidence-detection predictions was as follows: Coincidence detectors are driven by temporal coincidences across fibers. If this form of the temporal information in the AN responses was most important for encoding formant frequencies, then the counts (or rate) of the model coincidence-detector cells should be as successful as the AN rate-plus-timing predictions in explaining the behavioral results. Thus, if the coincidence-detectors counts are *not* as successful as the AN rate-plus-timing predictions, then the coincidence detectors are *not* extracting the temporal information that is critical for explaining the thresholds. Predictions based on both rate and timing of the coincidence detectors illustrate the extent to which timing information is preserved (but not "extracted" into response rates) by these model cells.

III. RESULTS

Figure 2 illustrates model and behavioral thresholds as a function of SNR. The goal of this work was to model the trends in performance, not absolute threshold values. Thresholds for optimal detectors are often significantly lower than empirical thresholds, but if the correct trends across different stimulus conditions are observed, then straightforward methods can be used to elevate the thresholds to match the data (e.g., by making the model less optimal). However, if a model predicts thresholds that are either higher than observed thresholds or have incorrect trends across conditions, then that model can be eliminated.

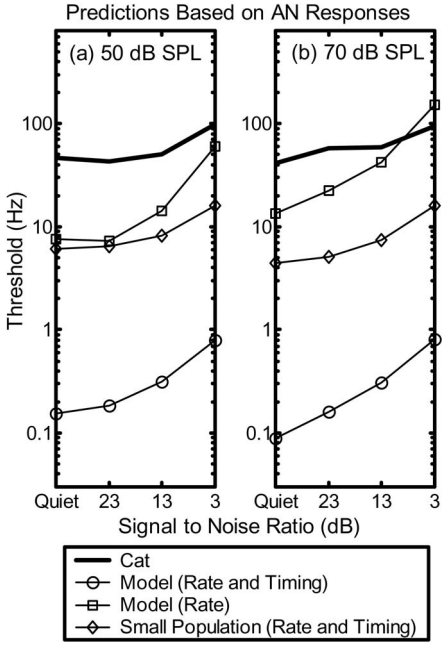


FIG. 2. Formant-frequency discrimination thresholds at various SNRs. Stimulus (combination of vowel and noise) levels were randomly varied (roved) with level uniformly distributed over a ±5-dB range in increments of 1 dB. Average vowel levels were (a) 50 dB SPL rms and (b) 70 dB SPL rms. Thick lines are cat performances, no lesions [(Hienz *et al.*, 1998), open bars of their Fig. 3]. Model threshold predictions were based on rate-only information for the population of model fibers with CFs between 1200 and 2500 Hz (squares), on both rate and timing information for the same population of model fibers (circles), and on both rate and timing information of a small group ($N=76$) of model fibers all having the same CF of 1720 Hz (diamonds).

Cat performance (thick line in Fig. 2, from Hienz *et al.*, 1998[2]) shows similar thresholds in quiet, in low-level noise, and in medium-level noise (the curve is almost flat), and the threshold in high-level noise is about twice as high as for the other noise levels. In contrast, the model's predicted thresholds based on the average response rate of the AN model fibers (squares) increased progressively as noise level increased across the entire range of noise levels tested. Performance based on both average rate and timing information of the AN model fibers (circles, based on all the model fibers with CFs between 1200 and 2500 Hz) also showed an increasing trend in thresholds across all noise levels. However, the predicted thresholds based on a single frequency channel (diamonds, the set of AN fibers with CF equal to 1720 Hz) were similar for quiet and for low and medium noise levels, and then increased for the high noise level, in agreement with the cat performance. This improved similarity between performance predicted by the responses of a subset of AN model fibers and psychophysical data is consistent with previous studies (Young and Sachs, 1979; Delgutte and Kiang, 1984a, b; Miller *et al.*, 1987; Tan and Carney, 2005).

The threshold predictions shown in Fig. 2 were made using the roving-level paradigm used by Hienz *et al.* (1998) for the behavioral tests. Model thresholds were also estimated using fixed-level stimuli to investigate the influence of

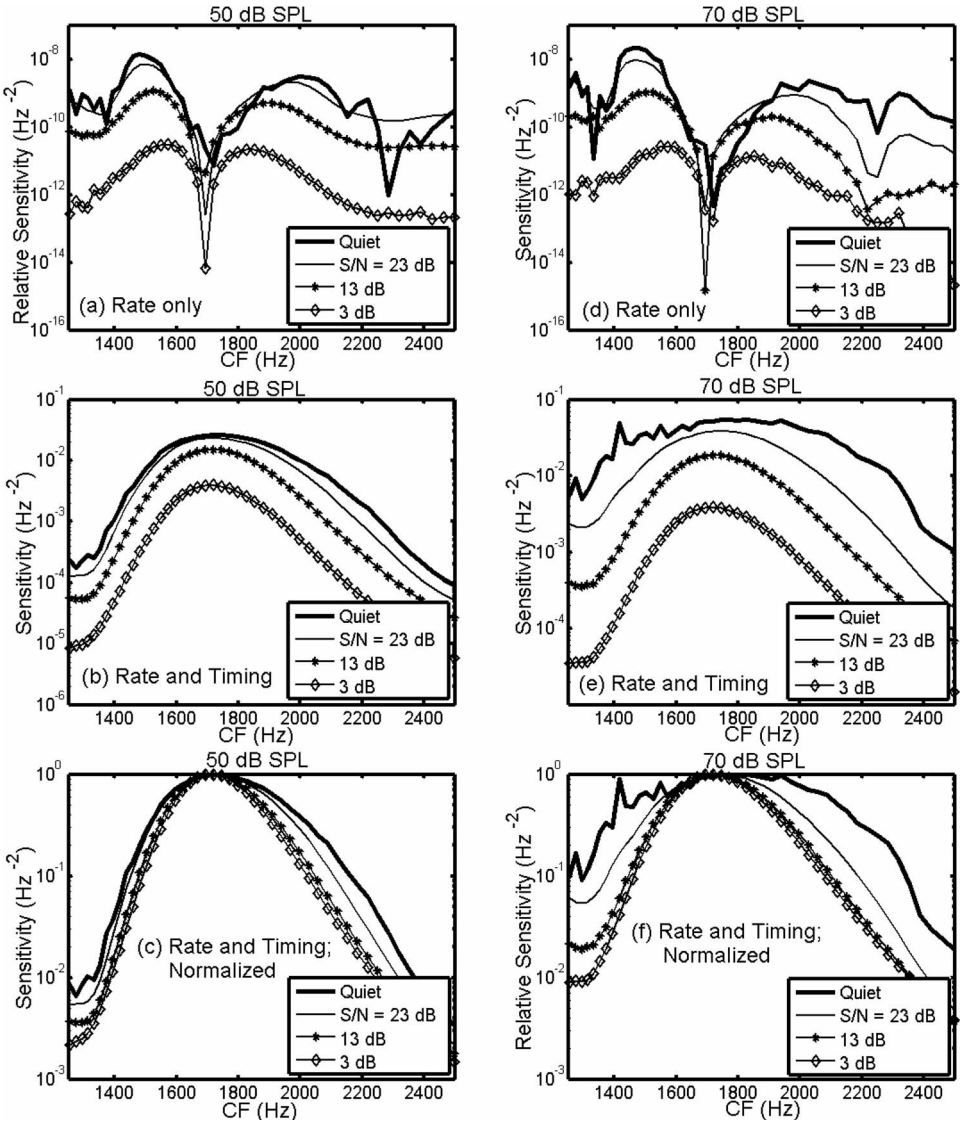


FIG. 3. Sensitivities computed for individual CF channels [Eqs. (4) and (6) computed for one CF at a time]. Roving vowel plus masker levels averaged 50 (left column) and 70 dB SPL (right column), with SNRs of 3 dB (diamonds), 13 dB (asterisks), 23 dB (thin lines), and quiet (bold). Results are based on the model using average rate information (a and d), or using both rate and timing information [(b) and (e)]. Panels (c) and (f) show the sensitivity metric (rate and timing information) normalized by the peak values to better illustrate changes in the width of the curves across SNRs.

the rove on model predictions (not shown). As expected, thresholds based on rate were most affected by the rove for vowels in quiet (thresholds were increased by about 150% due to the rove), and the elevation of threshold due to the rove dropped to about 10% for the lowest SNR studied (3 dB) for both 50 and 70 dB SPL vowels. Model thresholds based on both rate and timing were relatively unaffected by the rove; thresholds were elevated by approximately 1% or less for all SNRs for predictions based on the AN population, and by less than 1% for predictions based on the subset of fibers tuned near the formant frequency for the 50-dB vowels. The effect of the rove was slightly larger for the 70-dB SPL vowels (up to 2% threshold elevation for vowels in quiet based on rate and timing of the subset of model fibers).

The sensitivity of each AN model fiber for predictions made with average-rate information was calculated by evaluating Eq. (6) for individual CF channels (i.e., sensitivity was computed separately for each of the 50 different AN model CFs) and examining the results across CF [Figs. 3(a) and 3(b)]. (Note that Fig. 3 illustrates how sensitivity varies as a function of CF, but the actual contributions of the individual channels to the sensitivity predicted with the population response [Fig. 2; Eq. (4)] also depend upon the combined variance computed from the responses of all the included AN channels, i.e., the second term in the denominator of Eq. (4).) In general, Fig. 3 illustrates that sensitivity decreased quickly as the noise level increased, although there was a smaller difference between the sensitivities in quiet and low-level noise.

The drop in sensitivity for model responses based on rate at CFs around 1700 Hz [Figs. 3(a) and 3(d)] was due to the greater sensitivity of AN rates to changes in energy on the skirts of their tuning curves as compared to the peaks, where rates begin to saturate at mid to high stimulus levels (e.g., Whitfield, 1967; Siebert, 1968). The response rates for CFs tuned near the formant frequency are relatively high, as expected [Figs. 4(a) and 4(c)]; however, the *changes* in rate for fibers tuned near the formant peak as formant frequency is varied are relatively small [Figs. 4(b) and 4(d)]. The largest changes in rate for small changes in formant frequency occur for fibers tuned 100–200 Hz below or above the formant frequency [Figs. 4(b) and 4(d)]; therefore, these fibers contribute the most average-rate information [Figs. 3(a) and 3(d)]. The rates near the formant frequency are not com-

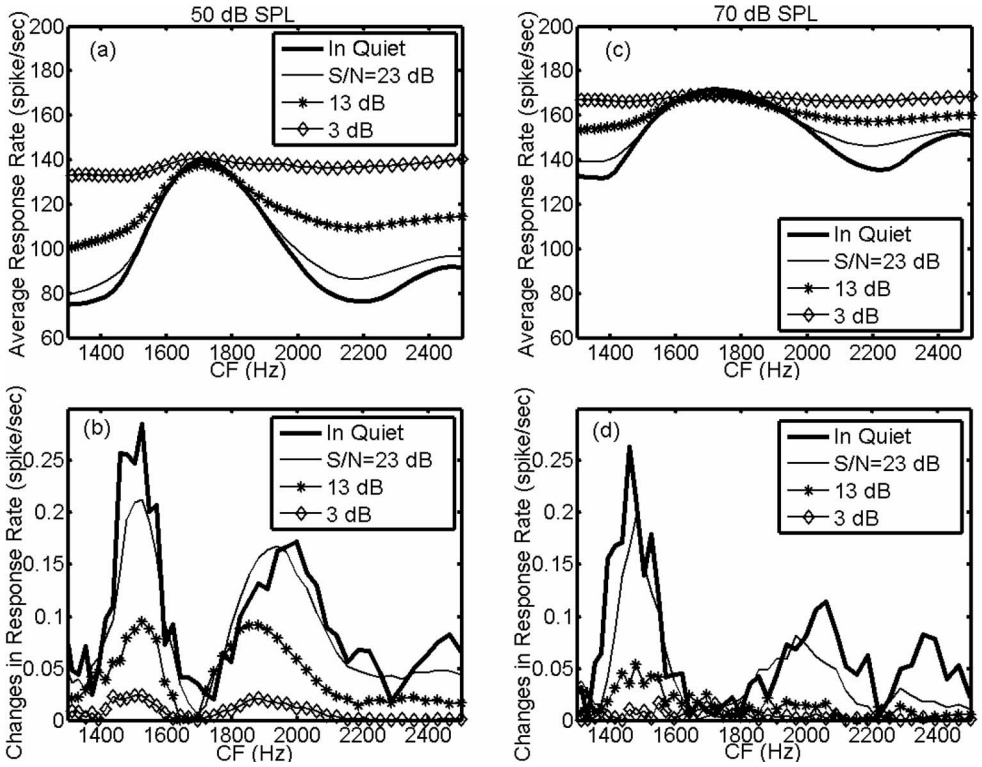


FIG. 4. Average model AN response rates [(a) and (c)] and changes in these rates [(b) and (d)] at SNRs of 3 dB (diamonds), 13 dB (asterisks), 23 dB (thin line), and in quiet (bold line) with fixed vowel levels of 50 dB SPL [(a) and (b)] and 70 dB SPL [(c) and (d)]. Changes in average AN response rates are for a 1-Hz increase (from 1700 to 1701 Hz) in the second formant frequency.

pletely saturated (i.e., rates for CFs near the formant frequency are higher in response to the 70-dB SPL vowel than in response to the 50-dB SPL vowel); however, the decreasing slope of AN rate-level functions as level increases toward saturation means that the sensitivity of these fibers decreases as level increases. In addition to this effect of saturation on fibers stimulated by the higher levels near formant peaks, the higher rates of these fibers are associated with higher variance, which also limits the sensitivity of the average discharge rate (Colburn *et al.*, 2003).

The sensitivity of each AN model fiber when using both average rate and timing information at each signal-to-noise ratio is shown in Figs. 3(b) and 3(e) and is found by computing the sensitivity using Eq. (4) for each CF channel. The sensitivity of the AN model fibers near F2 (1700 Hz) was always higher than that of other model fibers at each noise level. The sensitivity curve was wider in quiet (bold line) or in low-level noise (thin line) than in high-level noise (diamonds). The change in width of the relative sensitivity plots is made clearer by showing results normalized to the peak sensitivity [Figs. 3(c) and 3(f)]. The responses of the model AN fibers with CFs away from F2 were increasingly dominated by the masker noise as noise level increased [see Figs. 4(a) and 4(c)]. This result is consistent with the suggestion that when the whole population of AN model fibers was used in the predictions, more fibers contributed significantly in quiet or at low noise levels than at high noise levels (Miller *et al.*, 1987).

Figure 5 shows predicted thresholds based on the coincidence-detection mechanism [Eq. (7)]. If coincidence detection extracts the timing information that is relevant for this formant-frequency discrimination task, then the rate (or counts) of coincidence detectors should provide better predictions of the trends in the thresholds than do the rates of

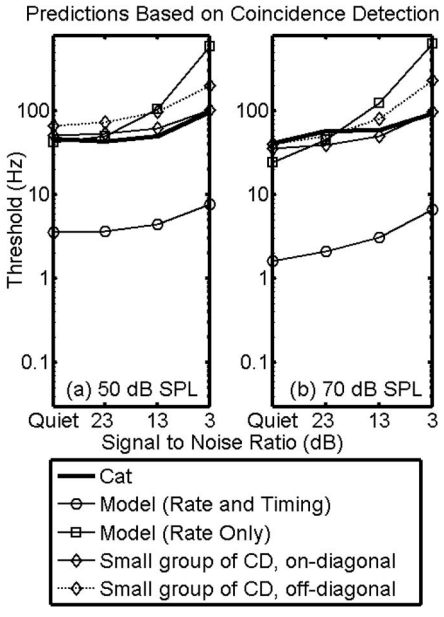


FIG. 5. Thresholds for formant-frequency discrimination based on the coincidence-detection mechanism [Eq. (7)] at various SNRs. Stimulus (combination of vowel and noise) levels were randomly varied (roved) with level uniformly distributed over a ±5-dB range, in increments of 1 dB. Average vowel levels were (a) 50 and (b) 70 dB SPL rms. Cat performance is shown with thick lines [Hienz *et al.* (1998), open bars of their Fig. 3]. Model threshold predictions used only rate information for the population of coincidence detectors that received inputs from model AN fibers with CFs between 1200 and 2500 Hz (squares), used both rate and timing information for the same population of coincidence detectors (circles), and used both rate and timing information of small groups of coincidence detectors [on- and off-diagonal, see Fig. 6(a)] (diamonds).

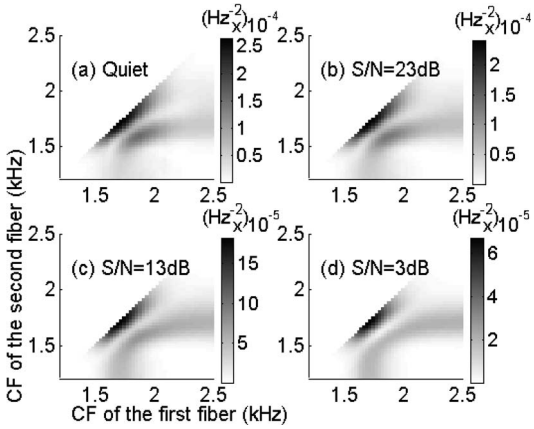


FIG. 6. Sensitivities of coincidence detectors to change in the second formant frequency. Roving-level vowel and masker levels were uniformly distributed over a ±5-dB range, with 1-dB increments and an average level of 50 dB SPL rms. Each panel corresponds to one SNR: (a) quiet, (b) 23 dB, (c) 13 dB, and (d) 3 dB.

AN fibers. The coincidence-detection threshold predictions based on count (Fig. 5, squares), however, are very similar to those based on AN rate (Fig. 2), suggesting that the coincidence detectors are not using the temporal information in the AN population (i.e., converting AN timing information into coincidence detector counts).

The trend in predicted thresholds obtained using both count and timing information of a large population of coincidence detectors (Fig. 5, circles) is nearly flat for quiet, low, and medium background noise levels, and the threshold at the high noise level is about two times the threshold at the medium noise level. This trend agrees with the psychophysical data, which simply suggests that the coincidence detectors preserve timing information present in the AN response. Thus, the timing information that contributes to the predictions based on AN responses is not lost in the process of coincidence detection, but it is apparently not "used" by the coincidence detectors, either (in which case it would be reflected in the coincidence-detector counts).

Thresholds were also predicted using the rate and timing information of subsets of the coincidence-detector population. The selection of the smaller groups of coincidence detectors was based on the results shown in Fig. 6, which demonstrates the sensitivity of each coincidence detector to the formant-frequency change. In general, each panel of Fig. 6 (corresponding to one signal-to-noise ratio) depicts two groups of coincidence detectors with relatively high sensitivity. The first group is the one on the diagonal, which represents coincidence detectors that receive inputs from two AN fibers that have matched CFs. The other group is away from the diagonal, representing coincidence detectors that receive inputs from two AN fibers with different CFs, and it has lower sensitivity than the first group. The coincidence detector with the highest sensitivity in each of these two groups was used to compute the formant-frequency discrimination thresholds. In Fig. 5, the solid line with diamonds corresponds to the performance of the most sensitive coincidence detector from the first group (on-diagonal, with matched AN CFs of 1745 Hz), and the dashed line with diamonds corresponds to the performance of the coincidence-detector from the second group (off-diagonal, convergence of ANs with CFs of 1595 and 1881 Hz). The performance of each of these coincidence detectors, based on both count and timing information, shows a trend that is similar to the cat performance and to the prediction with the original population of coincidence detectors (line with circles in Fig. 5). As shown above for the AN model predictions, these results suggest that a subset of the model coincidence-detector cells can predict the trends in the behavioral performance; however, the fact that both rate and timing of the coincidence detectors had to be included again illustrates that the coincidence detectors were not extracting the critical temporal information from the AN responses. Predictions that used only the rate of a subset of the coincidence detectors (not shown) would have been even higher than the rate-based predictions for the whole population of model coincidence detectors, and therefore higher than the behavioral thresholds, especially for the lower SNRs.

IV. DISCUSSION

This study provides predictions of performance limits in formant-frequency discrimination that used the response patterns of AN model fibers and responses of cross-frequency coincidence detectors. Quantitative tools developed to evaluate coding strategies for simple acoustic stimuli, such as tones and tones in noise (Heinz, 2000), were applied here to complex stimuli. The results of this study provide insight into the relative contributions of rate and timing information to speech coding in noise. When the background noise level increased, the thresholds predicted by using average-rate information of model AN fibers increased more than behavioral thresholds did, whereas model thresholds based on both rate and timing information increased in a manner more consistent with the behavioral data (Fig. 2). This result suggests that timing information was required to explain trends in psychophysical performance in this task.

A prominent feature of the AN model sensitivity patterns based on rate information was the dip at CFs near 1700 Hz [Figs. 3(a) and 3(d)], in contrast to the simple peak at 1700 Hz in the sensitivity pattern based on temporal and rate information [Figs. 3(b) and 3(e)]. The dip in the rate-based sensitivity was further explored by examining the average-rate responses (Fig. 4). The dominant factor in the decrease in sensitivity for fibers tuned near the formant frequency was the reduced *change* in rate for different formant frequencies for these fibers due to rate saturation. The model fiber sensitivity to the formant-frequency change was significantly reduced by high-level background noise [Figs. 3(a), 3(d), 4(b), and 4(d)] when predictions were based on only average-rate information (Young and Barta, 1986). Thus, as has been seen in previous studies (e.g., Heinz *et al.*, 2001a, b; Colburn *et al.*, 2003), the sensitivity pattern across a population of AN fibers is typically more complex for rate-based information [Figs. 3(a) and 3(d)] than it is for temporally based information [Figs. 3(b) and 3(e)].

For the predictions based on both rate and timing information, the sensitivity of AN model fibers as a function of

CF [Figs. 3(b) and 3(d)] indicated that the model fibers with CFs near but not too close to the second formant frequency may have contributed to the overall sensitivity at low noise levels, yet they contributed relatively less when the noise level was high because their response was overwhelmed by the high-level noise. Because of this behavior, the prediction based on a large population of model fibers degraded more for high-level noise than did the prediction based on a small subset of fibers [Figs. 3(c) and 3(f)]. As suggested by our previous results and by others, it is realistic and reasonable to predict psychophysical thresholds with a small population of AN model fibers with CFs close to the signal frequency when the task involves discrimination of spectral changes over a small frequency range (Young and Sachs, 1979; Sruloviczand Goldstein, 1983; Delgutte and Kiang, 1984a, b; Miller *et al.*, 1987; Tan and Carney, 2005).

The coincidence-detection mechanism did not specifically extract the temporal cues from the AN model responses related to the formant-frequency discrimination task. If it had, then the coincidence detection predictions based on rate (or count) would have explained the behavioral results as well as the AN predictions based on both rate and timing. In fact, the threshold trend of the coincidence-detector count prediction degraded more with decreasing SNR than did the prediction based on both rate and timing information in AN responses, and it was similar to the threshold trend obtained using only the average rate of the AN responses. Future studies must therefore consider other temporal aspects of the AN responses, such as information conveyed by phase locking to the envelope of the stimulus after narrow-band filtering in the auditory periphery (Tan and Carney, 2005).

An interesting question regarding these results concerns the change in the predicted thresholds, and thus in the model AN responses, at the highest noise level, that is, why does the threshold based on both temporal and rate measures increase noticeably at the highest noise level, especially as compared to the threshold based on rate only, which changes more gradually across all noise levels? Synchrony information is generally robust across a wide range of sound levels, even in response to signals in the presence of noise (Rhode *et al.*, 1978). However, the temporal responses related to the envelope of the vowel complex would be expected to change as noise level increases, because the bandwidth of the peripheral filters increases with sound level and because adaptation shapes the responses of AN fibers. The degree of modulation in response to the signal in the AN responses is reduced as level increases due to rate saturation (Smith and Brachman, 1980; Joris and Yin, 1992), and filter broadening due to the added noise tends to further reduce the largest envelope modulations in the AN responses [i.e., wideband noises have a broader amplitude-modulation spectra, with lower amplitudes at low frequencies, as compared with narrow-band noises, which have narrower and higher amplitude-modulation spectra, especially at low modulation frequencies (Lawson and Uhlenbeck, 1950; Dau *et al.*, 1997)]. The representation of modulation cues and the influence of masking noise on these cues in AN responses to speech sounds are topics for future study.

At every SNR, the threshold predicted by the coincidence-detection mechanism was higher than the prediction based on the response patterns of the AN model fibers, as expected, because the coincidence-detection process decreased the total amount of information carried by the AN model fiber response patterns. Similarly, for both the AN model fiber and the coincidence-detector predictions, thresholds based on smaller populations were always higher than the thresholds based on the full set of model fibers or coincidence detectors. This increase in thresholds was simply due to the fact that the amount of information is related to the number of fibers or cells included in the population response [Eqs. (1), (4), and (6)].

The work presented here was all based on formant-frequency discrimination of F2 in synthesized speech stimuli. It would be interesting to test predictions for lower (first formant) or higher (third formant) frequencies. To extend the study to higher frequencies, it would be important to include nonsaturating low-spontaneous-rate AN fibers in the simulations, because temporal information conveyed by the synchronization of AN fibers rolls off at high frequencies (Johnson, 1980). The distinct roles of rate and temporal coding in these frequency ranges and the diverse contributions of different spontaneous-rate populations of AN fibers (e.g., Heinz *et al.*, 2001b) suggest that different aspects of the AN model responses may dominate threshold predictions for speech cues in different frequency ranges.

ACKNOWLEDGMENTS

This work was supported by NIH-NIDCD Grant No. R01-01641. We gratefully acknowledge the comments of Sean Davidson, Michael Heinz, Paul Nelson, and two very helpful reviewers on a previous version of this manuscript and the editorial assistance of Susan Early.

[1]Computer code used for the simulations presented here is available at http://web.syr.edu/~lacarney

[2]Cat thresholds illustrated in the figures for comparison to model thresholds were all based on Hienz *et al.*'s (1998) results for cats without olivocochlear bundle (OCB) lesions. The AN model (Tan and Carney, 2003) was based on AN recordings made in cats without OCB lesions. The role of the OCB (or lesions of the OCB) in AN responses to speech, in either awake or anesthetized animals or in behavioral responses to speech, is still not completely understood; therefore, it seemed most straightforward to focus comparisons on results for intact animals.

Carney, L. H., Heinz, M. G., Evilsizer, M. E., Gilkey, R. H., and Colburn, H. S. (**2002**). "Auditory phase opponency: A temporal model for masked detection at low frequencies," Acust. Acta Acust. 88, 334–347.

Colburn, H. S. (1969). "Some physiological limitations on binaural performance," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Colburn, H. S. (1977). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. Supplementary material," AIP Document No. PAPS JASMA-61–525–98.

Colburn, H. S., Carney, L. H., and Heinz, M. G. (**2003**). "Quantifying the information in auditory-nerve responses for level discrimination," J. Assoc. Res. Otolaryngol. **04**, 294–311.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Delgutte, B., and Kiang, N. Y.-S. (**1984a**). "Speech coding in the auditory nerve: I. Vowel-like sounds," J. Acoust. Soc. Am. **75**, 866–878.

Delgutte, B., and Kiang, N. Y.-S. (**1984b**). "Speech coding in the auditory

nerve: V. Vowels in background noise" J. Acoust. Soc. Am. **75**, 908–918.
Flanagan, J. L. (**1955**). "A difference limen for vowel formant frequency," J. Acoust. Soc. Am. **27**, 613–617.
Hawks, J. W. (**1994**). "Difference limens for formant patterns of vowel sounds," J. Acoust. Soc. Am. **95**, 1074–1084.
Heinz, M. G. (2000). "Quantifying the effects of the cochlear amplifier on temporal and average-rate information in the auditory nerve," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001a**). "Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve," Neural Comput. **13**, 2273–2316.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001b**). "Rate and timing cues associated with the cochlear amplifier: Level discrimination based on monaural cross-frequency coincidence detection," J. Acoust. Soc. Am. **110**, 2065–2084.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2002**). "Quantifying the implications of nonlinear cochlear tuning for auditory-filter estimates," J. Acoust. Soc. Am. **111**, 996–1011.
Hienz, R. D., Aleszczyk, C. M., and May, B. J. (**1996**). "Vowel discrimination in cats: Acquisition, effects of stimulus level, and performance in noise," J. Acoust. Soc. Am. **99**, 3656–3668.
Hienz, R. D., Stiles, P., and May, B. J. (**1998**). "Effects of bilateral olivocochlear lesions on vowel formant discrimination in cats," Hear. Res. **116**, 10–20.
Johnson, D. H. (**1980**). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," J. Acoust. Soc. Am. **68**, 1115–1122.
Joris, P. X., and Yin, T. C. T. (**1992**). "Responses to amplitude-modulated tones in the auditory nerve of the cat," J. Acoust. Soc. Am. **91**, 215–232.
Keithley, E. M., and Schreiber, R. C. (**1987**). "Frequency map of the spiral ganglion of cat," J. Acoust. Soc. Am. **81**, 1036–1042.
Kewley-Port, D., and Watson, C. S. (**1994**). "Formant-frequency discrimination for isolated English vowels," J. Acoust. Soc. Am. **95**, 485–496.
Kewley-Port, D., Li, X., Zheng, Y., and Neel, A. (**1996**). "Fundamental frequency effects on thresholds for vowel formant discrimination," J. Acoust. Soc. Am. **100**, 2462–2470.
Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.
Lawson, J. L., and Uhlenbeck, G. E. (**1950**). *Threshold Signals*, Vol. **24** of *Radiation Laboratory Series* (McGraw-Hill, New York).
Liberman, M. C. (**1978**). "Auditory-nerve response fromcats raised in a low-noise chamber," J. Acoust. Soc. Am. **63**, 442–455.
Liberman, M. C. (**1982**). "The cochlear frequency map for cat: Labeling auditory-nerve fibers of known characteristic frequency," J. Acoust. Soc. Am. **72**, 1441–1449.
Liu, C., and Kewley-Port, D. (**2004**). "Formant discrimination in noise for isolated vowels," J. Acoust. Soc. Am. **116**, 3119–3129.
Lyzenga, J., and Horst, J. W. (**1995**). "Frequency discrimination of bandlimited harmonic complexes related to vowel formants," J. Acoust. Soc. Am. **98**, 1943–1955.
Mermelstein, P. (**1978**). "Difference limens for formant frequencies of steady-state and consonant-bound vowels," J. Acoust. Soc. Am. **63**, 572–580.
Miller, M. I., Barta, P. E., and Sachs, M. B. (**1987**). "Strategies for the representation of a tone in background noise in the temporal aspects of the discharge patterns of auditory-nerve fibers," J. Acoust. Soc. Am. **81**, 665–679.
Rhode, W. S., Geisler, C. D., and Kennedy, D. K. (**1978**). "Auditory-nerve fiber responses to wide-band noise and tone combinations," J. Neurophysiol. **41**, 692–704.
Sachs, M. B., and Young, E. D. (**1979**). "Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate," J. Acoust. Soc. Am. **66**, 470–479.
Sachs, M. B., Voigt, H. F., and Young, E. D. (**1983**). "Auditory-nerve representation of vowels in background noise," J. Neurophysiol. **50**, 27–45.
Siebert, W. M. (**1965**). "Some implication of the stochastic behavior of primary auditory neurons," Kybernetika **2**, 206–215.
Siebert, W. M. (**1968**). "Stimulus transformation in the peripheral auditory system," in *Recognizing Patterns*, edited by P.A. Kolers and M. Eden (MIT, Cambridge, MA), pp. 104–133.
Siebert, W. M. (**1970**). "Frequency discrimination in the auditory system: place or periodicity mechanisms?" Proc. IEEE **58**, 723–730.
Smith, R. L., and Brachman, M. L. (**1980**). "Response modulation of auditory-nerve fibers by AM stimuli: Effects of average intensity," Hear. Res. **2**, 123–133.
Srulovicz, P., and Goldstein, J. L. (**1983**). "The central spectrum: A synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum," J. Acoust. Soc. Am. **73**, 1266–1276.
Tan, Q. (2003). "Computational and statistical analysis of auditory peripheral processing for vowel-like signals," Ph.D. dissertation, Boston University.
Tan, Q., and Carney, L. H. (**2003**). "A phenomenological model for the responses of auditory-nerve fibers: II. Nonlinear tuning with a frequency glide," J. Acoust. Soc. Am. **114**, 2007–2020.
Tan, Q., and Carney, L. H. (**2005**). "Encoding of vowel-like sounds in the auditory nerve: Model predictions of discrimination performance," J. Acoust. Soc. Am. **117**, 1210–1222.
Whitfield, I. C. (**1967**). *The Auditory Pathway* (Edward Arnold, London).
Wier, C. C., Jesteadt, W., and Green, D. M. (**1977**). "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 178–184.
Young, E. D., and Barta, P. E. (**1986**). "Rate responses of auditory-nerve fibers in noise near masked threshold," J. Acoust. Soc. Am. **79**, 426–442.
Young, E. D., and Sachs, M. B. (**1979**). "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers," J. Acoust. Soc. Am. **66**, 1381–1403.

Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery

Muhammad S. A. Zilany and Ian C. Bruce[a)]
Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, L8S 4K1, Canada

(Received 17 February 2006; revised 5 June 2006; accepted 19 June 2006)

This paper presents a computational model to simulate normal and impaired auditory-nerve (AN) fiber responses in cats. The model responses match physiological data over a wider dynamic range than previous auditory models. This is achieved by providing two modes of basilar membrane excitation to the inner hair cell (IHC) rather than one. The two modes are generated by two parallel filters, component 1 (C1) and component 2 (C2), and the outputs are subsequently transduced by two separate functions. The responses are then added and passed through the IHC low-pass filter followed by the IHC-AN synapse model and discharge generator. The C1 filter is a narrow-band, chirp filter with the gain and bandwidth controlled by a nonlinear feed-forward control path. This filter is responsible for low and moderate level responses. A linear, static, and broadly tuned C2 filter followed by a nonlinear, inverted and nonrectifying C2 transduction function is critical for producing transition region and high-level effects. Consistent with Kiang's two-factor cancellation hypothesis, the interaction between the two paths produces effects such as the C1/C2 transition and peak splitting in the period histogram. The model responses are consistent with a wide range of physiological data from both normal and impaired ears for stimuli presented at levels spanning the dynamic range of hearing. [DOI: 10.1121/1.2225512]

PACS number(s): 43.64.Bt, 43.64.Pg, 43.64.Wn [WPS] Pages: 1446–1466

I. INTRODUCTION

Modeling auditory-nerve (AN) fiber responses provides a useful tool for testing our understanding of the underlying mechanical and physiological processes in the auditory periphery. In this regard, several efforts had been made to develop computational models (Deng and Geisler, 1987; Goldstein, 1990; Kates, 1991; Carney, 1993; Kates, 1995; Goldstein, 1995; Robert and Eriksson, 1999; Zhang *et al.*, 2001; Meddis *et al.*, 2001; Bruce *et al.*, 2003; Sumner *et al.*, 2003; Tan and Carney, 2003) that integrate data and theories from a wide range of research in the cochlea. Describing and understanding the mechanisms behind the nonlinearities in the cochlea such as compression, two-tone rate suppression, etc., was the main focus of these modeling efforts. In addition, these models can be used in studying how simple and complex sounds are encoded in the peripheral auditory system (e.g., Carney and Yin, 1988; Carney, 1994), in ways that might not be feasible in physiological experiments. Broader applications of these models include using them as front-ends in many research areas such as speech recognition in noisy environments (e.g., Ghitza, 1988; Tchorz and Kollmeier, 1999), computational modeling of auditory scene analysis (e.g., Brown and Cooke, 1994), modeling of neural circuits in the auditory brain-stem (e.g., Hewitt and Meddis, 1992), design of hearing-aid amplification schemes (e.g., Miller *et al.*, 1999; Bruce, 2004; Bondy *et al.*, 2004), and design of speech processors for cochlear implants (e.g., Wilson *et al.*, 2005).

In this study, a phenomenological model of the auditory periphery developed by Bruce *et al.* (2003) to simulate the low- and moderate-level responses in normal and impaired ears has been modified and extended to also account for the high-level responses of AN fibers across a wide range of characteristic frequencies (CFs). The motivation for this work is to provide a more accurate (quantitative and qualitative) description of the response properties of AN fibers to simple and complex stimuli for acoustic stimulus sound pressure levels spanning the dynamic range of hearing. As hearing aids amplify signals to compensate for hearing loss, accurately modeling high-level effects will be useful in suggesting and testing new strategies for hearing-aid signal-processing schemes.

Understanding the complex changes in responses of the mammalian AN fibers to high-level stimuli has been attempted by several researchers over the last few decades. One of the most striking nonmonotonic behaviors in AN responses to tones at high levels is the component 1-component 2 (C1/C2) transition (Kiang *et al.*, 1965, 1969; Kiang and Moxon, 1972; Gifford and Guinan, 1983; Wong *et al.*, 1998; Heinz and Young, 2004), which is characterized by a sharp transition of up to 180° in the phase-level function and sometimes also by a dip in the rate-level function (Kiang, 1984; Liberman and Kiang, 1984). The low-level response that occurs before the phase shift is referred to as the C1 response, and C2 is the high-level response that appears after the phase shift. Kiang (1990) offered a two-factor cancellation hypothesis to explain the rate dip and phase shift at this C1/C2 transition. According to his hypothesis, the C1 and C2 responses are caused by two different response generating factors. One of them, C1, dominates at low and mod-

[a)]Electronic mail: ibruce@ieee.org

erate levels, is narrowly tuned, and is vulnerable to trauma. The other, C2, dominates at high levels, is broadly tuned, is resistant to trauma, and is shifted in phase by 180° with respect to the C1 response. At the level where both are approximately equal in amplitude, they tend to cancel each other and cause the C1/C2 transition phenomenon to occur in AN responses.

The earliest physical interpretation of the C1/C2 transition phenomenon was that the C1 response is normally generated by the interaction between the outer hair cells (OHCs) and the tallest row of inner hair cell (IHC) stereocilia, while C2 requires only the shorter rows of IHC stereocilia (Liberman and Kiang, 1984). A selective loss of tall stereocilia from the IHC reduces the C1 without affecting the C2 responses, and in fact, total loss of tall IHC stereocilia (which is always associated with severe damage to the OHCs) completely eliminates the C1 response (Fig. 7, Liberman and Kiang, 1984). Liberman and Kiang (1984) argued that if the tall IHC stereocilia are coupled to the tectorial membrane whereas the shorter rows are not, total loss of the tall IHC stereocilia removes C1 from the AN responses because of the elimination of mechanical coupling between the OHCs and the tall IHC stereocilia. Liberman and Kiang (1984) observed that a normal C2 response requires neither functioning tall stereocilia on the IHCs nor functioning stereocilia on the OHCs. A functional relationship between the C2 and the short IHC stereocilia has been suggested by the observation that both are resistant to trauma (Liberman and Kiang, 1984).

Sewell (1984b) also argued that there could be two transduction mechanisms in the IHCs, based on AN fiber recordings during which furosemide was administered to cats. Application of furosemide, which is known to cause decreases in the endocochlear potential (EP), reduced the discharge rate of the C1 component, whereas the C2 component was largely unaffected and is thus likely be generated by a different transduction mechanism that does not require a large EP (Sewell, 1984b). Kiang *et al.* (1986) utilized these results in reassessing the conceptual physical model of Liberman and Kiang (1984), but still no definitive conclusion was reached regarding the physical source of the C2 component. Cheatham and Dallos (1998) showed that the notch and phase shift observed in the AN fiber responses has a correlate in the measurement of IHC receptor potentials, whereas measurements from the organ of Corti or OHCs do not demonstrate this phenomenon. This would indicate that the C1/C2 interactions may not be present in the basilar membrane (BM) vibrations. A review by Robles and Ruggero (2001) notes that, while notches and phase shifts are observed in some BM vibration data, they are not consistently found in all BM recordings. However, there is strong evidence for two or more modes of BM vibration and/or excitation of the IHCs at high sound pressure levels (e.g., Robles and Ruggero, 2001; Lin and Guinan, 2000, 2004). Combining this observation with the evidence for two separate transduction functions, we suggest in this paper that the IHC could be subjected to two modes of excitation from the BM with the frequency tuning and phase response corresponding to C1 and C2, and that each of these modes of excitation has its own transduction function, such that the C1/C2 interaction occurs within the IHC.

A second high-level effect is the emergence of harmonics in the period histogram in response to tones. For frequencies below 5–7 kHz, spikes tend to be time-locked to a particular phase of the input tone stimulus waveform (Kiang *et al.*, 1965; Johnson, 1980; Kiang, 1990). Consequently, the period histogram of responses to low-frequency tones exhibits a single peak at low levels. However, at high stimulus levels, the histograms of responses to tones below 1 kHz may show two or even three peaks per stimulus cycle, which is referred to as peak splitting (Kiang and Moxon, 1972; Johnson, 1980; Ruggero and Rich, 1983; Kiang, 1984; Ruggero and Rich, 1989; Kiang, 1990; Cai and Geisler, 1996). With further increases in the tone level, the histograms again show only a single peak, but now phase-shifted from the original peak at low stimulus level. Kiang (1990) postulated that if one of the two responses (C1 or C2) contains a second harmonic distortion, peak splitting can be qualitatively explained by his two-factor cancellation hypothesis. In response to a tone at the transition level, the two fundamental components cancel each other as the responses are out of phase; however, the uncancelled second harmonic component becomes significant within a few dB of that transition level. Since AN fibers only phase-lock to the positive portion of the cycle of a tone, the period histogram shows two peaks corresponding to the positive phases of the harmonic. Consistent with this, in this paper we show quantitatively that the peak-splitting phenomenon can be explained directly by the two-factor cancellation hypothesis if the C2 waveform is not subject to rectification by the IHC transduction function, in contrast to the strong rectification of the C1 waveform by its IHC transduction function.

Moderate and high sound pressure level effects in the C1 response itself include compression, two-tone rate suppression, and shifts in best frequency (BF).[1] All of theses effects are thought to arise from a single nonlinear, time-varying mechanism in the cochlea, referred to as the cochlear amplifier (CA) (Patuzzi and Robertson, 1988; Patuzzi, 1996; Holley, 1996). The compressive nonlinearity of the cochlea is strongly dependent on stimulus frequency, as is evident from data showing BM gain as a function of frequency (Ruggero *et al.*, 1997; Cooper and Rhode, 1997). At frequencies near CF, the gain grows systematically larger as a function of decreasing stimulus level; however, there is no change in the gain at frequencies well removed from CF. As a result, both the sharpness of tuning and the peak (or BF) change as a function of stimulus level. At the highest stimulus levels, sharpness of tuning significantly decreases and responses in the base of the cochlea reach their maxima at frequencies about one-half octave lower than at the lowest stimulus levels (Robles and Ruggero, 2001). Responses of AN fibers with CFs ranging from mid to high frequencies show similar shifts in BF to lower frequencies as a function of increasing sound level (Anderson *et al.*, 1971; Møller, 1977). This level-dependent BF shift relates to the interaction between the change in the shape of the impulse-response envelope with sound pressure level and the trend of the instantaneous-

frequency (IF) glide in the impulse response (Carney *et al.*, 1999; Tan and Carney, 2003). However, frequency glides themselves are nearly independent of stimulus sound pressure level over the full dynamic range of hearing (e.g., de Boer and Nuttall, 1997; Recio *et al.*, 1998; Carney *et al.*, 1999) and even maintain their general form postmortem (e.g., Recio *et al.*, 1998). Although the glides are largely invariant with changes in stimulus level, they are strongly dependent on CF (Carney and Yin, 1988; Carney *et al.*, 1999; Shera, 2001). The goal of this study is to incorporate all of these response properties in the model, including high-level effects such as the C1/C2 transition, peak splitting, and BF shift.

Toward this goal, several efforts have been made to model the nonlinearities of the cochlea that can be characterized in several ways such as transmission-line models (e.g., Giguère and Woodland, 1994), mechanical cochlear models (e.g., de Boer and Viergever, 1982), models using parallel filter paths with static nonlinearities (e.g., Goldstein, 1995; Lin and Goldstein, 1995; Meddis *et al.*, 2001), and models using time-varying nonlinear filters (e.g., Carney, 1993; Zhang *et al.*, 2001; Bruce *et al.*, 2003)—for a comparison of these approaches see the recent review by Lopez-Poveda (2005). The phenomenological approaches of (i) parallel filter paths with static nonlinearities and (ii) time-varying nonlinear filters lend themselves most readily to incorporation into models of the entire auditory periphery because of their relative simplicity and their computational efficiency. Both of these approaches are able to describe most of the cochlear nonlinearities observed in physiological data, although the time-varying nonlinear filter approach appears to better explain stimulus dependent changes in the phase of BM vibrations and AN fiber responses. For explaining high-level nonlinearities such as the C1/C2 transition and peak splitting, only Lin and Goldstein (1995) conducted a detailed investigation with a computational model that incorporates the two-factor hypothesis. This model is of the type with parallel filter paths and static nonlinearities, and consequently has two pathways that can interact to produce effects like the C1/C2 transition and peak splitting. However, because their model only incorporates static nonlinearities, Lin and Goldstein (1995) found it necessary to introduce an *ad hoc* stimulus-dependent phase adaptation mechanism to produce the correct stimulus-dependent phase changes. Additionally, the C1/C2 phenomenon occurs in the model BM response of Lin and Goldstein (1995) before IHC transduction, which may not be consistent with all of the physiological data summarized earlier.

In this paper, we propose that the two excitatory inputs to the IHC are provided by the outputs of two separate filters, C1 and C2. The output of the C1 filter closely resembles the primary mode of vibration of the BM which is dependent on the action of the cochlear amplifier. The physiological correlate of the output of the C2 filter is not clear, but we have found that the model is capable of generating C2 threshold tuning curves similar to those of Liberman and Kiang (1984) if the C2 filter is set to be the same as the C1 filter with complete OHC impairment, i.e., representing the passive BM response rather than the active BM response resulting from the action of the cochlear amplifier. From the above-noted arguments, we also propose the existence of two separate transduction functions following the two filters, and the outputs of these two transduction functions interact to produce the C1/C2 transition in the model. We have implemented this conceptual model by adding a parallel-path C2 filter to the computational auditory periphery model of Bruce *et al.* (2003). The 180° phase shift of the C2 response is incorporated into the C2 transduction function, which only receives excitation from the C2 filter. The C1/C2 interaction consequently occurs within the IHC, rather than in the BM response. An advantage of this approach is, in addition to appearing more consistent with the physiological data, that the C2 filter and transduction function properties can be adjusted to explain the C1/C2 transition and peak-splitting data, without having any effect on C1 responses at low to moderate sound pressure levels. An important consequence of this arrangement is that the low-frequency tail of the model tuning curves is still produced by the C1 filter, not the C2 filter, in contrast to the class of models using parallel filters with static nonlinearities.

The C1 filter of the Bruce *et al.* (2003) model is a gammatone filter that does not produce the frequency glides and BF shifts observed in the physiological data. Therefore, in this paper the C1 filter has been replaced with a chirp filter (Tan and Carney, 2003) that is able to produce these effects. We have made some improvements over the chirp filter proposed in Tan and Carney (2003). First, the order of the filter is reduced to half of that proposed in Tan and Carney (2003) to produce more realistic frequency tuning curves, including tails in both normal and impaired conditions. Second, during implementation the zeros of the filter are moved along with the poles according to the control signal to give more realistic level-independent frequency glides in the impulse response and to improve the simulation of the phase properties of the cochlea. A comparison with recent time-varying nonlinear filter type models is provided in Table I. The computational details of the proposed model are described in Sec. II, and Sec. III evaluates the model against published AN responses.

II. DESCRIPTION OF THE MODEL

A schematic diagram of the proposed model is given in Fig. 1. Each section of the model provides a phenomenological description of the major functional components of the auditory-periphery, from the middle-ear to the auditory nerve.

The input to the middle-ear (ME) is the instantaneous pressure waveform of the stimulus in units of Pa, sampled at 500 kHz. The ME filter is followed by three parallel filter paths: the C1 and C2 filters in the signal path and the broadband filter in the control path. The feed-forward control path regulates the gain and bandwidth of the C1 filter to account for several level-dependent properties in the cochlea. The C1 filter has been designed with an asymmetrical orientation of the poles and zeros (different damping coefficients) in the complex plane to produce IF glides in the impulse response of the C1 filter. The parallel-path C2 filter has been imple-

TABLE I. Comparison among several recent models.

	Zhang *et al.* (2001)	Bruce *et al.* (2003)	Tan and Carney (2003)	Proposed model
Middle-ear filtering	No	Yes	Yes	Yes
Signal-path filter	Gammatone fourth order	Gammatone fourth order	Chirp filter twentieth order	Chirp filter tenth order
High-level effects:				
(i) Broadened tuning	Yes	Yes	Yes	Yes
(ii) C1/C2 transition	No	No	No	Yes
(iii) Peak splitting	No	No	No	Yes
(iv) BF shift with level	No	No	Yes	Yes

mented based on Kiang's two-factor cancellation hypothesis. The C2 filter is linear and similar to the C1 filter with its broadest possible tuning, i.e., the C2 filter is the completely OHC-impaired version of the C1 filter. The transduction function following the C2 filter is such that at low levels its output is significantly lower than the output of the corresponding C1 response, whereas at high sound pressure levels the C2 output dominates and the C1 and C2 outputs are out of phase. At levels within the transition region, both outputs are approximately equal and tend to cancel each other. In addition, the C2 response is not subject to rectification, unlike the C1 response at high levels, such that the peak-splitting phenomenon also results from the C1/C2 interaction. The combined response of the two transduction functions following the C1 and C2 filters is passed through a seventh-order IHC low-pass filter. The IHC output drives the IHC-AN synapse which provides the instantaneous synaptic release rate as output. Finally the discharge times are produced by a renewal process that includes refractory effects.

A. Middle-ear filter

The ME section of the model is simplified from that of Bruce *et al.* (2003), which was created by combining the ME model of Peake *et al.* (1992) with that of Matthews (1983) and deriving a digital-filter implementation of the resulting continuous-time transfer function. Because of difficulties in ensuring stability with the previous implementation, here the continuous-time transfer function is reduced from an eleventh-order system to a fifth-order system. A fifth-order digital filter is created using the bilinear transformation for a sampling frequency of 500 kHz with the frequency axis prewarped to give a matching frequency response at 1 kHz. To ensure stability, this digital filter is implemented in a second-order system form by cascading the following digital filters:

$$\mathrm{ME}_1(z) = 0.0127\left(\frac{1.0000 + 1.0000z^{-1}}{1.0000 - 0.9986z^{-1}}\right), \tag{1}$$

$$\mathrm{ME}_2(z) = \frac{1.0000 - 1.9998z^{-1} + 0.9998z^{-2}}{1.0000 - 1.9777z^{-1} + 0.9781z^{-2}}, \tag{2}$$

$$\mathrm{ME}_3(z) = \frac{1.0000 - 1.9943z^{-1} + 0.9973z^{-2}}{1.0000 - 1.9856z^{-1} + 0.9892z^{-2}}. \tag{3}$$

B. Feed-forward control path

The purpose of the control path is to reflect the active processes in the cochlea. The gain and bandwidth of the C1 filter are varied according to the control signal to account for several level-dependent response properties of the cochlea. The control path consists of several stages: (i) a time-varying third-order gammatone filter with a broader bandwidth than the signal-path C1 filter; (ii) a nonlinear function (Boltzmann function) followed by a third-order low-pass filter to control the dynamic range and the time course of compression; and (iii) a nonlinear function to convert the output of the low-pass filter to a time-varying time constant for the C1 filter. The proposed model retains the same control path from

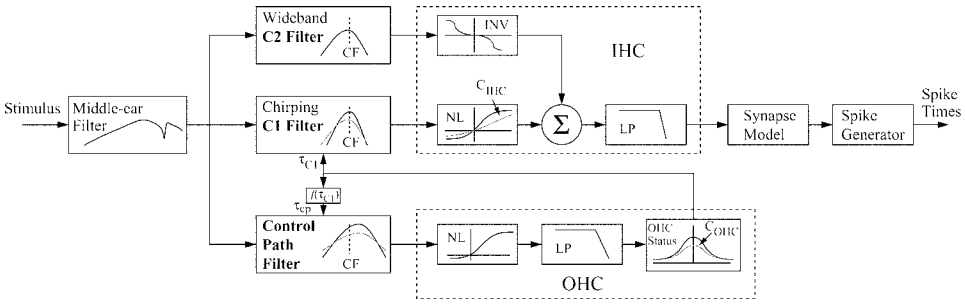


FIG. 1. Schematic diagram of the auditory-periphery model. The input to the model is an instantaneous pressure waveform of the stimulus in Pa and the output is the AN spike times in response to that input. The model includes a middle-ear filter, a feed-forward control path, a signal-path C1 filter, and a parallel-path C2 filter, the inner hair-cell (IHC) section followed by the synapse model and the discharge generator. Abbreviations: outer hair cell (OHC), low-pass (LP) filter, static nonlinearity (NL), characteristic frequency (CF), inverting nonlinearity (INV). C_{OHC} and C_{IHC} are scaling constants that control OHC and IHC status, respectively.

TABLE II. Parameter values.

Parameters	Description	Values
Middle-ear filter		Eqs. (1)–(3)
Control path		
$\mathrm{WBorder_{cp}}$	Order of the wide-band Gammatone filter	3
τ_{cpmax}	Estimated time-constant at low sound level	Eq. (4)
τ_{cpmin}	Estimated time-constant at high sound level	Eq. (5)
K_{cp}	Gain of the wide-band filter	$(\tau_{\mathrm{cp}}/\tau_{\mathrm{cpmax}})^3 \cdot 10 \times 10^3 \cdot \max(1, \mathrm{CF}/5 \times 10^3)$
$x0$	Parameter in Boltzmann function	7.6
$x1$	Parameter in Boltzmann function	5
$s0$	Parameter in Boltzmann function	12
$s1$	Parameter in Boltzmann function	5
$\mathrm{shift_{cp}}$	Parameter in Boltzmann function	0.125
$\mathrm{cutoff_{cp}}$	Cut-off frequency of control-path low-pass filter	600 Hz
$\mathrm{LPorder_{cp}}$	Order of the control-path low-pass filter	2
dc	Estimated dc shift of low-pass filter output at high-level	0.37
R_{C1}	Ratio of lower bound of τ_{C1} to τ_{C1max}	0.05
τ_{cp}	Time-constant for the control-path wide-band filter	Eq. (8)
τ_{C1}	Time-constant for the signal-path C1 filter	Eq. (8)
$\mathrm{C_{OHC}}$	Scaling constant for impairment in the OHC	$0 \leqslant \mathrm{C_{OHC}} \leqslant 1$
$\tau_{\mathrm{C1\text{-}impaired}}$	Time-constant for the C1 filter with impairment	Eq. (22)
Signal-path C1 filter		
$\mathrm{NBorder_{C1}}$	Order of the narrow-band C1 filter (chirp)	10
P_{01}	Real part of the pole nearest to the imaginary axis at low stimulus levels	Eq. (9)
P_{11}	Real part of the pole nearest to the imaginary axis at high stimulus levels	Eq. (16)
P_w	Imaginary part of the pole closest to the imaginary axis	Eq. (10)
P_a	Relative real parts of the poles	Eq. (12)
P_b	Relative imaginary parts of the poles	Eq. (11)
P_{shift}	Displacement of the real parts of the poles due to control signal	Eq. (14)
Z_L	Locations of the zeros on the real axis at low stimulus levels	Eq. (13)
Z_H	Locations of the zeros on the real axis at high stimulus levels	
Z_{shift}	Displacement of the zeros due to control signal	Eq. (14)
Parallel-path C2 filter	Linear and same as the broadest possible C1 filter	as in Sec. II D
Inner hair cell (IHC)		
$V_{\mathrm{ihc,C1}}$	Output of the C1 transduction function	Eqs. (17) and (18)
$V_{\mathrm{ihc,C2}}$	Output of the C2 transduction function	Eq. (19)
A_{ihc0}	Constant in C1/C2 transduction function	0.1
B_{ihc}	Constant in C1/C2 transduction function	2000
C_{ihc}	Constant in C1 transduction function	1.74
D_{ihc}	Constant in C1 transduction function	6.87×10^{-9}
$\mathrm{C_{IHC}}$	Scaling constant for impairment in the IHC	$0 \leqslant \mathrm{C_{IHC}} \leqslant 1$
$\mathrm{LP_{ihc}}$	Low-pass filter in the IHC	Same as Zhang *et al.* (2001)
Synapse model and discharge generator	Same as Bruce *et al.* (2003)	

Bruce *et al.* (2003), except for a minor modification in the parameters, which are given in Table II.

The control-path wide-band filter is a third-order time-varying gammatone filter, the center frequency and bandwidth of which are higher than those of the signal-path C1 filter. The implementation of this filter follows that of Zhang *et al.* (2001) and Bruce *et al.* (2003). The broader bandwidth of the control-path filter produces two-tone rate suppression in the model output. The wide-band filter has a center frequency corresponding to a shift of 1.2 mm basal to the fiber CF along the basilar membrane (Zhang *et al.* 2001) to be consistent with the physiological observation that the tip of a suppression tuning curve is at a slightly higher frequency than the tip of the excitatory tuning curve (Delgutte, 1990).

The tuning of the control-path gammatone filter is determined based on the tuning of the signal-path gammatone filter used in the earlier models (Zhang *et al.*, 2001; Bruce *et al.*, 2003). The maximum time constant of the control-path wide-band filter (i.e., at low stimulus levels) is denoted by τ_{cpmax}, given as

$$\tau_{\mathrm{cpmax}} = \tau_{\mathrm{wide}} + 0.2 \times (\tau_{\mathrm{narrow}} - \tau_{\mathrm{wide}}), \tag{4}$$

where τ_{narrow} and τ_{wide} denote the maximum and minimum time constants for the *signal-path* gammatone filter from

the previous models. The former is computed according to $\tau_{narrow}=2Q_{10}/(2\pi CF)$ and $\log_{10}(Q_{10})=0.4708\log_{10}(CF/10^3)+0.4664$, where CF is in hertz. The latter is determined by $\tau_{wide}=\tau_{narrow}\times$ ratio, where ratio $=10^{-gain_{CA}(CF)/(20\times3.0)}$ and $gain_{CA}$ (CF) is defined in Eq. (6). Similarly, the minimum time constant of the control-path filter (i.e., at very high stimulus levels) is defined as

$$\tau_{cpmin}=\tau_{cpmax}\times \text{ratio}. \tag{5}$$

As in Bruce *et al.* (2003), the CA gain varies as a function of CF, increasing from 15 dB at low CFs to 52 dB at high CFs [Eq. (B3) of Bruce *et al.* (2003)]:

$$gain_{CA}(CF)=\max\{15,52(\tanh(2.2\log_{10}(CF/10^3)+0.15)+1.0)/2\}, \tag{6}$$

where CF has the units of hertz. Bruce *et al.* (2003) modified somewhat the control path of Zhang *et al.* (2001) by removing a symmetric nonlinearity between the wide-band filter and the OHC Boltzmann nonlinearity. This symmetric nonlinear function was included in the Zhang *et al.* (2001) to increase the dynamic range of compression in the model, but this introduced undesired distortion products in both the control-path and signal-path output in response to multi-tone or vowel stimuli. Bruce *et al.* (2003) avoided the distortion products by producing compression in the control path dynamically rather than by a static nonlinearity.[2] Here we have further modified the gain of the control-path wide-band filter to vary as a function of CF according to the factor $(\tau_{cp}/\tau_{cpmax})^3\cdot 10\times 10^3\cdot\max(1,CF/5\times10^3)$, where CF is in units of hertz. This produces a compression threshold that is at a roughly constant level above the model auditory nerve fiber threshold at each CF (Sachs *et al.*, 1989).

The asymmetrical saturating nonlinearity that follows the wide-band filter is implemented with a second-order Boltzmann function with an asymmetry of 7:1 [Eqs. (11) and (12) of Zhang *et al.* (2001)]:

$$BN(V)=\frac{1}{1-shift_{cp}}\times\left\{\frac{1}{1+e^{-(V-x0)/s0}(1+e^{-(V-x1)/s1})}-shift_{cp}\right\}, \tag{7}$$

where

$$shift_{cp}=\frac{1}{1+e^{x0/s0}(1+e^{x1/s1})},$$

where V and BN are the outputs of the wide-band filter and the Boltzmann function, respectively. The parameters of this function are adjusted to replicate the reported physiological data in the literature (e.g., Ruggero *et al.*, 1997) and are given in Table II. The cut-off frequency of the low-pass filter following the Boltzmann function was reduced to 600 Hz in Bruce *et al.* (2003) from that of 800 Hz in Zhang *et al.* (2001), and the order of this filter is also reduced from three to two in the model presented in this paper.

The nonlinear function that converts the output of the low-pass filter to a time-varying time constant for the signal-path C1 filter, τ_{C1}, is similar to that of Zhang *et al.* (2001) and Bruce *et al.* (2003):

$$\tau_{C1}=\tau_{C1max}\left[R_{C1}+(1-R_{C1})\times\left(\frac{\tau_{C1min}/\tau_{C1max}-R_{C1}}{1-R_{C1}}\right)^{|V_{LP}|/dc}\right],$$

where $V_{LP}(t)$ is the output of the OHC low-pass filter, dc is an estimate of the dc component of the control path output at high-levels, R_{C1} is the ratio of an asymptotic lower bound of the time constant to bound of the time constant to τ_{C1max}. τ_{C1max} and τ_{C1min} are the estimated time constants of the C1 filter at low and high levels, respectively, and will be defined in the next section. The parameters of this function are also described in Table II. Finally, the time-varying time constant for the control-path wide-band filter is found by

$$\tau_{cp}=f(\tau_{C1})=a\cdot\tau_{C1}+b, \tag{8}$$

where the constants are suitably defined as $a=(\tau_{cpmax}-\tau_{cpmin})/(\tau_{C1max}-\tau_{C1min})$ and $b=\tau_{cpmax}-a\cdot\tau_{C1max}$. Here the constants a and b are selected so that the range of variations possible for the time constants of the C1 filter (between τ_{C1max} and τ_{C1min}) are mapped into the range (between τ_{cpmax} and τ_{cpmin}) for the control-path filter.

At low stimulus sound pressure levels, the control-path output, τ_{C1}, is almost equal to τ_{C1max} such that the tuning is sharp, the gain is high, and the filter behaves linearly. At moderate levels, the control signal deviates substantially from τ_{C1max}, dynamically varying between τ_{C1max} and τ_{C1min}. The tuning of the C1 filter becomes broader and the gain is reduced, simulating the compressive and suppressive cochlear nonlinearity. At very high stimulus sound pressure levels, the control signal saturates, essentially equal to τ_{C1min}, and the C1 filter is again effectively linear with broad tuning and low gain.

C. C1 filter

The C1 filter produces the tuning properties for the model BM response that provides input to the C1 IHC transduction function. It consists of two second-order poles (i.e., two poles at the same position) and one first-order pole (a single pole), their complex conjugates and a fifth-order zero on the real axis. The configuration of this filter (i.e., the relative placement of the poles and zeros) is similar to the recent model proposed by Tan and Carney (2003), however the filter order and the equations controlling the initial placement and the movements of the poles and zeros in the complex plane according to the control signal are substantially different. The filter order has a great impact on the sharpness of tuning; if the filter order is too high, then the filter remains fairly sharply tuned even for high-sound pressure level stimuli or in the case of OHC impairment. To make the impaired and high-level tuning more realistic (i.e., reasonably broadly tuned), the order of the C1 filter has been reduced

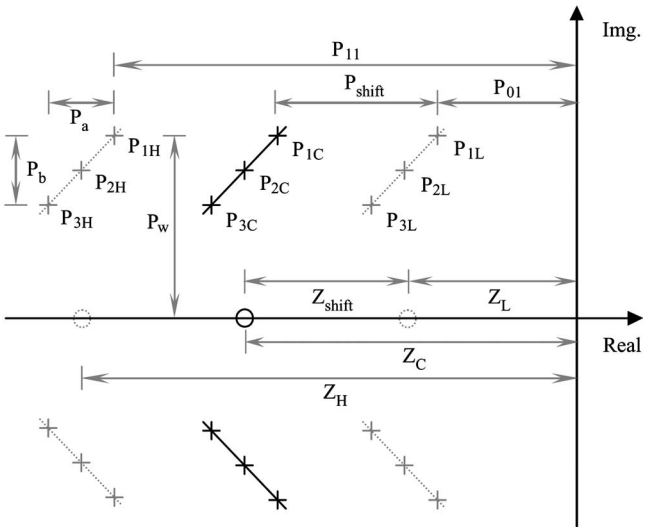


FIG. 2. Configuration of the location of poles and zeros of the model's signal-path BM (C1) filter. Five pairs of poles (P_{1C}, P_{3C} are second order and P_{2C} is a first-order pole and their complex conjugates) along with five zeros (Z_C on the real axis) in the complex plane represent the C1 filter pole and zero position as a function of the control-signal, which is mapped to the shift in pole position P_{shift} and zero position Z_{shift}. P_{1L} represents the closest that the pole can be to the imaginary axis, i.e., at low stimulus levels, and P_{1H} indicates the furthest possible location in the completely impaired cochlea or at very high stimulus levels in the normal case. Similar interpretation holds for other poles such as P_{2L} and P_{2H}, respectively, and also for P_{3L} and P_{3H}. Z_L and Z_H indicate the closest and furthest possible locations of the fifth-order zeros in the complex plane.

from 20 to 10. The model proposed by Tan and Carney (2003) is limited to CFs up to 4 kHz, however, the model presented here is able to simulate responses for AN fibers with CFs as high as 40 kHz.[3]

The configuration (i.e., the structure of the relative positions) and the limits of movement of the poles and zeros of the C1 filter are shown in Fig. 2. The CF-dependent pole-zero locations for *low-level stimuli* are governed by the desired Q_{10} values and tuning curve tail shape, and are described by the following (also, see Fig. 2):

$$P_{01} = 1/\tau_{\text{C1max}} = 0.7 \times (2\pi\text{CF})/(2Q_{10}), \tag{9}$$

$$P_w = 1.01 \times 2\pi\text{CF} - 50, \tag{10}$$

$$P_b = 0.2343 \times 2\pi\text{CF} - 1104, \tag{11}$$

$$\log_{10}(P_a - 2000) = \log_{10}(\text{CF}) \times 0.9 + 0.55, \tag{12}$$

$$\log_{10}(Z_L - 500) = \log_{10}(\text{CF}) \times 0.7 + 1.6, \tag{13}$$

where CF is in units of hertz. P_{01} denotes the real part of the pole nearest to the imaginary axis, P_w is the imaginary part of the pole closest to the imaginary axis, P_b is the relative imaginary parts of the poles, P_a specifies the relative real parts of the poles, and Z_L is the location of the zeros on the real axis; for simplicity, all zeros are located at the same position in the complex plane. The selection of these functions is motivated by a number of physiological observations reported in the literature and thus reflect the following phenomena in the model.

First, physiological studies showed that the shape of neural tuning curves gradually changes with CF (Kiang *et al.*, 1965). Low-CF fibers have shallow, symmetric tuning, and high-CF fibers have sharp, asymmetrical tuning with extended low-frequency tails. Single AN fibers therefore appear to behave as bandpass filters, with asymmetric filter shapes. The frequency selectivity of AN fibers is similar to that of the BM and IHCs, from which their frequency selectivity is almost certainly derived (Russell and Sellick, 1978; Sellick *et al.*, 1982; Narayan *et al.*, 1998; Robles and Ruggero, 2001). Although the BM and IHC data are primarily from species such as chinchilla and guinea pig, a method for estimating BM amplitudes from AN responses gives similar results for cats (van der Heijden and Joris, 2003). The relative positions of the poles and zeros of the C1 filter have been chosen carefully to replicate this aspect of tuning in the model.

Second, several studies reported frequency modulation in the impulse response of the BM (Robles *et al.*, 1976; de Boer and Nuttall, 1997; Recio *et al.*, 1997) and AN fibers (Carney *et al.*, 1999; Lin and Guinan, 2000). The signal-path C1 filter with asymmetrical transfer function has been configured to reflect the IF glides in its impulse response. The interactions among the poles and zeros give rise to the glides, which were extensively explained in Tan and Carney (2003). The frequency glides in the impulse response of the C1 filter of the proposed model show downward glides for CFs below 750 Hz, constant glides for CFs ranging from 750 to 1500 Hz, and upward glides for CFs above 1500 Hz, which are qualitatively consistent with the AN data (Carney *et al.*, 1999).

Third, the zeros of the C1 filter are chosen in such a way that the magnitude response of the BM for a wide range of CFs reported in van der Heijden and Joris (2003) can be achieved by this model, and consequently the tails of the tuning curves become more realistic.

The location of the poles and zeros are varied along the real axis in the negative direction according to the time-varying control signal to account for the compressive and suppressive nonlinearity in the model. Since the output of the control path is the time constant for the C1 filter, we define mapping functions that convert the time constant to a corresponding shift of the location of poles (P_{shift}) and zeros (Z_{shift}) from the initial positions,

$$P_{\text{shift}} = \left(\frac{1}{\tau_{\text{C1}}} - \frac{1}{\tau_{\text{C1max}}}\right), \tag{14}$$

$$Z_{\text{shift}} = \frac{2\pi\text{CF}}{\tan((\phi_{ZL} - \phi_{PL} + \phi_{PC})/5)} - Z_L. \tag{15}$$

Here $\phi_{ZL} = 5 \cdot \tan^{-1}(2\pi\text{CF}/(-Z_L))$ and $\phi_{PL} = \Sigma_{i=1}^{10} \tan^{-1}(2\pi\text{CF} - \text{Img}(P_{iL})/(-\text{Real}(P_{iL})))$ denote the sum of the phases at CF (Hz) due to the zeros and poles of the C1 filter, respectively, at low levels. Similarly, ϕ_{PC} represents the total phase contribution by the poles of the C1 filter at CF after the poles have been moved according to the control signal. With the movement of these poles and zeros in the complex plane, the phase response of the C1

filter at CF becomes invariant to stimulus sound pressure level. Unlike Tan and Carney (2003), the locations of the zeros of the C1 filter are moved along the real axis according to the control signal to account for the following two important level-dependent response properties in the model.

First, at the base of the cochlea, BM responses to tones with near-CF frequency grow nonlinearly, and phases vary systematically with sound pressure level (Geisler and Rhode, 1982; Cooper and Rhode, 1992; Nuttall and Dolan, 1993; Ruggero *et al.*, 1997; Rhode and Recio, 2000). However, at frequencies well below CF, response magnitudes grow linearly with stimulus sound pressure level and phases at those frequencies are invariant with respect to the stimulus sound pressure level. The same pattern of phase shift with sound pressure level was described in the phases of IHCs (Dallos, 1986; Cheatham and Dallos, 1989, 1998) and AN fibers (Anderson *et al.*, 1971; Liberman and Kiang, 1984; Cai and Geisler, 1996; Ruggero *et al.*, 1996). To address these level-dependent phase properties in the model, zeros are moved along with the poles (along the real axis toward negative infinity) in such a way that at all sound pressure levels the phase response of the C1 filter remains unaltered at that particular CF, i.e., the phase at CF is independent of the control signal. So, the phase is affected by the control signal only for tone frequencies other than CF. However, for tones well below CF, the displacement of the poles and zeros of the C1 filter from the initial positions is negligible, and thus phase responses remain unaffected by the stimulus sound pressure level. The tone frequencies around CF cause significant displacement of the poles and zeros in the complex plane of the C1 filter, and hence the phase response varies systematically with sound pressure level.

Second, model responses show that as the control signal becomes significant (i.e., the corresponding displacements of poles and zeros in the complex plane are greater) at moderate and high levels, the frequency glides in the impulse response of the BM and AN become level-dependent if the locations of zeros are not moved along with the poles within the C1 filter. So, the movement of the positions of zeros along with the poles make the frequency glides in the impulse response of the BM and AN fibers almost level-independent, consistent with physiological observations (e.g., Recio *et al.*, 1997; de Boer and Nuttall, 1997; Carney *et al.*, 1999).

The maximum possible movement of the poles and zeros depends on the cochlear-amplifier gain, i.e., the gain of the CA defines the relation between the tuning at low and high levels. So, within the C1 filter, the real part of the pole nearest to the imaginary axis at high levels (when the control signal is saturated) is described as

$$P_{11} = \frac{1}{\tau_{\mathrm{C1min}}} = \frac{1}{\tau_{\mathrm{C1max}} \cdot 10^{-\mathrm{gain_{CA}(CF)}/(20\times 2.5)}}. \tag{16}$$

Hence, the maximum possible displacement of the poles and zeros from the initial positions (i.e., at low levels) along the real axis in the negative direction in the complex plane, $(P_{11}-P_{01})$ and (Z_H-Z_L), respectively, are limited by the CA gain (Fig. 2).

D. C2 filter

A C2 filter, parallel to the C1 filter, has been introduced according to Kiang's two-factor cancellation hypothesis to implement the C1/C2 transition and peak-splitting phenomena in the model. In order to comply with the hypothesis proposed by Kiang (1990) and other experimental observations (e.g., Gifford and Guinan, 1983; Sewell, 1984b; Liberman and Kiang, 1984; Wong *et al.*, 1998), the C2 filter should possess the following two important characteristics, which are implemented accordingly in the model.

First, Liberman and Kiang (1984) showed that C2 tuning curves are very broad (almost flat). The thresholds of a C2 tuning curve for a particular fiber are defined as the levels at which the responses of the fiber undergo an abrupt phase shift of about 180° as a function of the stimulus frequency. In addition, it was reported in Wong *et al.* (1998) that above the C2 threshold, AN fibers have poor frequency selectivity and respond to many frequency components of a vowel stimulus. So, the C2 filter has to be very broad. In the proposed model, the C2 filter (tenth order) has been chosen to be identical to the broadest possible C1 filter. Hence the C2 filter has been implemented by placing the poles and zeros in the complex plane at the positions for the C1 filter when the OHCs are completely impaired. Implementation of impairment in the cochlea will be discussed in Sec. II G.

Second, in acoustically traumatized cats, C2 responses remain unaltered, while C1 responses are significantly attenuated (Liberman and Kiang, 1984; Heinz and Young, 2004). Sewell (1984b) also showed that C2 responses are less sensitive to decreases in EP produced by furosemide. Similarly, Gifford and Guinan (1983) reported that C1 responses can be suppressed by stimulation of the crossed olivocochlear bundle, whereas C2 responses cannot. All of these observations imply that the C1 response is dependent on OHC function but C2 is not. To reflect this finding in the model, the C2 filter is linear, static (i.e., the tuning of the C2 filter remains fixed across all levels), and is followed by a separate IHC transduction function.

E. The inner hair cell (IHC)

The function of the IHC is to transduce the mechanical responses of the BM to an electrical potential that leads to neurotransmitter release across the IHC-AN synapse. The ratio of the ac to dc components of the IHC output affects the synchrony of the fiber's response to tones (Dallos, 1985; Palmer and Russell, 1986). In acoustically traumatized cats, Liberman and Kiang (1984) found severe damage to both the OHC stereocilia and to the tallest row of IHC stereocilia with only shorter IHC stereocilia being normal. They hypothesized that the tallest row of IHC stereocilia are responsible for the generation of C1 responses and C2 responses are produced by the shorter IHC stereocilia. Hence acoustic trauma can reduce or eliminate C1 responses with little effect on C2 responses. The above-mentioned finding, together with the phase shift being a fixed phenomenon of the C1/C2 transition, imply that C1 and C2 represent two different modes of stimulation of the IHC that are 180° out of phase. The IHC section of the model has been modified accordingly

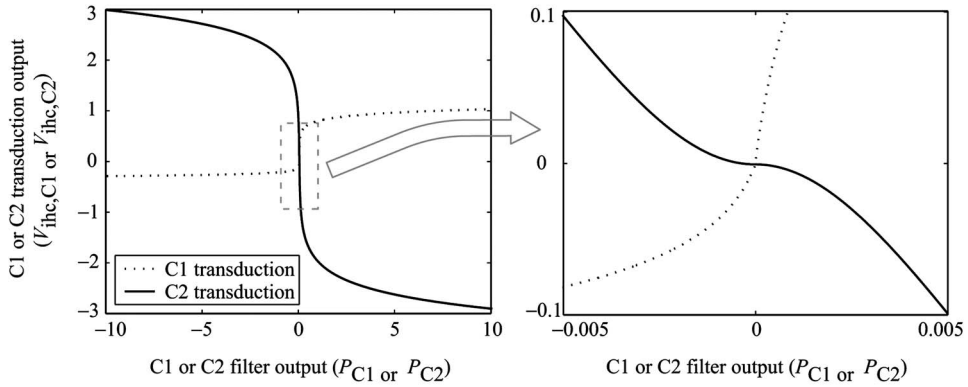


FIG. 3. C1 and C2 transduction functions. Left panel: transduction outputs are shown across a large range of inputs so that the saturation in both functions is evident. Right panel: zoomed-in version to illustrate that at low levels the growth of the C1 response is sharply increasing, whereas the growth of C2 is shallow.

from previous models (Zhang *et al.*, 2001; Bruce *et al.*, 2003) and is shown in Fig. 1. The IHC model consists of two transduction functions following the C1 and C2 filters with the outputs being summed before low-pass filtering by the IHC.

The C1 transduction function following the C1 filter is the same as that of Zhang *et al.* (2001) and Bruce *et al.* (2003). This function produces a dc component in the IHC's of high-CF model fibers, providing nonsynchronized synaptic drive to such fibers. A logarithmic compressive function is used [Eqs. (15) and (16) of Zhang *et al.* (2001)]:

$$V_{\mathrm{ihc,C1}} = A_{\mathrm{ihc}}[P_{\mathrm{C1}}]\log(1 + B_{\mathrm{ihc}}|P_{\mathrm{C1}}|), \tag{17}$$

where P_{C1} represents the output of the signal-path C1 filter, and

$$A_{\mathrm{ihc}}[P_{\mathrm{C1}}] = \begin{cases} A_{\mathrm{ihc0}} & \text{for } P_{\mathrm{C1}} > 0 \\ -\dfrac{|P_{\mathrm{C1}}|^{C_{\mathrm{ihc}}} + D_{\mathrm{ihc}}}{3|P_{\mathrm{C1}}(t)|^{C_{\mathrm{ihc}}} + D_{\mathrm{ihc}}} A_{\mathrm{ihc0}} & \text{for } P_{\mathrm{C1}} < 0. \end{cases} \tag{18}$$

Here A_{ihc0}, B_{ihc}, C_{ihc}, and D_{ihc} are the constants [Table I of Zhang *et al.* (2001)] that were adjusted to achieve realistic IHC response properties. The asymmetry of this nonlinear function varies smoothly from 1:1 (at low levels) to 3:1 (at high levels) as a function of the input signal, P_{C1}.

The C2 transduction function following the C2 filter should have the following two important properties to comply with the hypothesis proposed by Kiang (1990). First, the C2 output is shifted in phase by 180° with respect to the C1 response. Second, the C2 transduction function should be such that the C1 output is significantly greater in amplitude than the C2 output at low and moderate levels, while at high levels the C1 output is almost negligible compared to the C2 output. In the proposed model, the C2 transduction function has been implemented in such a way as to ensure these two response properties:

$$V_{\mathrm{ihc,C2}} = \mathrm{signum}(-P_{\mathrm{C2}}) \cdot 2 \cdot A_{\mathrm{ihc0}} \log(1.0 + 5 \times 10^{-5} B_{\mathrm{ihc}}(P_{\mathrm{C2}} \cdot \mathrm{CF})^2), \tag{19}$$

where P_{C2} denotes the output of the C2 filter and CF is expressed in units of hertz. A_{ihc0} and B_{ihc} are the same constants as in the C1 transduction function.

It is to be noted that the C2 transduction function differs from the C1 transduction function in two aspects. First, it is symmetrical and inverted (with respect to the C1 output), whereas the C1 transduction output is highly asymmetric (3:1) at high levels. Second, the output of the C2 filter, P_{C2}, is transformed to $[5 \times 10^{-5}(P_{\mathrm{C2}} \cdot \mathrm{CF})^2]$ to make it sharply increasing around 90–100 dB SPL, whereas the output of the C1 filter, P_{C1}, is directly used in the C1 transduction function. In addition, the slope of the growth for C2 transduction is twice that of the C1 transduction function. This CF-dependent transformation along with the higher slope of the growth function ensure the C1/C2 transition occurs within a few decibels across a wide range of CFs (Wong *et al.*, 1998). For comparison, these two functions are plotted in Fig. 3. Note that the growth of $V_{\mathrm{ihc,C1}}$ and $V_{\mathrm{ihc,C2}}$ with *stimulus sound pressure level* also differ because of the different gains of the C1 and C2 filters at low and moderate sound pressure levels.

The output of the two transduction functions, $V_{\mathrm{ihc,C1}}$ and $V_{\mathrm{ihc,C2}}$, are added together and passed through the IHC low-pass filter to give the IHC potential V_{ihc}. The low-pass filter of the IHC section is a seventh-order filter with a cut-off frequency of 3800 Hz (Zhang *et al.*, 2001). This was chosen to match the cat data published by Johnson (1980) that describes the fall-off in pure-tone synchrony with increasing tone frequency.

F. Synapse model and discharge generator

The IHC-AN synapse and the discharge generator section of the proposed model are similar to those of Zhang *et al.* (2001). The nonlinear IHC-AN synapse is described by the time-varying three store diffusion model of Westerman and Smith (1988). The spontaneous rate, adaptation properties, and the rate-level behavior of the model are determined by the synapse model. The synapse gain, which describes the relationship of the IHC potential to the synaptic release rate, varies as a function of CF to make the model thresholds match the empirical data from cats—see Fig. 4(a). In the proposed model, the parameters of this function are slightly modified from Bruce *et al.* (2003):

$$K_{\mathrm{CF}} = 10^{0.29\mathrm{CF}/10^3 + 0.4}, \tag{20}$$

where CF is in units of hertz.

Previous versions of the model with a low-order gammatone filter (Carney, 1993; Zhang *et al.*, 2001; Bruce *et al.*, 2003) used a CF-dependent delay to match the spike laten-

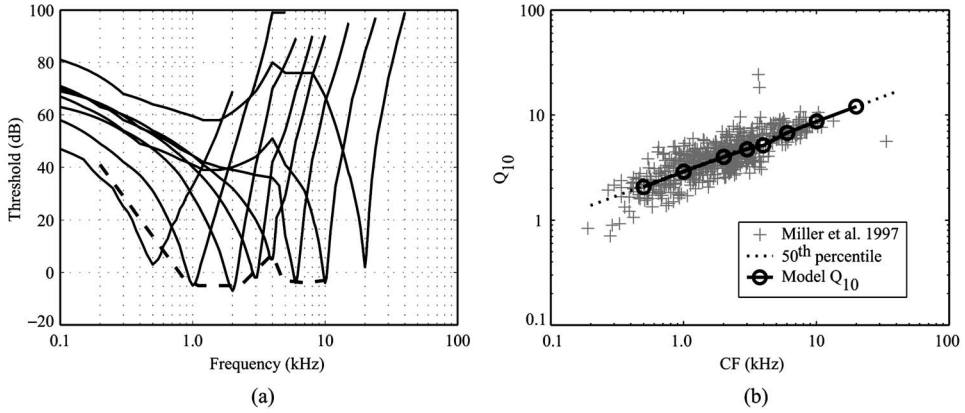


FIG. 4. (a) Model tuning curves for AN fibers with CFs ranging from 500 Hz to 20 kHz, threshold being the tone level that produces a rate of 10 spikes/s more than the spontaneous rate. The dashed line shows the best threshold curve (BTC) based on the cat AN population data of Miller *et al.* (1997). (b) Q_{10} values vs CF, measured from the model tuning curves and compared to the cat AN population data of Miller *et al.* (1997).

cies of the model to measured data. As the C1 filter is tenth order, we have modified the delay function that matches model reverse-correlation (revcor) functions to the measured functions (Carney and Yin, 1988):

$$\alpha(\mathrm{CF}) = 3.0e^{(-x_{\mathrm{CF}}/12.5)}, \tag{21}$$

where x_{CF} is the distance (mm) from apex of the basilar membrane (Liberman, 1982) and α(CF) represents the delay in ms that includes acoustic, traveling-wave, synaptic, and neural delays.

Finally, discharge times are produced by a renewal process that includes refractory effects and is driven by the synapse output. The implementation of the discharge generator section of the proposed model is the same as that of Zhang *et al.* (2001) and Bruce *et al.* (2003).

G. Modeling OHC and IHC impairment

It has been reported in some anatomical investigations (e.g., Liberman and Dodds, 1984) that noise-induced impairment in the cochlea causes damage to both the outer and inner hair cell stereocilia. Damage to the OHC stereocilia causes both broadened and elevated AN fiber threshold tuning curves (Kiang *et al.*, 1976; Robertson, 1982; Liberman and Dodds, 1984), whereas damage to the IHC results only in elevation of the tuning curves, without any substantial broadening (Liberman and Dodds, 1984). Furthermore, OHC lesions are associated with the reduction in the compressive nonlinearity in the responses of the BM (Rhode, 1973; Robles and Ruggero, 2001) and also with the reduction in the two-tone suppressions of the AN responses (Schmiedt *et al.*, 1980; Salvi *et al.*, 1982; Miller *et al.*, 1997). In this model, both the OHC and IHC sections of the normal model have been modified to address hair cell damage in the cochlea (Bruce *et al.*, 2003).

The effects of the OHC status are incorporated in the model by introducing a scaling factor $0 \leqslant C_{\mathrm{OHC}} \leqslant 1$ to the control path output (more specifically to the time constant of the C1 filter), as proposed by Bruce *et al.* (2003). $C_{\mathrm{OHC}}=1$ simulates the normal functioning of OHC in the model and $C_{\mathrm{OHC}}=0$ indicates complete impairment in the OHC. As the behavior (e.g., gain and bandwidth) of the C1 filter is determined by the control signal, scaling down the control path output causes both increased bandwidth (broadened tuning) and decreased gain (elevated threshold) in the C1 responses, which are qualitatively consistent with physiological studies of damage to the OHC only. Thus, the control signal has been modified according to the degree of impairment in the OHC and is given by the following relation [Eq. (8) of Bruce *et al.* (2003)]:

$$\tau_{\mathrm{C1\text{-}impaired}} = C_{\mathrm{OHC}}(\tau_{\mathrm{C1}} - \tau_{\mathrm{C1min}}) + \tau_{\mathrm{C1min}}. \tag{22}$$

The effects of C_{OHC} on the tuning curves and different responses of the model will be discussed in Sec. III.

Liberman and Kiang (1984) reported that acoustically traumatized cats show significant attenuation in C1 responses, while C2 responses are robust and resistant to trauma. So, the C2 transduction function remains unaffected in the impaired cochlea, and thus impairment in the IHC has been addressed in the model by introducing a scaling constant ($0 \leqslant C_{\mathrm{IHC}} \leqslant 1$) in the C1 transduction function only, similar to Bruce *et al.* (2003). This constant varies from 0 to 1 according to the degree of impairment in the IHC: $C_{\mathrm{IHC}}=1$ corresponds to the normal functioning of IHC and $C_{\mathrm{IHC}}=0$ indicates complete impairment.

III. RESULTS

The following describes some of the basic response properties of the model. The main focus here is to illustrate the high-level effects to tones. The responses of the model to pairs of tones are similar to the results shown in Zhang *et al.* (2001) and are not repeated here. The responses to other complex stimuli, particularly speech, have been considered in development of the model but are beyond the scope of this paper.

A. Tuning curves and Q_{10} values

The tuning characteristics of an AN fiber at threshold are represented by the frequency threshold curve. Figure 4 shows the model tuning curves of AN fibers across a wide range of CFs for parameters simulating a normal cochlea. The same paradigm used in Liberman (1978) is employed here to plot these tuning curves. These normal tuning curves have the following characteristics.

First, low-CF fiber tuning curves have shallow slopes on both sides of the CF and are symmetric, whereas the tuning of high-CF fibers become increasingly asymmetric with steep high-frequency slopes and less steep low frequency slopes with a long broadly tuned tail on the low-frequency

side of the CF, which are consistent with tuning curves from cat (e.g., Kiang *et al.* 1965; Liberman, 1978).

Second, for AN fibers with CF$\geqslant$1 kHz, the model thresholds at CF are a good match with the best threshold curve (BTC) defined from a population of data published in Miller *et al.* (1997) [dashed line in Fig. 4(a)]. In the low-CF region (below 1 kHz), the thresholds are slightly lower than the BTC, which could be addressed by adjusting the gain of the ME filter or the synapse gain in the IHC-AN section. However, this is not done here because there is some variability in the best thresholds of low-CF fibers in cats.

Third, the notch immediately above 4 kHz in the ME filter is reflected in the tuning curves as an upward notch (Bruce *et al.*, 2003). The slightly lower values of the model Q_{10} around 4 kHz are due to the notch in the ME filter.

To simulate impaired tuning due to hair cell damage in the model, the scaling constant C_{OHC} and C_{IHC} are set according to the degree of the impairment in the OHC and IHC sections, respectively. Figure 5 shows the C1 tuning curves of AN fibers with CFs at 0.5, 3.0, and 15.0 kHz for five different degrees of impairment in the OHC (C_{OHC}=1.0, 0.75, 0.5, 0.25, and 0.0, corresponding to 0%–100% impairment in the OHC), and have the following important characteristics.

First, the C1 tuning curves of the impaired fibers are elevated in threshold and broadened in tuning, consistent with the physiological data (Liberman and Dodds, 1984). The extreme case with complete OHC lesion (C_{OHC}=0) gives the highest elevation and broadest tuning possible by the model. Also, the compression and suppression nonlinearities in the BM responses are completely lost as the C1 filter becomes linear (results not shown).

Second, since the C1 filter is asymmetrical in shape and is responsible for the glides in the IF of its impulse response, impairment in the OHC causes the center frequency of the filter to shift from its location at low levels (Robles and Ruggero, 2001). With impairment in the OHC, CF shifts to a lower frequency for high-CF fibers (>1.5 kHz), shifts upward for low-CF fibers (<0.75 kHz), and does not shift significantly for the mid-CF fibers (0.75<CF<1.5 kHz), consistent with the experimental data (Liberman and Mulroy, 1982; Sewell, 1984a; Liberman, 1984).

Third, thresholds in the low-frequency tails of the tuning

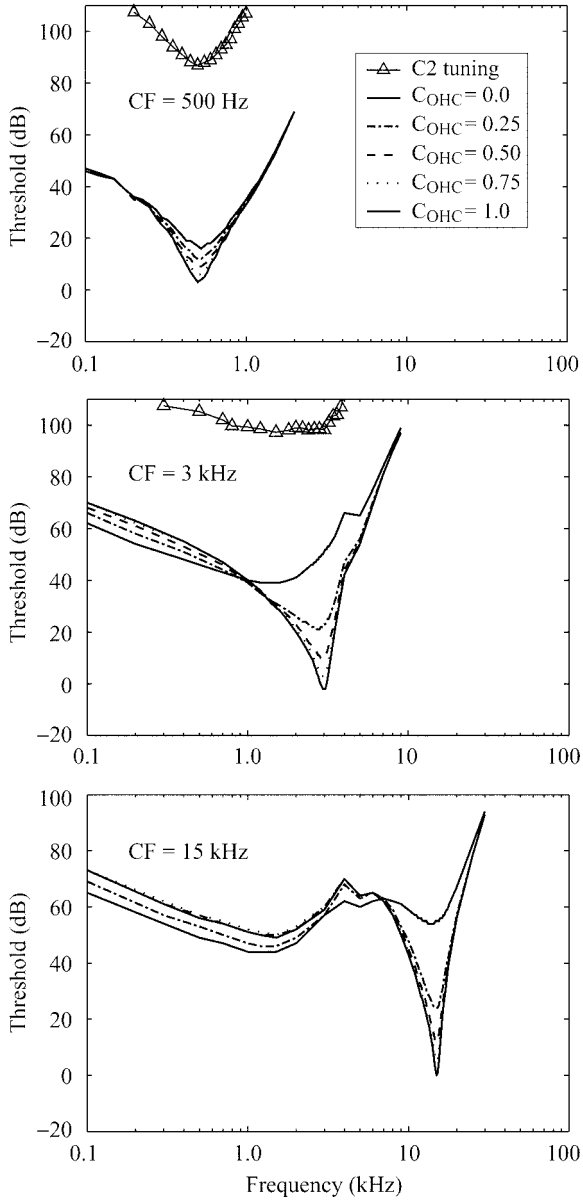


FIG. 5. Model C1 tuning curves of three representative AN fibers with CFs at 0.5, 3.0, and 15.0 kHz as a function of impairment in the OHC: with no impairment C_{OHC}=1.0 and for complete impairment C_{OHC}=0.0. For each fiber, tuning curves are shown for five different degrees of impairment in the OHC with a step of 0.25. Also plotted are the C2 tuning curves (triangles) for the model fibers with CFs at 0.5 and 3.0 kHz.

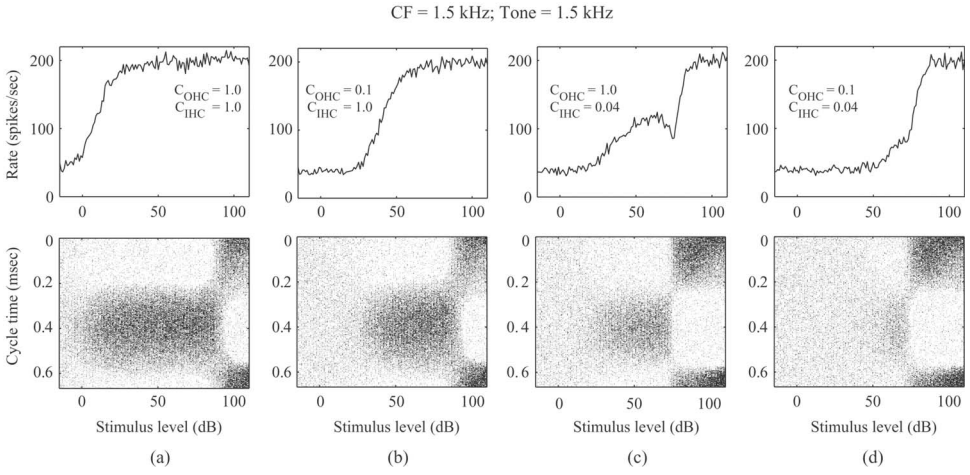


FIG. 6. Model AN fiber discharge rate and phase raster plot vs level. Top panel: mean discharge rates for a model fiber with CF at 1.5 kHz in response to a tone at CF for four different conditions of the cochlea: (a) normal OHC, normal IHC; (b) impaired OHC, normal IHC; (c) normal OHC, impaired IHC; and (d) impaired OHC, impaired IHC. Lower panel: phase raster plots display period histograms along the ordinate at various sound levels. At each sound level, a dot is placed at the stimulus phase corresponding to each action potential in the AN fiber response.

curves decrease slightly with increasing impairment, which indicates the hypersensitivity of the tail. For fibers with higher CFs, the tuning curves become "W-shaped," consistent with some physiological observations (Liberman and Kiang, 1984). In the model, the hypersensitivity of the tail is due to the movement of the C1 filter zeros to their impaired position, which acts to slightly increase the filter gain in the tail.

Also plotted in Fig. 5 are C2 tuning curves (triangles) for the model fibers with CFs at 0.5 and 3.0 kHz. These C2 tuning curves are consistent with C2 tuning curves measured in cats (e.g., Fig. 2 of Liberman and Kiang, 1984). The C2 threshold here is defined as the level at which a 90° phase shift occurs in the AN response—see Sec. III F for more details. The C2 tuning for the fiber with CF at 15.0 kHz cannot be determined because the synchrony to tones with frequencies around CF is insufficient to reliably determine the phase of the AN response, consistent with Fig. 6D of Kiang *et al.* (1986). Note that the C2 thresholds are *not* equal to the C1 thresholds with complete OHC impairment. Although the C1 and C2 filters are identical in this case, the shallower growth of the C2 transduction function at low to moderate levels (see Fig. 3) means that the outputs of the C1 and C2 transduction functions do not become comparable in magnitude until higher sound pressure levels.

B. Mean discharge rate and phase versus level

Figure 6 shows the rate-level functions for a model AN fiber with CF at 1.5 kHz in response to a tone at CF. Four different conditions of the cochlea are considered: (a) normal hair cells; (b) impaired OHC and normal IHC; (c) normal OHC and impaired IHC; and (d) impaired OHC and impaired IHC. To illustrate whether the responses are associated with C1 or C2, phase raster plots are shown below each rate-level function plot. The phase raster displays show the period histogram along the ordinate, and the abscissa indicates the sound pressure level of the stimulus in dB SPL. Each ordinate corresponds to one cycle of the stimulus and each dot in the period histogram indicates the stimulus phase at which a spike occurs in the AN fiber responses. Note that the model fiber considered here is a low-threshold fiber.

The normal rate-level function shown in Fig. 6(a) has a narrow dynamic range and saturates around the sound level at which the BM becomes compressive, so the rate-level function does not exhibit any effects from BM compression. The phase raster for this case shows a transition from the C1 phase to the C2 phase, such that the discharge rate remains at approximately the same level. Consequently, no notch is observed in the rate-level function in this example model fiber.

Impairment of the OHC with the IHC being normal [shown in Fig. 6(b)] gives almost the same dynamic range in the rate-level function as the normal fiber, and the responses are shifted to the higher level without any substantial change in the slope, consistent with the model described in Fig. 1 of Heinz and Young (2004). The phase raster plot also appears unchanged for suprathreshold stimuli.

Impairment in the IHC only [Fig. 6(c)] substantially attenuates the C1 amplitude, producing a rate-level curve with a much shallower initial slope than normal. Here, a notch becomes more evident in the rate-level function and there is an abrupt change of phase around the stimulus level of the notch, consistent with physiological data (e.g., see Fig. 6B of Liberman and Kiang, 1984). Following the initial shallow slope is a steep high-level response associated with the C1/C2 transitions, as shown by the corresponding phase raster plot. Note that the maximum discharge rate for the C1 response is reduced, whereas the overall maximum discharge rate (produced by the C2 response) is similar to that of normal fibers, consistent with physiological data (Liberman and Kiang, 1984; Miller *et al.*, 1999; Heinz and Young, 2004).

Severely impaired AN fibers (with both OHC and IHC impairment) have very steep rate-level functions (Heinz and Young, 2004). In the model response [Fig. 6(d)], this steep slope at high levels is due to the C2 responses only, which is evident from the corresponding phase raster plot.

C. Instantaneous frequency (IF) glides

To compare model results with the data published by Carney and Yin (1988), measured and model revcor functions of an AN fiber with a CF of 1914 Hz are shown in Figs. 7(a) and 7(b), respectively, for six different levels ranging from 30 to 80 dB SPL. The same technique used in Carney *et al.* (1999) has been employed here to estimate the model revcor function. Model revcor functions are in good agreement with the published data and have the following properties: (i) For all levels shown, the zero-crossing points of the revcor functions are almost identical (dotted line), which confirms that the IF glides are independent of the input stimulus levels. The locations of the zeros in C1 filter are moved along with the poles according to the control signal to ensure this property in the model. (ii) The revcor function waveforms are more quickly damped out at higher levels, i.e., the duration of the IF trajectory is shorter at higher levels and longer at lower levels, because of the broader tuning at higher stimulus levels and sharper tuning at lower levels, respectively. (iii) Responses to high-level stimuli show the peak of the impulse response occurring at shorter latencies, and relatively longer latencies to low-level stimuli. So, fibers that show upward frequency glides in their impulse responses ($CF > 1.5$ kHz) have more energy in the low-frequency part of the glides with increased stimulus level, i.e., the BF is shifted to lower frequencies at high levels. Similarly, low-CF fibers that show downward glides in their impulse responses ($CF < 0.75$ kHz) have the BF shifted to higher frequencies with increased stimulus sound pressure levels. All of these observations are consistent with the physiological data (Evans, 1981; Carney *et al.*, 1999).

Figure 7(c) and 7(d) show the IF trajectories of the measured and model revcor functions, respectively, for three different CFs exhibiting downward, constant and upward glides in their impulse responses across three different noise levels. The model responses are compared with the data published in Carney and Yin (1988). Figure 8 shows the slopes of the IF trajectories of the model revcor functions, which are in good agreement with the lower values of the measured data

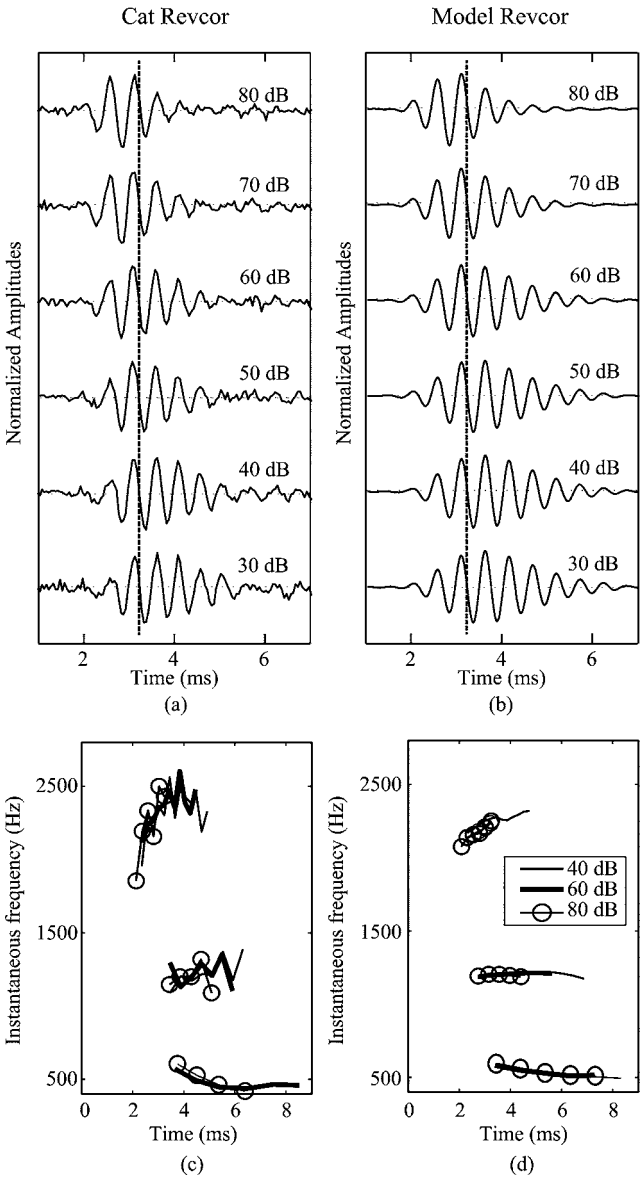


FIG. 7. Revcor functions and IF glides. (a) Measured (unit 86100-25, Carney and Yin, 1988) and (b) model revcor functions for an AN fiber with CF of 1914 Hz at six different sound levels ranging from 30 to 80 dB SPL. All revcor functions are normalized to their peak magnitude. (c), (d) The measured (Carney and Yin, 1988) and model IF glides, respectively, in the impulse response of the three different AN fibers with CFs at 508 Hz (unit 86100-2), 1210 Hz (unit 86095-30), and 2305 Hz (unit 86100-26) for three sound pressure levels (40, 60, and 80 dB SPL). The overlap of the IF profiles of a particular model fiber for different levels indicates that these glides are level independent.

(Carney *et al.*, 1999) across a wide range of CFs. However, the proposed model cannot produce the steeper slopes that are seen in the data in the high-CF region.

D. Reverse correlation filters and level dependence of BF

The revcor filter determined from the Fourier transform of the revcor function to broadband noise gives an estimate of a linear filter that describes the response of an AN fiber to that particular noise signal (de Boer and Kuyper, 1968; Carney and Yin, 1988). The measured and model revcor filters for an AN fiber with CF at 1914 Hz are shown in Figs. 9(a) and 9(b), respectively, across a wide range of noise levels. Each revcor filter is normalized to its maximum magnitude and for clarity a 1 dB shift is introduced between filters at different levels. As in Carney and Yin (1988), several signal-processing techniques are used to improve the estimate of the revcor filter.

The model revcor filter becomes increasingly broad and BF is shifted to lower frequency as the noise level is increased. These are consistent with the measured data for AN fibers in the range CF$>$1.5 kHz (Møller, 1977; de Boer and de Jongh, 1978; Evans, 1981; Carney and Yin, 1988) and also with the BM responses at the base of the cochlea (Rhode, 1971; Ruggero *et al.*, 1997). BF here is estimated as the frequency corresponding to the peak of the revcor filter. The monotonic BF shift as a function of sound pressure level is consistent with the monotonic IF glide in the impulse response.

The phase of the revcor filter response changes systematically with the sound pressure level of the noise stimulus for frequencies above and below CF. The measured and model phases for the same CF are shown in Figs. 9(c) and 9(d), respectively, for stimulus levels ranging from 30 to 80 dB SPL. Here, the phase at 80 dB SPL is chosen as reference for all frequencies shown. The model phase response shows a reasonably good quantitative match to the reported data, with both being fairly invariant with stimulus level at CF and increasingly lagging and leading below and above CF, respectively, with decreasing stimulus sound pressure level.

E. Level dependence of phase properties for high-CF fibers

For high-CF AN fibers (CF$>$4 kHz), level-dependent changes in phase cannot be accurately measured due to the lack of synchrony to tones around CF. To evaluate the model's phase properties for high CFs, the phases are computed from the output of the C1 filter that corresponds to the BM response. Since the stages followed by the C1 filter do not introduce any level-dependent phase change in the AN responses, the C1 filter output reflects the level-dependent phase changes in the model responses and confirms that this property is a result of the nonlinear interaction in the cochlea. High-CF fibers show stronger phase changes with level than low-CF fibers, as the CA gain is greater at high CFs.

Figures 10(a) and 10(b) show the level-dependent phase changes in the measured and model responses, respectively, across a wide range of stimulus frequencies for an AN fiber with CF at 10 kHz. To compare with the BM phase data from chinchilla published in Fig. 14 of Ruggero *et al.* (1997), phases are referenced to 80 dB SPL. The model responses are in good agreement with the measured data.

F. C1/C2 transition

Figures 11(a) and 11(c) show the measured (Fig. 1, Wong *et al.*, 1998) and model phase raster displays, respectively, for an AN fiber with a CF of 2.03 kHz to three different tones at 0.6, 1.07, and 2.03 kHz, respectively. The dura-

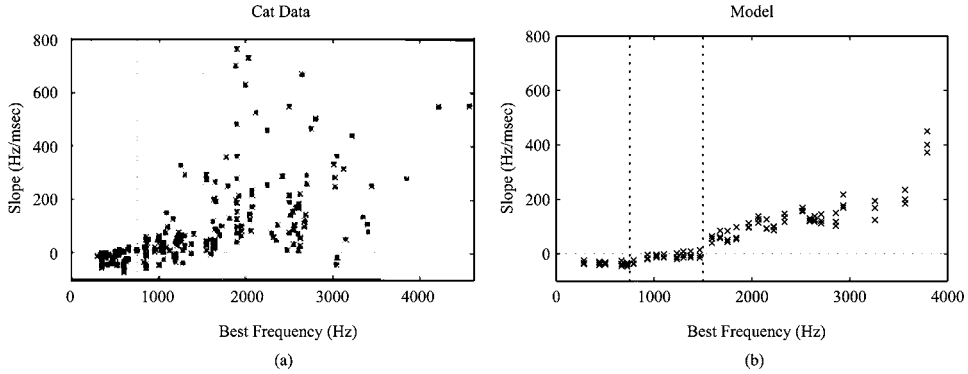


FIG. 8. Slope of the first-order regressions of IF trajectories of the cat (Carney *et al.*, 1999) and model revcor functions. In each case, the best frequency for each fiber was determined by averaging the estimated best frequencies across the SPLs studied. Vertical dotted lines at 750 and 1500 Hz are used to demarcate the trend in the IF glide over different frequency regions. (a) Measured data from 214 fibers in cats across a wide range of SPLs (reprinted from Fig. 7 of Carney *et al.*, 1999, with permission from the Acoustical Society of America©) and (b) model IF slope for three different sound pressure levels at 40, 60 and 80 dB SPL, respectively, for each fiber.

tion of the stimulus tone is 400 ms and 4 repetitions of the same stimulus are used to plot the period histogram, to be able to compare with the results published in Wong *et al.* (1998). The first 20 ms in each repetition is excluded when constructing the period histogram to avoid the transient response at the stimulus onset.

The Fourier transform of the period histogram gives the phase of the stimulus at which the fiber is responding and the plot of phase versus level is constructed by determining the phase for each level of the stimulus at that frequency. The phase at the lowest level is arbitrarily set as 0°, because it is the change in phase that is significant. The level at which the phase shift just exceeds 90° is referred to as the C2 threshold (Wong *et al.*, 1998). Below each phase raster plot is the corresponding phase versus level plot. Measured and model responses are shown in Figs. 11(b) and 11(d), respectively. Like the data, model response shows the expected pattern of phase with level. For the tones below CF, the C1/C2 transition occurs at relatively higher levels of the stimulus than that for the tone at CF, which is consistent with the data. The model responses here show slightly more rapid transitions than the example AN data, but there exists substantial vari-

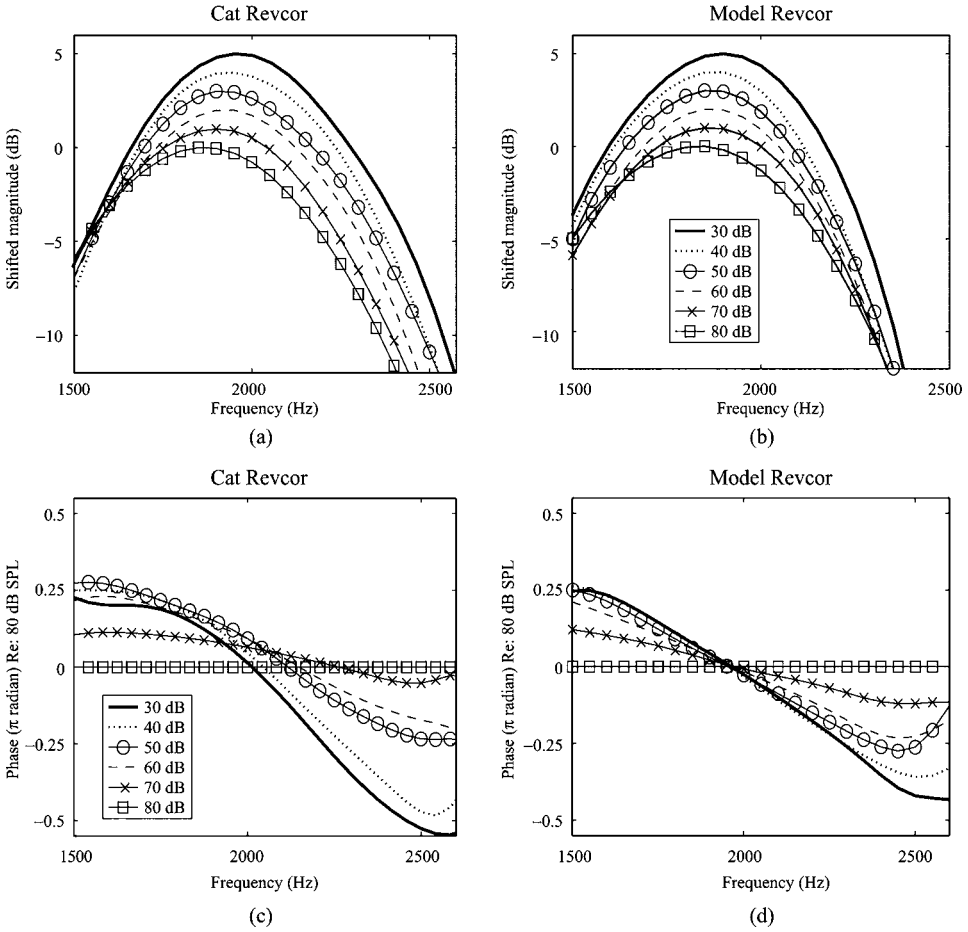


FIG. 9. Revcor filters: (a), (c) measured (unit 86100-25, 1914 Hz, Carney and Yin, 1988) and (b), (d) model responses. Upper panels (a), (b) show the magnitude of the filters computed in response to a wide-band noise at several stimulus levels. Each revcor filter is normalized by its peak magnitude, and for clarity a 1 dB shift is introduced between filters at different levels. Lower panels (c), (d) show the level-dependent phase response of the revcor filters where the phases are forced to be between $-\pi$ and $+\pi$. In both cases, the phases vary systematically with stimulus level for frequencies above and below CF.

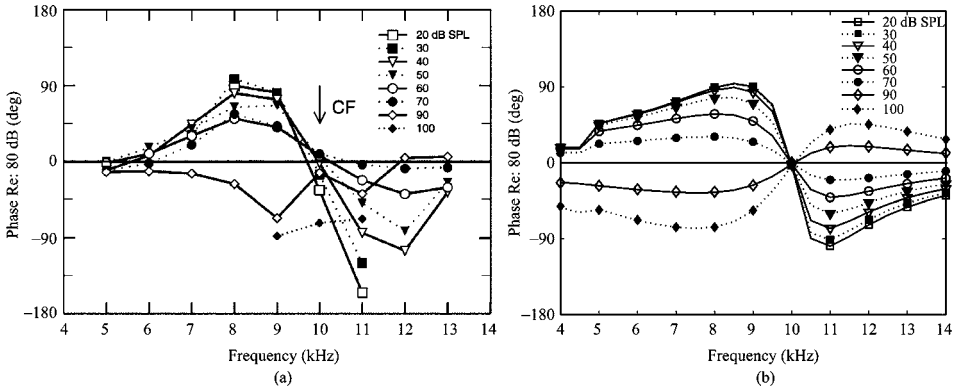


FIG. 10. (a) Measured (reprinted from Fig. 14 of Ruggero *et al.*, 1997, with permission from the Acoustical Society of America©) and (b) model phase responses of an AN fiber (CF = 10 kHz) at the level of the BM. The model phases are measured from the output of the C1 filter. Phases are referenced to the phase of the response at each frequency at 80 dB SPL, as in Ruggero *et al.* (1997).

ability in the steepness *and direction* of the phase transition across AN fibers (Wong, 1998; Heinz, private communication). Similar variability in phase transition behavior is seen in the model predictions depending on the frequency of the tone relative to CF and the parameters for the C2 IHC transduction function.

G. Peak splitting

Figure 12 shows the period histograms for a single AN fiber with CF of 1.69 kHz responding to five different continuous tones at 0.25, 0.5, 1.0, 1.69, and 2.0 kHz, respectively. For comparison with the published data (Fig. 6, Kiang, 1990), both measured and model responses are presented. Each histogram shows the averages for 15 s samples for two cycles of the continuous tone. At low tone levels, the histogram is flat during spontaneous activity and a single peak per stimulus cycle gradually emerges with the peak shifted slightly with increasing sound pressure level. At high levels, the phase shift is as high as 180°, corresponding to the switch from C1 to C2 responses. In addition, extra peaks appear in the histograms of both measured and model responses to tones below 1 kHz only. Note that the phase of

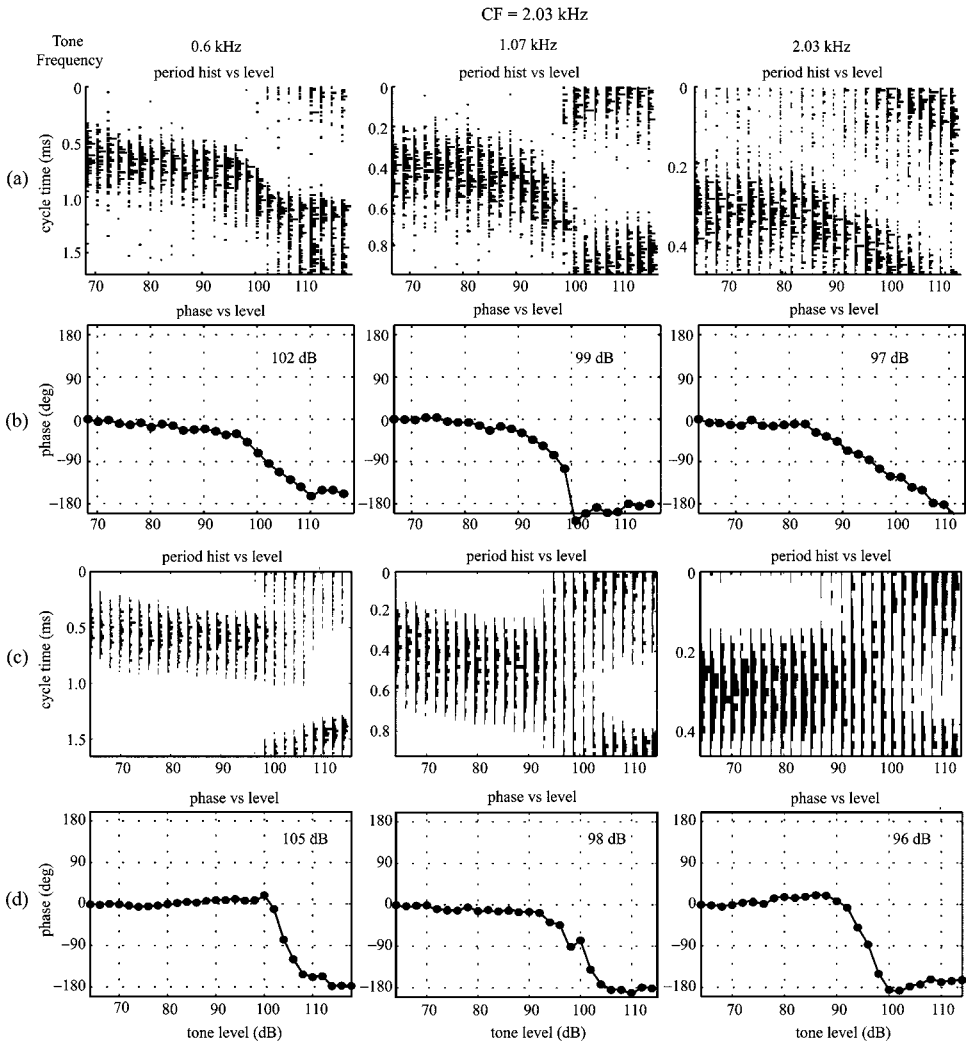


FIG. 11. Measured (reprinted from Fig. 1 of Wong *et al.*, 1998©, with permission from Elsevier) and model responses of an AN fiber with CF at 2.03 kHz in response to three different continuous tones at 0.6, 1.07, and 2.03 kHz, respectively. Phase raster plots in (a) and (c) (corresponding to measured and model responses, respectively) display period histograms along the ordinate at various sound levels. Three columns show data for three different tone frequencies. (b), (d) The average phases of the tone in the measured and model responses, respectively, vs stimulus level. The level at which phase shift just exceeds 90° (C1/C2 threshold) is shown in the upper right corner in the average phase plots.

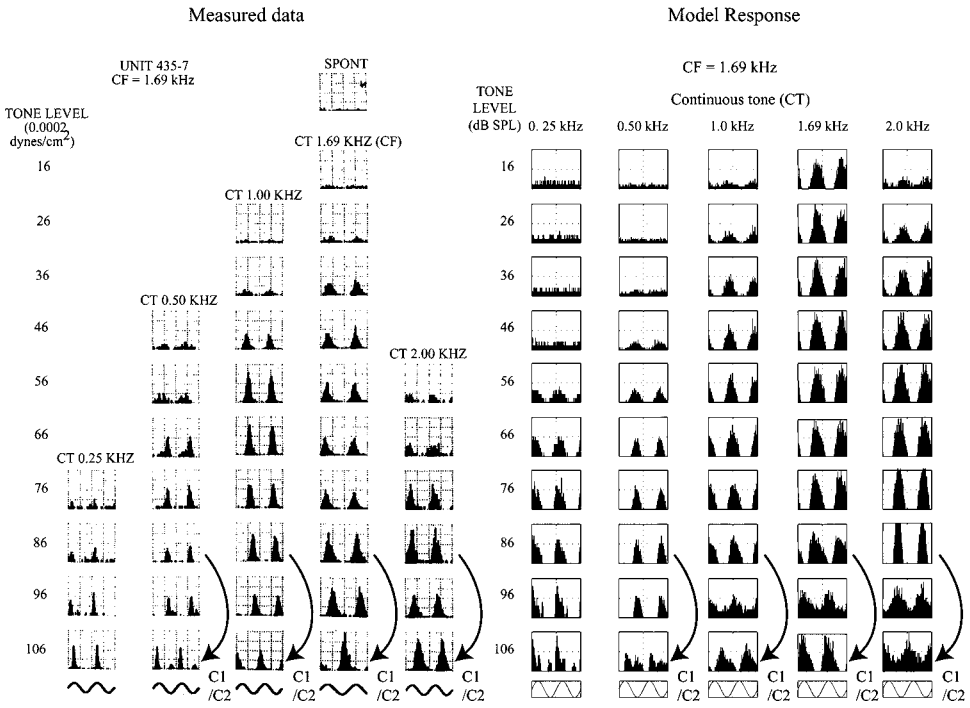


FIG. 12. Period histograms for a single AN fiber with CF at 1.69 kHz in response to five different continuous tones at 0.25, 0.50, 1.0, 1.69, and 2.0 kHz, respectively. The left panel shows the reported data (reprinted from Fig. 6 of Kiang, 1990©, with permission from Elsevier) and the right panel shows the model responses. Curved arrows labeled "C1/C2" indicate where C1/C2 transitions are observed. Responses to tones of 15 s duration measured to construct the histogram. Each column contains data for a particular tone at different sound pressure levels. For histograms in the same column, the ordinate is scaled by the same factor to give a clear pattern. However, across rows the scaling factor varies. The abscissa for each column is the time period corresponding to two cycles of that particular tone.

the AN model response relative to the tone phase does not always match the relative phase observed in the corresponding data. The phase response of the model is affected by the delay function described in Sec. II F, which is set to match the average delay observed in the revcor data at each CF rather than attempting to match the phase data for tones.

Figure 13 shows the histograms for both measured (Fig. 7, Kiang, 1990) and model AN fiber (CF at 0.240 kHz) responses to a tone at CF. For comparison the histograms are shown for the levels from 90 to 104 dB SPL with 2 dB steps. At around 92 dB SPL, a separate smaller peak appears in the model responses that is almost opposite in phase to the primary peak. With the increase in tone level, the small secondary peak grows, and around 98 dB SPL there are two

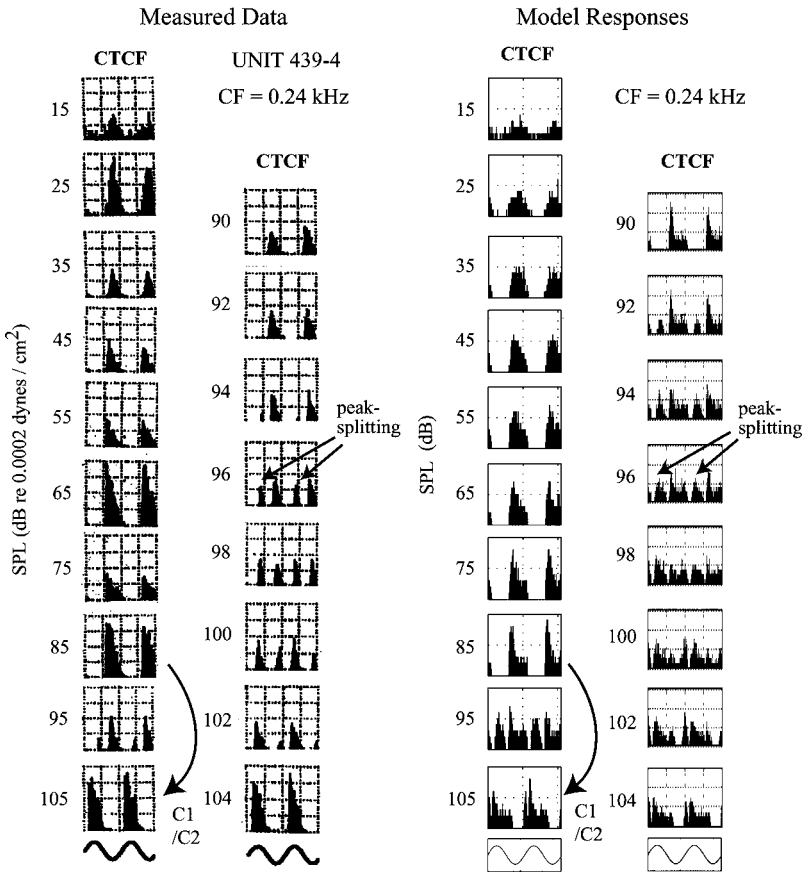


FIG. 13. Period histograms for a single AN fiber with CF at 0.240 kHz responding to a continuous tone at CF. Both measured (modified from Fig. 7 of Kiang, 1990©, with permission from Elsevier) and model responses are shown. The tone duration for constructing the histograms is 15 s. In each case, the left column contains data in 10 dB steps with increasing sound pressure levels downwardly and the right column shows data in 2 dB steps. Curved arrows labeled C1/C2 indicate where C1/C2 transitions are observed. Straight arrows labeled "peak splitting" indicate an example histogram where the emergence of strong harmonic peaks is observed. For histograms in the same column, the ordinate is scaled by the same factor to give a clear pattern. The abscissa for each column is the time period corresponding to two cycles of that particular tone.

distinctly spaced separate peaks per stimulus cycle. With further increase in tone level, the secondary peak dominates and the primary peak diminishes completely, i.e., there is again only a single peak per cycle and thus the overall response is shifted to the opposite phase relative to the phase at low levels. Overall, model responses are in good agreement with the reported data, although the model responses still show some spread of synchrony above 90 dB SPL. This is due to the C1 and C2 outputs being nearly but not exactly 180° out of phase at these stimulus levels. At higher levels (not shown), the C1 and C2 outputs become exactly 180° out of phase and the model fiber again shows synchrony over a narrow duration of the stimulus cycle.

IV. DISCUSSION

This paper presents a computational AN model that gives a phenomenological description of the auditory-periphery for a wide range of CFs across various sound levels. The signal-path C1 filter together with the feed-forward control path reflects the important response properties of the AN at low and moderate levels, such as compression, two-tone rate suppression, level-dependent phase responses, level-independent IF glides in the impulse response and BF shift with sound pressure level. The interaction between the two paths, C1 and C2, explains most of the transition region (90–100 dB SPL) effects, such as the C1/C2 transition and peak splitting, that are the main focus of this modeling study. For levels above the transition region, the C2 output dominates and is responsible for the very high-level effects in the AN fiber responses, such as the loss of synchrony capture due to poor frequency selectivity and BF shift (e.g., synchrony goes to lower formant to speech responses, Wong *et al.*, 1998), and lack of vulnerability to acoustic trauma (Liberman and Kiang, 1984; Wong *et al.*, 1998).

A. Comparison with previous models

As discussed in Sec. I, some previous studies have considered implementing Kiang's two-factor cancellation hypothesis in models of BM and IHC processing. Cheatham and Dallos (1998) built an *ad hoc* mathematical model to demonstrate the transition region effects at the level of the IHC (Fig. 9, Cheatham and Dallos, 1998). They presumed that the two factors interact at the input to the IHC transducer. A C1 response having both fundamental and second harmonic components is added with a C2 response having only a fundamental component at the input to the IHC, *before the transduction process*. A similar approach was taken by Mountain and Cody (1999), with three modes of excitation adding before IHC transduction. The architecture of the model presented in this paper is slightly different in the sense that the two paths interact *after* they are transduced by two separate functions. The justification of having two separate filters followed by their corresponding transduction functions was discussed earlier.

Lin and Goldstein (1995) investigated inclusion of a C1/C2 transition mechanism in their multiple-bandpass-nonlinearity (MBPNL) model. At low levels, the MBPNL model response is dominated by the high-gain band-pass tip path, whereas at very high levels, the low-pass tail path dominates because of the expander. The two paths interact nonlinearly within a transition region. To simulate the C1/C2 transition and peak-splitting phenomena in the MBPNL model, the two paths have to be in anti-phase relation, consistent with the Kiang's two-factor hypothesis (Lin, 1994; Lin and Goldstein, 1995). However, for predicting responses to paired clicks (Goblick and Pfeiffer, 1969), it is required that the tip path leads the tail path by 90°. As the MBPNL model is static, there is no change in the filter phases over time that could explain the different phase behavior for continuous tone pairs versus paired clicks. Thus Lin and Goldstein (1995) proposed an *ad hoc* phase adaptation mechanism that produces a temporal transition of the two paths' phase responses from an initial quadrature relation to a final anti-phase relation, with a time constant of 0.2 ms. In contrast, the model presented in this paper does not require any *ad hoc* phase adaptation, because phase adaptation is a natural result of the implementation of the cochlear amplifier in this model. The time-varying C1 filter dynamically adapts its phase at low and moderate levels according to the control-path signal; a low-pass filter in the control path produces an approximately 0.2 ms time constant for the onset of the compressive nonlinearity (Fig. 8, Recio *et al.*, 1998) and the resulting phase changes. Thus, for paired clicks the anti-phase relationship of the C1 and C2 outputs does not have time to develop because of the ~0.2 ms time constant for the onset of C1 gain and phase changes. However, for continuous tones the anti-phase relationship of the C1 and C2 outputs has time to develop and consequently produces the C1/C2 transition and peak-splitting phenomena in the model.

The model presented here will be useful in simulating responses to complex stimuli, particularly for speech. It has been reported that all frequency components of the response to a vowel stimulus undergo the C1/C2 transition level simultaneously and this level is determined by the level at which the largest component reaches the C1/C2 transition, rather than the level of an individual harmonic component of the vowel (Fig. 3, Wong *et al.*, 1998). This observation rules out the possibility of separate processing of each component of complex stimuli, otherwise individual components might undergo C1/C2 transitions separately. In the proposed model, the C1/C2 transition occurs when the C1 filter gain and phase match the C2 filter gain and phase *and* the output of the C2 transduction function grows to have a similar magnitude to the output of the C1 transduction function. The model's wide-band control path determines how the gain and phase of the C1 filter approach those of the C2 filter at high levels; the control path behavior is dependent on the overall spectrum of a wide-band stimulus, rather than just one frequency component, and consequently all vowel components should undergo the C1/C2 transition simultaneously. It is unlikely that the phase-transition-modified MBPNL model (Lin and Goldstein, 1995) could explain the vowel data.

B. Physical interpretation of the model architecture

While the proposed model architecture was created in such a way to be compatible with as much of the physiologi-

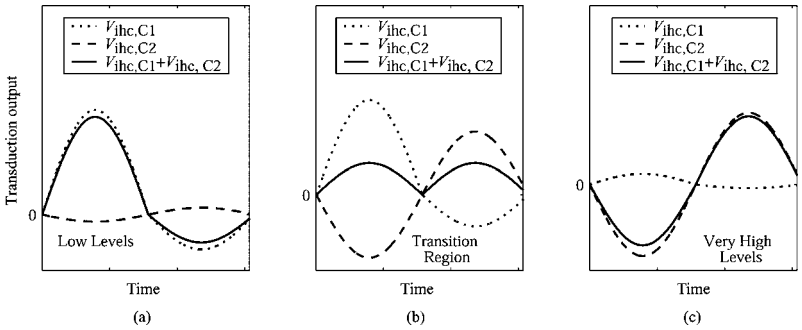


FIG. 14. Illustration of how the C1/C2 transition and peak-splitting phenomena occur in the IHC section of the model. The dotted, dashed, and solid lines represent the output of the C1 transducer, C2 transducer, and the combined responses, respectively. Left, middle, and right panels give representative responses before the low-pass filter of the IHC at low, transition, and very high levels, respectively.

cal data as possible, the C2 filter and transduction function behavior were chosen to provide a good phenomenological description of the AN fiber data rather than being based on a particular interpretation of the physical source of the C2 response. We did investigate alternative model architectures such as omitting the C2 filter and just having the C1 filter provide the input to the two separate C1 and C2 transduction functions, but none of these other approaches produced satisfactory results. The apparent need for a separate C2 filter suggests that the physical source of the C2 response is a second traveling wave on the BM, as has been proposed by Lin and Guinan (2004). The fact that our C2 filter is identical to our C1 filter with complete OHC impairment indicates that this second traveling wave could be subject only to the passive mechanical properties of the cochlea, i.e., is not affected by the cochlear amplifier, which has been suggested previously by Lin and Guinan (2004). To explain the difference between response properties in the tip and tail regions of AN fiber tuning curves, Lin and Guinan (2000) proposed that there may be two modes of excitation in the tip region and another two in the tail region. With our model architecture, the differences in the response properties within the tip and tail regions result from the difference in the control path behavior for stimuli within the tip and tail regions. That is, our phenomenological description of the cochlear amplifier achieved by the control path's influence on the C1 filter tuning can explain the tip and tail differences, and only a single C2 filter is required as well.

A physical interpretation of the separate C2 transduction function is more difficult. While Liberman and Kiang (1984) observed that the C2 response and the shorter IHC stereocilia are both fairly resistant to acoustic trauma, it is unclear why transduction through the shorter stereocilia should not be affected by a reduction in the EP (Sewell, 1984b). It is also not obvious why the C2 transduction pathway would only be driven by the C2 excitation from the BM and not from the C1 excitation. Cheatham and Dallos (1998) speculated on the possibility of a change in the coupling of the IHC stereocilia with the tectorial membrane as a result of a change in the magnitude of basilar membrane vibrations, but they did not consider the existence of two different modes of basilar membrane vibration caused by two traveling waves. Further investigation with biophysical models of cochlear micromechanics and IHC transduction is clearly required.

C. Interpretation of the peak-splitting results

The peak-splitting phenomenon has been reported in cats mostly for low CFs in response to stimulus frequencies below 1 kHz (Kiang and Moxon, 1972; Johnson, 1980; Kiang, 1984, 1990; Cai and Geisler, 1996). In contrast, it has been found in chinchilla that high-CF fibers are more likely to show peak splitting than lower CFs (Ruggero and Rich, 1983, 1989). However, irrespective of species, the occurrence of peak splitting, the number of peaks, and the distances between the peaks in the period histogram of the AN response are dependent upon the stimulus frequency (Cai, 1995; Cai and Geisler, 1996). Two types of peak splitting have been observed in the data. One is symmetrical with the two peaks being approximately one half cycle apart; the range of sound pressure levels over which it has been observed is narrow, and it is almost always accompanied by abrupt changes in the phase-level function and occasionally by the notches in the rate-level function (Kiang and Moxon, 1972; Kiang, 1984, 1990). The other is asymmetrical where peaks are unevenly spaced and more than two peaks may appear in the histogram. The peaks maintain their existence during a fairly large range of stimulus sound pressure levels and thus the multiple peaks do not seem to be related in any way to notches in the rate-level functions (Cai and Geisler, 1996). Both types of peak splitting were observed in several species. However, the exact mechanism behind the generation of these responses is not known yet.

As stated in Sec. I, Kiang (1990) hypothesized that if one of the two responses (C1 or C2) contains a second harmonic distortion, peak splitting could be a direct result of the C1/C2 interaction. He argued that in the transition region the fundamental components of the C1 and C2 responses will cancel each other, but the second-harmonic component will remain uncancelled. In this paper, we have described a model consistent with this hypothesis that quantitatively addresses these issues. The C1 transduction function is a saturating, asymmetrical nonlinearity, and hence its output has strong harmonic distortions and a dc offset at high stimulus levels. However, the symmetrical C2 transduction function has a strong fundamental component with no dc offset and insignificant harmonic distortions in the transition region.

Figure 14 shows example response waveforms obtained at the outputs of the two transduction functions in our model IHC, with the C1 output shown by dotted lines, the C2 output shown by dashed lines, and their summed output shown by solid lines. At low and moderate levels [Fig. 14(a)], the summed output is completely dominated by the C1 responses and thus the period histogram exhibits only a single peak in the AN responses. Within the transition region [Fig. 14(b)], the C1 and C2 amplitudes are comparable and the fundamental-frequency components tend to cancel each

other. However, the second harmonic of the C1 response (produced by the C1 IHC transduction function asymmetry) remains uncancelled, and consequently the period histogram in the transition region exhibits two peaks per cycle (peak splitting) corresponding to the two positive peaks in the combined output. Figure 14(c) illustrates that as the growth of the C2 response is sharp, the primary (C1) peak is cancelled completely just a few decibels above the transition threshold, at which point the anti-phase C2 output dominates.

In the model presented here, the combined output of the two transduction functions is passed through the seventh-order IHC low-pass filter with a cut-off frequency of 3800 Hz to describe the fall-off in pure tone synchrony above 1 kHz (Zhang *et al.*, 2001). Consistent with the experimental observations, the model histogram shows symmetrical peak splitting to tones below 1 kHz only. For tone frequencies above 1 kHz, the second harmonic will have a frequency above 2 kHz and consequently will be greatly attenuated by the IHC low-pass filter. In this case, the period histogram will have a phase transition with no obvious peak splitting. However, the dc component of the C1 response (produced by the C1 transduction function asymmetry) will pass through the IHC filter, such that a broadening of the period histogram will occur and no notch will occur in the rate-level function, as is seen in Figs. 6(a) and 6(b).

From the above-mentioned observations, it is apparent that a notch in the rate-level function could be produced in either of two circumstances. First, if an IHC has a C1 transduction function that is more linear than normal, then no second-harmonic or dc offset will be produced in the output. In this case, the C1 response will be more completely cancelled by the anti-phase C2 response, and the AN fiber discharge rate will drop, producing a notch in the rate-level function until the C2 response grows large enough to dominate the response. Second, if an IHC has a C2 transduction function that is more asymmetric than normal, then the C2 response may have harmonic distortions and a dc offset similar to the C1 response, such that they can cancel each other and likewise produce a notch in the rate-level function. The former case is illustrated in the model simulations with IHC impairment shown in Fig. 6(c). A notch appears in the rate-level function at 75 dB SPL with this implementation of IHC impairment, and an abrupt change of phase is observed in the phase raster. Impairment modeled as a shallower slope in the C1 transduction function results in less rectification of the C1 waveform. Consequently, in the transition region the fundamental frequency component dominates in the C1 waveform and tends to cancel with the anti-phase C2 waveform, and the lack of harmonic distortion or dc offset in the C1 results in the notch in the rate-level function and the absence of peak splitting in the phase raster. Variability in C1 and C2 transduction function shape between different IHCs in experimental animals could consequently explain why notches of different depths occur in different AN fiber rate-level functions, and why the peak-splitting phenomenon is only strong in rate-level functions without notches (Kiang, 1990).

D. C1/C2 transitions in high-CF fibers

It has been reported that as CF increases there is a narrower range of stimulus frequencies that can elicit an abrupt phase shift in the cat AN responses (Fig. 3, Liberman and Kiang, 1984). In contrast, in chinchilla the abrupt phase transitions are seen in high-CF fibers to even very low stimulus frequencies (Ruggero *et al.*, 1996). However, Lin and Guinan (2000) reported that in high-CF cat AN responses to clicks, a phase reversal of the short-latency peak occurs at high levels [Fig. 10(F) of Lin and Guinan, 2000]. Presumably, the short-latency peaks in high-CF fibers are due to excitation in the tuning-curve tail (Schoonhoven *et al.*, 1994). So, the short-latency click responses can be compared to the tail-frequency tone responses. The fact that the phase reversals in cat high-CF fibers have been seen with clicks, but not with tones, might be due to the dearth of cat data for stimuli above 100 dB SPL. Clicks can be applied at very high sound levels, but tones produce acoustic trauma at these high levels. If this is the case, then in response to low-frequency tones the C1/C2 transition will be observed for high-CF fibers at very high levels in cats, and thus, chinchilla and cat would be similar in this regard. In this study, the C2 transduction function sharply increases the C2 output at high levels and consequently, in high-CF fibers the combined output is dominated by the C2 response even for low-frequency tones. So, the model can simulate phase transitions at very high levels for high-CF fibers to low-frequency tones.

V. CONCLUSION

The proposed model gives good quantitative and qualitative description of AN response properties for tones and broadband noise stimuli over a wide range of stimulus levels. While the model architecture is designed primarily to give a good phenomenological description of the AN response properties, it does give some insight into the possible physical mechanisms in the cochlea generating the C2 response. Investigation of model predictions for high-level clicks (e.g., Lin and Guinan, 2000) and complex stimuli, particularly speech (e.g., Wong *et al.*, 1998), should provide useful further evaluation of the model.

ACKNOWLEDGMENTS

The authors thank Laurel Carney, Michael Heinz, and Xuedong Zhang for helpful discussions during development of the model and comments on earlier versions of the manuscript. The suggestions of two anonymous reviewers were invaluable in improving this manuscript. This research was supported by NSERC Discovery Grant No. 261736 and the Barber-Gennum Chair Endowment.

[1]In this paper, best frequency (BF) is the frequency at which the fiber response is maximum, which can vary with sound level, whereas the characteristic frequency (CF) is the frequency at which the fiber shows the lowest threshold, which is a model parameter that is invariant to the stimulus sound pressure level. Note however that the *estimate of CF* that is measured from a model tuning curve may change as a function of hair cell impairment (see Fig. 5), although the model parameter CF has not changed.

[2]See Appendix B of Bruce *et al.* (2003). Note that there is an error in Eq. (B1) of Bruce *et al.* (2003). The corrected form of the equation is

$$\text{gain}_{cp}[n] = (\{1 + c1_{LP}[n]^2 - 2c1_{LP}[n]\cos((\omega_{cp} - \omega_{CF})/F_s)\}/\{2c2_{LP}[n]^2(1 + \cos((\omega_{cp} - \omega_{CF})/F_s))\})^{1/2}. \quad (23)$$

[3]Due to the lack of data on frequency glides for CFs above 4 kHz, model parameters for high CF fibers are determined using Q_{10} and threshold data from Miller *et al.* (1997).

Anderson, D. J., Rose, J. E., Hind, J. E., and Brugge, J. F. (**1971**). "Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: Frequency and intensity effects," J. Acoust. Soc. Am. **49**, 1131–1139.

Bondy, J., Becker, S., Bruce, I., Trainor, L., and Haykin, S. (**2004**). "A novel signal-processing strategy for hearing-aid design: Neurocompensation," Signal Process. **84**, 1239–1253.

Brown, G. J., and Cooke, M. (**1994**). "Computational auditory scene analysis," Comput. Speech Lang. **8**, 297–336.

Bruce, I. C. (**2004**). "Physiological assessment of contrast-enhancing frequency shaping and multi-band compression in hearing aids," Physiol. Meas **25**, 945–956.

Bruce, I. C., Sachs, M. B., and Young, E. D. (**2003**). "An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses," J. Acoust. Soc. Am. **113**, 369–388.

Cai, Y. (**1995**). "Temporal responses of the auditory-nerve fibers to single-tone and two-tone stimuli: Experimental and modeling studies," Ph.D. thesis, University of Wisconsin-Madison.

Cai, Y., and Geisler, C. D. (**1996**). "Temporal patterns of the responses of auditory-nerve fibers to low-frequency tones," Hear. Res. **96**, 83–93.

Carney, L. H. (**1993**). "A model for the responses of low-frequency auditory-nerve fibers in cat," J. Acoust. Soc. Am. **93**, 401–417.

Carney, L. H. (**1994**). "Spatiotemporal encoding of sound level: Models for normal encoding and recruitment of loudness," Hear. Res. **76**, 31–44.

Carney, L. H., McDuffy, M. J., and Shekhter, I. (**1999**). "Frequency glides in the impulse responses of auditory-nerve fibers," J. Neurophysiol. **105**, 2384–2391.

Carney, L. H., and Yin, T. C. T. (**1988**). "Temporal coding of resonances by low-frequency auditory nerve fibers: Single-fiber responses and a population model," J. Neurophysiol. **60**, 1653–1677.

Cheatham, M. A., and Dallos, P. (**1989**). "Two-tone suppression in inner hair responses," Hear. Res. **40**, 187–196.

Cheatham, M. A., and Dallos, P. (**1998**). "The level dependence of response phase: Observations from cochlear hair cells," J. Acoust. Soc. Am. **104**, 356–369.

Cooper, N. P., and Rhode, W. S. (**1992**). "Basilar membrane mechanics in the hook region of cat and guinea-pig cochlae: Sharp tuning and nonlinarity in the absence of baseline position shifts," Hear. Res. **63**, 163–190.

Cooper, N. P., and Rhode, W. S. (**1997**). "Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea," J. Neurophysiol. **78**, 261–270.

Dallos, P. (**1985**). "Response characteristics of mammalian cochlear hair cells," J. Neurosci. **5**, 1591–1608.

Dallos, P. (**1986**). "Neurobiology of cochlear inner and outer hair cells: Intracellular recordings," Hear. Res. **22**, 185–198.

de Boer, E., and de Jongh, H. R. (**1978**). "On cochlear encoding: Potentialities and limitations of the reverse correlation technique," J. Acoust. Soc. Am. **63**, 115–135.

de Boer, E., and Kuyper, P. (**1968**). "Triggered correlation," IEEE Trans. Biomed. Eng. **15**, 169–179.

de Boer, E., and Nuttall, A. N. (**1997**). "The mechanical waveform of the basilar membrane. I. Frequency modulations (glides) in impulse responses and cross-correlation functions," J. Acoust. Soc. Am. **101**, 3583–3592.

de Boer, E., and Viergever, M. A. (**1982**). "Validity of the Liouville-Green (or WKB) method for cochlear mechanics," Hear. Res. **8**, 131–155.

Delgutte, B. (**1990**). "Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level," Hear. Res. **49**, 225–246.

Deng, L., and Geisler, C. D. (**1987**). "A composite auditory model for processing speech sounds," J. Acoust. Soc. Am. **82**, 2001–2012.

Evans, E. F. (**1981**). "The dynamic range problem: Place and timing coding at the level of the cochlear nerve and nucleus," in *Neuronal Mechanisms of Hearing*, edited by J. Syka and L. Aitkin (Plenum, New York), pp.69–95.

Geisler, C. D., and Rhode, W. S. (**1982**). "The phases of basilar-membrane vibrations," J. Acoust. Soc. Am. **71**, 1201–1203.

Ghitza, O. (**1988**). "Temporal non-place information in the auditory-nerve firing patterns as a front-end for speech recognition in a noisy environment," J. Phonetics **16**, 109–123.

Gifford, M. L., and Guinan, J. J., Jr. (**1983**). "Effects of crossed-olivocochlear bundle stimulation on cat auditory-nerve fiber response to tones," J. Acoust. Soc. Am. **74**, 115–123.

Giguère, C., and Woodland, P. C. (**1994**). "A computational model of the auditory periphery for speech and hearing research. I. Ascending path," J. Acoust. Soc. Am. **95**, 331–342.

Goblick, T. J., and Pfeiffer, R. R. (**1969**). "Time-domain measurements of cochlear nonlinearities using combination-click stimuli," J. Acoust. Soc. Am. **46**, 924–938.

Goldstein, J. L. (**1990**). "Modeling rapid waveform compression on the basilar membrane as multiple-bandpass-nonlinearity filtering," Hear. Res. **49**, 39–60.

Goldstein, J. L. (**1995**). "Relations among compression, suppression, and combination tones in mechanical responses of the basilar membrane: Data and MBPNL model," Hear. Res. **89**, 52–68.

Heinz, M. G. (private communication).

Heinz, M. G., and Young, E. D. (**2004**). "Response growth with sound level in auditory-nerve fibers after noise-induced hearing loss," J. Neurophysiol. **91**, 784–795.

Hewitt, M. J., and Meddis, R. (**1992**). "Regularity of cochlear neucleus stellate cells: A computational modeling study," J. Acoust. Soc. Am. **93**, 3390–3399.

Holley, M. C. (**1996**). "Outer hair cell motility," in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 386–434.

Johnson, D. (**1980**). "The relationship between spike rate and synchrony in responses to auditory-nerve fibers to single tones," J. Acoust. Soc. Am. **68**, 1115–1122.

Kates, J. M. (**1991**). "A time-domain digital cochlear model," IEEE Trans. Signal Process. **39**, 2573–2592.

Kates, J. M. (**1995**). "Two-tone suppression in a cochlear model," IEEE Trans. Speech Audio Process. **3**, 396–406.

Kiang, N. Y., Liberman, M. C., Sewell, W. F., and Guinan, J. J., Jr. (**1986**). "Single unit clues to cochlear mechanisms," Hear. Res. **22**, 171–182.

Kiang, N. Y.-S. (**1984**). "Peripheral neural processing of auditory information," in *Handbook of Physiology, Section 1: The Nervous System*, edited by J. M. Brookhart and V. B. Mountcastle (American Physiological Society, Bethesda, MD), Vol. **III** Pt. 2, pp. 639–674.

Kiang, N. Y.-S. (**1990**). "Curious oddments of auditory-nerve studies," Hear. Res. **49**, 1–16.

Kiang, N. Y.-S., Baer, T., Marr, E. M., and Demont, D. (**1969**). "Discharge rates of single auditory-nerve fibers as a function of tone level," J. Acoust. Soc. Am. **46**, 106.

Kiang, N. Y.-S., Liberman, M. C., and Levine, R. A. (**1976**). "Auditory-nerve activity in cats exposed to ototoxic drugs and high-intensity sounds," Ann. Otol. Rhinol. Laryngol. **85**, 752–768.

Kiang, N. Y.-S., and Moxon, E. C. (**1972**). "Physiological considerations in artificial stimulation of the inner ear," Ann. Otol. Rhinol. Laryngol. **81**, 714–731.

Kiang, N. Y.-S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). "Discharge patterns of single fibers in the cat's auditory nerve," Res. Monogr. No. 35 (MIT, Cambridge, MA).

Liberman, M. C. (**1978**). "Auditory nerve response from cats raised in a low noise chamber," J. Acoust. Soc. Am. **63**, 442–455.

Liberman, M. C. (**1982**). "The cochlear frequency map for the cat: Labeling auditory-nerve fibers of known characteristic frequency," J. Acoust. Soc. Am. **72**, 1441–1449.

Liberman, M. C. (**1984**). "Single-neuron labeling and chronic cochlear pathology. I. Threshold shift and characteristic-frequency shift," Hear. Res. **16**, 33–41.

Liberman, M. C., and Dodds, L. W. (**1984**). "Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves," Hear. Res. **16**, 55–74.

Liberman, M. C., and Kiang, N. Y.-S. (**1984**). "Single-neuron labeling and chronic cochlear pathology. IV. Stereocilia damage and alterations in rate- and phase-level functions," Hear. Res. **16**, 75–90.

Liberman, M. C., and Mulroy, M. J. (**1982**). "Acute and chronic effects of acoustic trauma: Cochlear pathology and auditory nerve pathophysiology," in *New Perspectives on Noise-Induced Hearing Loss*, edited by R. P. Hamernik, D. Henderson, and R. Salvi (Raven, New York), pp. 105–135.

Lin, T. (**1994**), "Quantitative modeling of nonlinear auditory-nerve responses as two-factor interactions," Ph.D. thesis, Sever Institute of Technology, Washington University.

Lin, T., and Goldstein, J. L. (**1995**). "Quantifying 2-factor phase relations in non-linear responses from low characteristic-frequency auditory-nerve fibers," Hear. Res. **90**, 126–138.
Lin, T., and Guinan, J. J., Jr. (**2000**). "Auditory-nerve-fiber responses to high-level clicks: Interference patterns indicate that excitation is due to the combination of multiple drives," J. Acoust. Soc. Am. **107**, 2615–2630.
Lin, T., and Guinan, J. J., Jr. (**2004**). "Time-frequency analysis of auditory-nerve-fiber and basilar-membrane click responses reveal glide irregularities and non-characteristic-frequency skirts," J. Acoust. Soc. Am. **116**, 405–416.
Lopez-Poveda, E. A. (**2005**). "Spectral processing by the peripheral auditory system: Facts and models," Int. Rev. Neurobiol. **70**, 7–48.
Matthews, J. W. (**1983**). "Modeling reverse middle ear transmission of acoustic distortion signals," in *Mechanics of Hearing: Proceedings of the IUTAM/ICA Symposium*, edited by E. de Boer and M. A. Viergever (Delft University Press, Delft), pp. 11–18.
Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (**2001**). "A computational algorithm for computing nonlinear auditory frequency selectivity," J. Acoust. Soc. Am. **109**, 2852–2861.
Miller, R. L., Calhoun, B. M., and Young, E. D. (**1999**). "Contrast enhancement improves the representation of /ε/-like vowels in the hearing-impaired auditory nerve," J. Acoust. Soc. Am. **106**, 2693–2708.
Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (**1997**). "Effects of acoustic trauma on the representation of the vowel /ε/ in cat auditory nerve fibers," J. Acoust. Soc. Am. **101**, 3602–3616.
Møller, A. R. (**1977**). "Frequency selectivity of single auditory-nerve fibers in response to broad-band noise stimuli," J. Acoust. Soc. Am. **62**, 135–142.
Mountain, D. C., and Cody, A. R. (**1999**). "Multiple modes of inner hair cell stimulation," Hear. Res. **132**, 1–14.
Narayan, S. S., Temchin, A. N., Recio, A., and Ruggero, M. A. (**1998**). "Frequency tuning of basilar membrane and auditory nerve fibers in the same cochleae," Science **282**, 1882–1884.
Nuttall, A. L., and Dolan, D. F. (**1993**). "Two-tone suppression of inner hair cell and basilar membrane responses in the guinea pig," J. Acoust. Soc. Am. **94**, 3511–3514.
Palmer, A. R., and Russell, I. J. (**1986**). "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells," Hear. Res. **24**, 1–15.
Patuzzi, R. B. (**1996**). "Cochlear micromechanics and macromechanics," in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 186–257.
Patuzzi, R. B., and Robertson, D. (**1988**). "Tuning in the mammalian cochlea," Physiol. Rev. **68**, 1009–1082.
Peake, W. T., Rosowski, J. J., and Lynch, T. J., III (**1992**). "Middle-ear transmission: Acoustic versus ossicular coupling in cat and human," Hear. Res. **57**, 245–268.
Recio, A., Narayan, S. S., and Ruggero, M. A. (**1997**). "Wiener-kernel analysis of basilar-membrane response to white noise," in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 325–331.
Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). "Basilar-membrane responses to clicks at the base of the chichilla cochlea," J. Acoust. Soc. Am. **103**, 1972–1989.
Rhode, W. S. (**1971**). "Observations of the vibration of the basilar membrane in squirrel monkeys using the mössbauer technique," J. Acoust. Soc. Am. **49**, 1218–1231.
Rhode, W. S. (**1973**). "An investigation of postmortem cochlear mechanics using the Mössbauer effects," in *Basic Mechanisms in Hearing*, edited by A. R. Møller (Academic, New York), pp. 49–63.
Rhode, W. S., and Recio, A. (**2000**). "Study of mechanical motions in the basal region of the chinchilla cochlea," J. Acoust. Soc. Am. **107**, 3317–3332.
Robert, A., and Eriksson, J. L. (**1999**). "A composite model of the auditory periphery for simulating responses to complex sounds," J. Acoust. Soc. Am. **106**, 1852–1864.
Robertson, D. (**1982**). "Effects of acoustic trauma on stereocilia structure and spiral ganglion cell tuning properties in the guinea pig cochlea," Hear. Res. **7**, 55–74.
Robles, L., Rhode, W. S., and Geisler, C. D. (**1976**). "Transient response of the basilar membrane measured in squirrel monkeys using the mössbauer effect," J. Acoust. Soc. Am. **59**, 926–939.
Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.
Ruggero, M. A., and Rich, N. C. (**1983**). "Chinchilla auditory-nerve responses to low-frequency toens," J. Acoust. Soc. Am. **73**, 2096–2108.
Ruggero, M. A., and Rich, N. C. (**1989**). "Peak splitting: Intensity effects in cochlear afferent responses to low frequency tones," in *Cochlear Mechanisms: Structure, Function and Models*, edited by J. P. Wilson and D. T. Kemp (Plenum, New York), pp. 259–267.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**, 2151–2163.
Ruggero, M. A., Rich, N. C., Shivapuja, B. G., and Temchin, A. N. (**1996**). "Auditory-nerve responses to low-frequency tones: Intensity dependence," Aud. Neurosci. **2**, 159–185.
Russell, I. J., and Sellick, P. M. (**1978**). "Intracellular studies of hair cells in the mammalian cochlea," J. Physiol. (London) **284**, 261–290.
Sachs, M. B., Winslow, R. L., and Sokolowski, B. H. A. (**1989**). "A computational model for rate-level functions from cat auditory-nerve fibers," Hear. Res. **41**, 61–70.
Salvi, R., Perry, J., Hamernik, R. P., and Henderson, D. (**1982**). "Relationships between cochlear pathologies and auditory nerve and behavioral responses following acoustic trauma," in *New Perspectives on Noise-Induced Hearing Loss*, edited by R. P. Hamernik, D. Henderson, and R. Salvi (Raven, New York), pp. 165–188.
Schmiedt, R. A., Zwislocki, J. J., and Hamernik, R. P. (**1980**). "Effects of hair cell lesions on responses of cochlear nerve fibers. I. Lesions, tuning curves, two-tone inhibition, and responses to trapezoidal-wave patterns," J. Neurophysiol. **43**, 1367–1389.
Schoonhoven, R., Keijzer, J., Versnel, H., and Prijs, V. F. (**1994**). "A dual filter model describing single-fiber responses to clicks in the normal and noise-damaged cochlea," J. Acoust. Soc. Am. **95**, 2104–2121.
Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). "Measurement of basilar membrane motion in the guinea pig using Mössbaeur technique," J. Acoust. Soc. Am. **72**, 131–141.
Sewell, W. F. (**1984a**). "The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning curves," Hear. Res. **14**, 305–314.
Sewell, W. F. (**1984b**). "Furosemide selectively reduces one component in rate-level functions from auditory-nerve fibers," Hear. Res. **15**, 69–72.
Shera, C. A. (**2001**). "Frequency glides in click responses of the basilar membrane and auditory nerve: Their scaling behavior and origin in traveling-wave dispersion," J. Acoust. Soc. Am. **109**, 2023–2034.
Sumner, C. J., O'Mard, L. P., Lopez-Poveda, E. A., and Meddis, R. (**2003**). "A non-linear filter-bank model of the guinea-pig cochlear nerve: Rate responses," J. Acoust. Soc. Am. **113**, 3264–3274.
Tan, Q., and Carney, L. H. (**2003**). "A phenomenological model for the responses of the auditory-nerve fibers. II. Nonlinear tuning with a frequency glide," J. Acoust. Soc. Am. **114**, 2007–2020.
Tchorz, J., and Kollmeier, B. (**1999**). "A model of auditory perception as a front end for automatic speech recognition," J. Acoust. Soc. Am. **106**, 2040–2050.
van der Heijden, M., and Joris, P. X. (**2003**). "Cochlear phase and amplitude retrieved from the auditory-nerve at arbitrary frequencies," J. Neurosci. **23**, 9194–9198.
Westerman, L. A., and Smith, R. L. (**1988**). "A diffusion model of the transient response of the cochlear inner hair cell synapse," J. Acoust. Soc. Am. **83**, 2266–2276.
Wilson, B. S., Schatzer, R., Lopez-Poveda, E. A., Sun, X., Lawson, D. T., and Wolford, R. D. (**2005**). "Two new directions in speech processor design for cochlear implants," Ear Hear. **26**, 73S–81S.
Wong, J. C. (**1998**). "Nonlinearities in the representation of the vowel /ε/ in cat auditory nerve at sound levels near 100 dB SPL," Master's thesis, Johns Hopkins University.
Wong, J. C., Miller, R. L., Calhoun, B. M., Sachs, M. B., and Young, E. D. (**1998**). "Effects of high sound levels on responses to the vowel /ε/ in cat auditory nerve," Hear. Res. **123**, 61–77.
Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (**2001**). "A phenomenological model for the responses of auditory-nerve fibers. I. Nonlinear tuning with compression and suppression," J. Acoust. Soc. Am. **109**, 648–670.

Temporal integration of the contralateral acoustic-reflex threshold and its age-related changes

Michele B. Emmer[a] and Shlomo Silman[b]
Department of Speech Communication Arts & Sciences and Center for Auditory Research, Brooklyn College, CUNY and Au.D. Program in Audiology, Graduate Center, CUNY, 365 Fifth Avenue, New York, NY 10016

Carol A. Silverman[c]
Au.D. Program, Graduate Center, CUNY, New York, NY; Ph.D. Program in Speech and Hearing Sciences, Graduate Center, CUNY, New York, NY; Hunter College, CUNY; Department of Otolaryngology-Head & Neck Surgery, New York Medical College, Valhalla, NY; Departments of Otolaryngology-Head & Neck Surgery and Communication Sciences, New York Eye and Ear Infirmary, New York, NY; and Communication Sciences Program, Hunter College, CUNY

Harry Levitt[d]
Ph.D. Program in Speech and Hearing Sciences, Graduate Center, CUNY, 365 Fifth Avenue, New York, NY 10016

(Received 3 August 2005; revised 10 June 2006; accepted 12 June 2006)

Although numerous studies have investigated temporal integration of the acoustic-reflex threshold (ART), research is lacking on the effect of age on temporal integration of the ART. Therefore the effect of age on temporal integration of the ART was investigated for a broad-band noise (BBN) activator. Subjects consisted of two groups of adults with normal-hearing sensitivity: one group of 20 young adults (ten males and ten females, ages 18–29 years, with a mean age of 24 years) and one group of 20 older adults (ten males and ten females, ages 59–75 years, with a mean age of 67.5 years). Activating stimulus durations were 12, 25, 50, 100, 200, 300, 500, and 1000 ms. Significant main effects for duration and age were obtained. That is, as the duration increased, the acoustic reflex threshold for BBN decreased. The interactions of duration × age group and duration × hearing level were not significant. The result of pair-wise analysis indicated statistically significant differences between the two age groups at durations of 20 ms and longer. The observed age effect on temporal integration of the ART for the BBN activator is interpreted in relation to senescent changes in the auditory system. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2221415]

PACS number(s): 43.64.Ha, 43.64.Ld [WPS] Pages: 1467–1473

I. INTRODUCTION

Aging effects have been extensively documented for various parameters of the acoustic reflex. Although aging effects on the acoustic-reflex threshold (ART) for the tonal activator appear to be absent (Osterhammel and Osterhammel, 1979; Silman, 1979a; Gelfand and Piper, 1981; Thompson *et al.*, 1980; Silverman *et al.*, 1983; Silman *et al.*, 1987), several studies have shown significantly increased ARTs for the broad-band noise (BBN) activator in older than younger adults (Handler and Margolis, 1977; Silman, 1979a; Gelfand and Piper, 1981; Silverman *et al.*, 1983). Jakimetz *et al.* (1989) investigated the effect of signal bandwidth and spectral density on the ART in the elderly. Their findings revealed that as spectral density increased from one to five components, the mean ART decrease was 5.6 in the elderly subjects versus 8.1 in the younger subjects. Also, the plateau in ART decrease with increases in spectral density occurred earlier in older adults (five components) than in younger adults (seven components). These findings suggest that the auditory system of the elderly has a reduced ability to summate the sound energy from the components of a multitone complex signal, which are presented simultaneously within a particular bandwidth. This result was interpreted by Jakimetz *et al.* as a weakening of the frequency selectivity of the cochlea in the elderly following the reduction in the number of OHCs due to aging.

The growth in acoustic-reflex magnitude with increases in intensity of the tonal or BBN activator is decreased for older adults as compared with younger adults, and the acoustic-reflex growth function saturates in the older, but not the younger, adults (Thompson *et al.*, 1980; Silman and Gelfand, 1981a). Acoustic-reflex latencies for tonal and BBN activators are prolonged in older adults as compared with younger adults (Bosatra *et al.*, 1984).

Numerous studies have been done on temporal integration (threshold-duration function) for the ART (Woodford *et al.*, 1975; Barry and Resnick, 1976; Morgan *et al.*, 1977; Stelmachowitz and Seewald, 1977; Cacace *et al.*, 1991;

[a]Author to whom correspondence should be addressed. Electronic mail: memmer@brooklyn.cuny.edu
[b]Electronic mail: ssilman@brooklyn.cuny.edu
[c]Electronic mail: profaud@nyc.rr.com
[d]Electronic mail: harrylevitt@earthlink.net

TABLE I. Means, standard deviations (SDs), and ranges for the pure-tone air-conduction thresholds (dB HL).

	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
Younger adults						
Right ear						
Mean	3.0	2.0	2.8	1.3	1.0	1.0
SD	3.4	2.51	3.4	2.8	2.1	2.1
Range	0–10	0–5	0–10	0–10	0–5	0–5
Left ear						
Mean	2.8	2.5	3.3	1.0	1.8	2.5
SD	3.4	3.4	2.5	2.6	2.9	2.6
Range	0–10	0–10	0–5	0–10	0–10	0–5
Older adults						
Right ear						
Mean	10.3	10.0	10.0	9.3	11.0	15.5
SD	5.0	4.9	6.3	6.3	6.0	5.8
Range	5–20	5–20	5–20	5–20	0–20	0–20
Left ear						
Mean	9.6	9.3	7.5	8.8	11.8	16.3
SD	4.7	6.1	5.5	6.5	5.7	5.4
Range	0–20	0–20	0–20	0–20	0–20	0–20

Zwislocki, 2003). These studies yielded variable results in all likelihood because of small samples and/or lack of control for age. Thus far, however, aging effects on temporal integration of the ART have not been investigated. Such an investigation would further enhance our understanding and knowledge of the physiology of the acoustic reflex. Therefore, we studied the effect of age on temporal integration for the contralateral ART that is elicited with a BBN activating signal.

II. EXPERIMENTAL METHODS

A. Subjects

The subjects consisted of two groups of adults with normal-hearing sensitivity: one group of 20 young adults (ten males and ten females, ages 18–29 years, with a mean age of 24 years) and one group of 20 older adults (ten males and ten females, ages 59–75 years, with a mean age of 67.5 years).

All subjects in both groups met the following criteria for inclusion: (a) air-conduction thresholds less than or equal to 20 dB HTL at the octave frequencies from 250 to 8000 Hz (ANSI, 1996); (b) bone-conduction thresholds (ANSI, 1996) from 250 to 4000 Hz within 5 dB of the air-conduction thresholds; (c) tympanometric peak pressure within ±50 daPa; (d) peak-compensated static-acoustic admittance not less than 0.35 mmho (Jerger, 1970; Silman and Silverman, 1991); (e) contralateral acoustic reflexes present for the 500-, 1000-, and 2000-Hz tonal activators within the 90th percentiles (Silman and Gelfand, 1981b); (f) negative otoscopic findings; and (g) negative otologic and neurologic histories.

Table I shows the means, ranges, and standard deviations for the pure-tone air-conduction thresholds for the younger and older adult subjects.

B. Procedure

The order of ear tested was counterbalanced across subjects in each group. The activating stimulus durations (12, 25, 50, 100, 200, 300, 500, 1000 ms) were randomized. The interstimulus interval (ISI) was 6 s as this interval is optimal for the avoidance of overestimation of aging effects on the acoustic reflex (Jerger and Oliver, 1987). The contralateral mode of activating signal presentation was employed whereby the activating signal was presented through an insert phone in the test ear and the acoustic-admittance change was monitored in the contralateral ear with a digital strip-chart recorder. The activating stimulus was presented below the expected ART and the intensity of the activating stimulus was increased in 1-dB increments until a reliable ART was observed. The ART was determined by visual inspection of the strip-chart recorder. An acoustic-reflex response was considered to be present if its pattern was distinguishable from the background noise. The ART was defined as the lowest level at which a repeatable response was present on all three ascending trials (Silman *et al.*, 1978).

C. Instrumentation and calibration

The GS-16 Audiometer was calibrated according to ANSI (1996). An acoustic-immittance meter (GS-1723) with 220-Hz probe tone was calibrated according to ANSI (1987). The Hewlett Packard Spectrum Analyzer (Model 3582A) was used to measure the spectrum of the BBN stimulus.

Figure 1 shows a block diagram of the instrumentation and setup. The activating stimulus consisted of a BBN signal that was generated by a white-noise generator (Model S 81-02). The signal was routed first into a programmable attenuator (S 85-08) and up-down counter (S 41-28), for control of intensity. The output of the up-down counter was directed to the electronic switch (S 84-04) that controlled the rise and fall of the stimulus (3 ms rise/fall). The signal from the electronic switch was directed into an S 82-24 amplifier, the

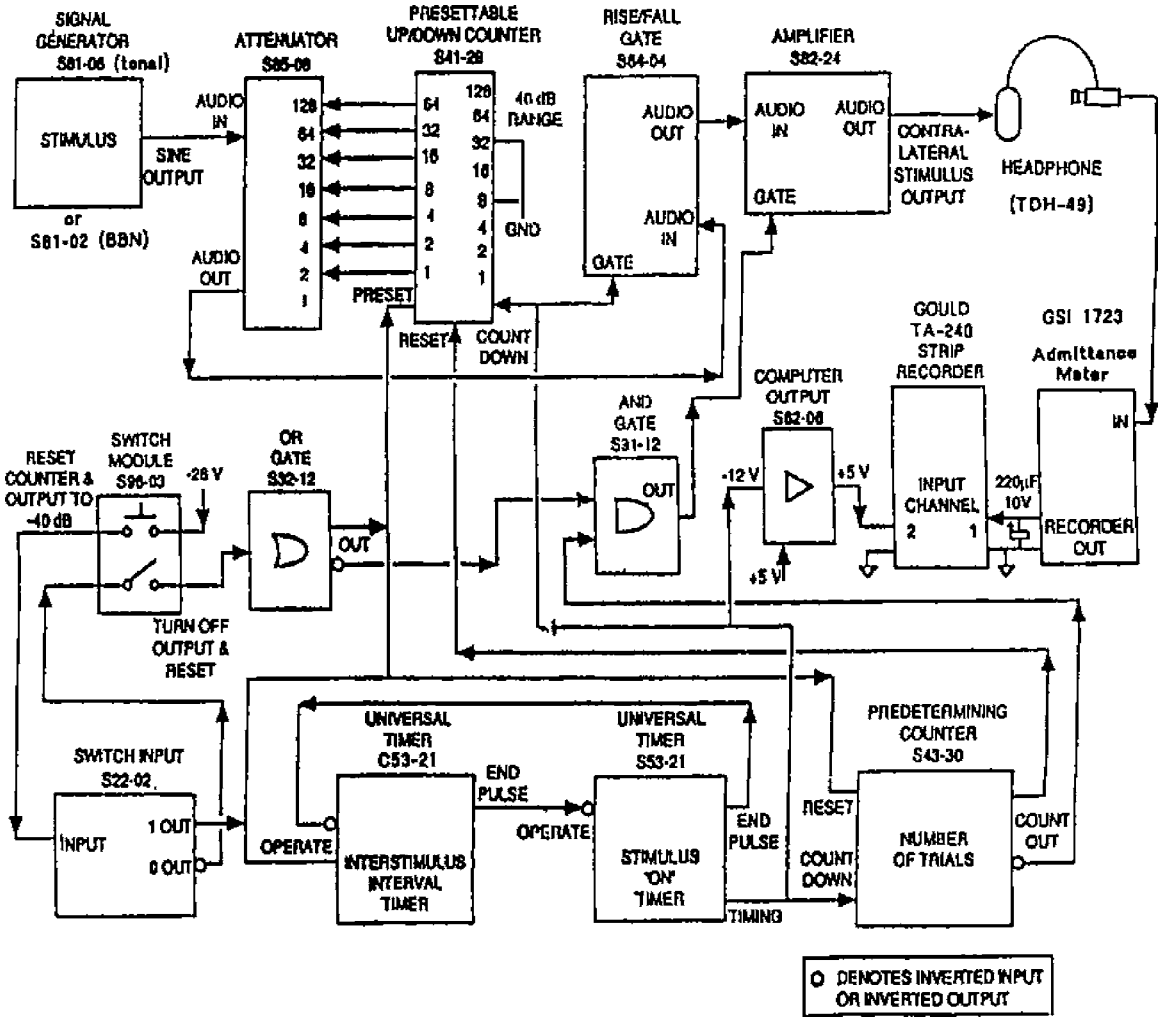


FIG. 1. Block diagram of instrumentation setup.

electrical output of which was routed to a TDH-49 transducer (encased in a Telephonics Corp. P/N 510C017-1 cushion) that delivered the activating signal. The activating stimulus duration and ISI were controlled by timers (Universal C 53-21 and S 53-21) that controlled the on-off cycle of the electronic switch.

The acoustic-admittance change for the 220-Hz probe tone was monitored with an electroacoustic immittance device (GSI 1723 Middle Ear Analyzer) connected to one channel of a two-channel digital strip-chart recorder (Gould-Easy Graf TA 240 Strip Chart Recorder with a resolution of 5000 Hz). The second channel of the recorder was used as an event marker and was controlled by the Coulbourn computer output (Model S 62-06).

The activating signal was calibrated with a digital sound-level meter (Quest Model OB-300), coupler (Model EC-9A), and frequency counter (Quest AA-175). All sound-pressure level (SPL) values were in dB *re* 20 μPa.

In order to account for the time constant of the sound-level meter for short-duration signals, the following procedure was performed: First, the intensity of a BBN activating signal was measured, using the fast mode, for a duration greater than 1 s; its value in SPL was recorded and its voltage value was noted on the oscilloscope (Model 510 3N Tektronix). Next, the intensity in dB SPL was measured for the signal at each of the eight activating durations. The difference between the recorded SPL at the specified activator duration and the SPL at greater than 1 s was calculated for each of the eight activator durations. These differences were assumed to reflect the time constant of the sound-level meter and were compared with the differences reported by the manufacturer. When deviations above or below those reported by the manufacturer were noted, corrections were applied to the presentation levels of the activator stimuli. That is, if measured SPL at a particular duration was above the manufacturer's value, we subtracted the difference between the manufacturer's value and measured SPL. The opposite was done for measured values below manufacturer's reported values.

The acoustic-admittance device was calibrated to a standard volume of 2 ml for static-acoustic admittance. The sensitivity of meter deflection during acoustic-reflex testing at the full scale of 0.5 ml (which recorded a full deflection) was equal to 0.5 ml, as specified by the manufacturer. This calibration was confirmed by coupling the probe assembly to a microsyringe (Gilmont Model S1200) set to a full-scale deflection of 0.5 ml. Such calibration insures that any meter

TABLE II. Means, standard deviations (SDs), and ranges for younger and older adults subjects for the BBN activator.

Duration (ms)	12	25	50	100	200	300	500	1000
Younger adults								
N	20	20	20	20	20	20	20	20
Mean	106.7000	101.3500	95.3000	89.4500	82.3000	76.3500	73.9000	73.9250
SD	9.3265	9.4369	8.6167	9.8767	8.3152	9.2766	6.8798	7.2770
Older adults								
N	20	20	20	20	20	20	20	20
Mean	105.4250	100.5500	96.750	90.5250	87.2250	83.8750	83.0250	82.4750
SD	7.3615	7.8050	8.7604	7.2828	5.9393	6.4438	5.6626	6.1761

deflection will correspond with the values (in μl) indicated in the manufacturer's instrument manuals and in our own calibration.

Duration, ISI, and rise and fall times of the experimental stimulus, controlled by the electronic switches (S 84-04) and timers (C 53-21 and S 53-21), were checked with a B&K Precision Dynascan 20-MHz digital storage oscilloscope (Model 2520). Spectral analysis of the BBN activating signal at the 12- and 1000-ms activating signal durations indicated the following: the spectrum was uniform up to approximately 6280 Hz (the 3-dB down point); beyond this frequency through 10 000 Hz, the filter roll-off was approximately 20 dB/oct.

The air-conduction and bone-conduction thresholds were obtained in an Industrial Acoustics Corporation double-walled sound-treated booth (Acoustics Systems, Model 18997A). The pure-tone air-conduction thresholds were obtained using a GS-16 audiometer with TDH-49 air-conduction transducers mounted in (MX-41AR) cushions. The bone-conduction thresholds were obtained with a Radioear (Model B-71) transducer.

Calibrations of the air- and bone-conduction stimuli were done periodically according to ANSI standards (ANSI, 1996) with biologic checks performed prior to each test session. The ambient noise inside the audiometric sound-treated suite did not exceed specified levels (ANSI, 1991).

Measurement of the latency of immittance change as a result of the activating signal was done using the procedure described by Silman and Gelfand (1982) involving measurement of the first detectable acoustic-immittance change in a cavity (micro-liter syringe Gilmont Model S 1200). Based upon the definition of latency of the electroacoustic immittance device as "the time from stimulus onset to the first detectable immittance change" (p. 126), the latency was established to be 12 ms, consistent with other commercially available devices.

The acoustic-immittance measurements and evaluation of the temporal integration for the ART were measured in a laboratory. This laboratory was found to be adequately quiet for those procedures as indicated by the octave-band measurements of ambient noise made with the digital sound-level meter (linear weighting). The ambient noise level at 1000 Hz was 40.4 dB SPL, which is well below the level that could elicit an acoustic reflex. The overall intensity of the ambient noise in the laboratory, using the C-weighting scale, was 64.7 dB SPL. An additional attenuation of approximately 20 dB was provided by the insert phones.

D. Statistical analysis

Prior to the investigation, we determined the proper number of subjects to be included in both groups, experimental and control, by using statistical power. The participant effect size and power were provided by formulas given in Cohen (1988). The effective sample size is given by Cohen's formula (8.3.4)

$$n' = \frac{\text{denominator } df}{u+1} + 1$$

and is equal to 20 subjects in each group. With a power of 0.80, the moderate effect size is equal to approximately 0.65.

The data were analyzed using a repeated measures analysis of covariance. The between-subjects factor was age and the within-subjects factor was duration of the BBN activator. The fixed or "constant" covariate was hearing threshold level at 250, 500, 1000, 2000, 4000, and 8000 Hz. It was preferable to use the multivariate design rather than the conservative corrected univariate Greenhouse-Geisser procedure. Because the within-subjects factor has more than two levels (there are eight durations), certain symmetry conditions, or assumptions, are made. If these assumptions are violated, then the statistical results for the univariate test (the Greenhouse-Geisser) are too likely to yield significant results.

In order to test for symmetry conditions, we used the following two tests: Box's test of equality of covariance matrices and Mauchly's test of sphericity. Box's test of equality determines whether the correlation between the two age groups, i.e., between the younger and older subjects, is equal. Mauchly's test of sphericity investigates the existence of a significant correlation within the pooled group of subjects.

An alpha level of 0.05 was used for all statistical tests.

III. RESULTS

Table II shows the means and standard deviations for younger and older adult subjects for the BBN activator.

The results of Box's test of equality of covariance matrices were significant ($p=0.043$), suggesting a lack of correlation between the groups. The results of Mauchly's test of sphericity were significant (Mauchley's $W=0.008$, χ^2

TABLE III. Results of multivariate tests of temporal integration for the ART for the BBN activator.

Effect	Hypothesis				
	Value	F	df	Error df	Significance (p)
Duration					
Pillai's trace	0.716	11.162	7.000	31.000	0.000
Duration * hearing level					
Pillai's trace	0.115	0.576	7.000	31.000	0.770
Duration * age group					
Pillai's trace	0.233	1.344	7.000	31.000	0.264

=165.038, df=27, p=0.000), suggesting the absence of correlation within the pooled group. The latter result violated the condition of symmetry; therefore, the analysis depended largely on the multivariate design, which does not rely on the symmetry condition.

Inspection of the multivariate results (see Table III) reveals a significant main effect for duration (p=0.000). That is, as the activator duration increases, the ART for the BBN activator decreases in level. This finding can be observed for both groups in Fig. 2.

Table IV (the between-subjects effect for temporal integration of the BBN ART) reveals a significant main effect for age (p=0.041). That is, the ART for the BBN activator was significantly higher for the older than younger adults. These differences between groups are illustrated in Fig. 2 as well, in which we see the ART for the BBN activator as a function of activator duration (ms) for both groups. Table IV also shows that the main effect of hearing level failed to achieve significance (p=0.154).

Although Table III revealed the lack of a significant interaction effect between activator duration and age (p =0.264), Fig. 2 appears to suggest such an interaction. The absence of a significant interaction between activator duration and age may be related to the fact that the functions for the two groups overlap from 12 to 200 ms. Perhaps this overlap impacted the statistical results.

In order to test this assumption, we further analyzed the difference in temporal integration of the ART for the BBN activator between the younger and older adults using pairwise analysis (see Table V). Pair-wise analysis indicates statistically significant differences between the two age groups at durations of 200 ms and longer. The differences in the mean ART for the BBN activator between the younger versus older subjects essentially become wider, and the significance levels overall become stronger, as activator duration increases beyond 200 ms. At 1000 ms, the difference widens to approximately 10 dB between the two groups (see Fig. 3). This 10-dB difference between the threshold for the BBN activator between the older and younger subjects is consistent with previous research (Djupesland and Zwislocki, 1971; Djupesland *et al.*, 1973; Jerger *et al.*, 1977; Bazarov and Moroz, 1975; Richards, 1975; Woodford *et al.*, 1975; Barry and Resnick, 1976).

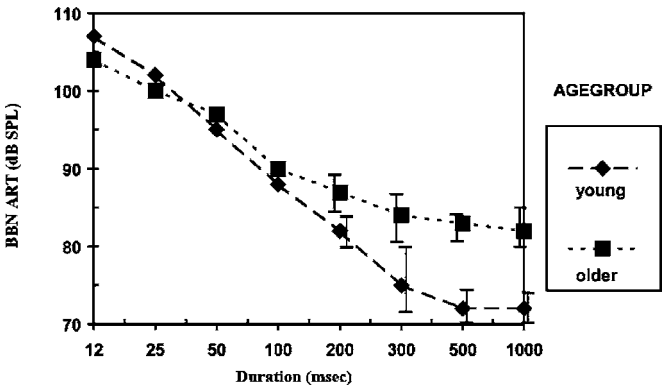


FIG. 2. ART for the BBN activator as a function of duration for younger versus older adults. Error bars begin at 200 ms, the duration at which significance begins. Error bars represent the standard error of the mean.

IV. DISCUSSION

Recall that the difference in temporal integration for threshold for older and younger subjects for the BBN activator in this study widened as a function of activator duration, becoming as wide as 10 dB at 1000 ms. This divergence is consistent with previous reports that showed differences ranging from 8 to 11 dB for temporal integration at 1000 ms (Handler and Margolis, 1977; Silman, 1979a, b; Gelfand and Piper, 1981; Wilson, 1981). It has been suggested that the difference in the ART for the BBN activator at 1000 ms is related to the slight differences in hearing sensitivity between younger adults with normal-hearing sensitivity and older adults with normal-hearing sensitivity (Jerger *et al.*, 1978a, Jerger, 1979). It was therefore assumed that temporal integration for older and younger adults is similar, and any observed differences are related to slight differences in hearing sensitivity rather than age. Several observations do not support this assumption. Firstly, results of ANCOVA counter the above argument that differences in temporal integration between the younger and older subjects are likely due to slight differences in hearing sensitivity. Even if hearing sensitivity

TABLE IV. Between-subjects effect for temporal integration of the BBN ART.

Source	Type III sum of squares	df	Mean square	F	Significance
Intercept	385 334.504	1	385 334.504	1044.246	0.000
Hearing Level	779.709	1	779.709	2.113	0.154
Age Group	1663.145	1	1663.145	4.507	0.041
Error	13653.275	37	369.007		

TABLE V. Pair-wise comparisons based on estimated marginal means.

Duration (ms)	(I) Age group	(J) Age group	Mean difference (I-J)	Standard error	Significance (p)	95% confidence interval for Difference[a]	
						Lower bound	Upper bound
12	Younger	Older	−2.812	5.402	0.606	−13.758	8.134
	Older	Younger	2.812	5.402	0.606	−8.134	13.758
25	Younger	Older	−2.438	5.591	0.665	−13.767	8.891
	Older	Younger	2.438	5.591	0.665	−8.891	13.767
50	Younger	Older	−7.209	5.511	0.199	−18.376	3.957
	Older	Younger	7.209	5.511	0.199	−3.957	18.376
100	Younger	Older	−8.369	5.465	0.134	−19.442	2.704
	Older	Younger	8.369	5.465	0.134	−2.704	19.442
200	Younger	Older	−10.193	4.586	0.032	−19.485	−0.900
	Older	Younger	10.193[b]	4.586	0.032	0.900	19.485
300	Younger	Older	−14.642[b]	5.010	0.006	−24.793	−4.490
	Older	Younger	14.642[b]	5.010	0.006	4.490	24.793
500	Younger	Older	−14.566[b]	3.961	0.001	−22.592	−6.540
	Older	Younger	14.566[b]	3.961	0.001	6.540	22.592
1000	Younger	Older	−13.700[b]	4.274	0.003	−22.361	−5.040
	Older	Younger	13.700[b]	4.274	0.003	5.040	22.361

[a]Adjustment for multiple comparisons: least significant difference (equivalent to no adjustment).
[b]The mean difference is significant at the 0.05 level.

was significantly different between the two groups, such a difference would have no effect on the findings. This point will be further elaborated upon below.

Secondly, pair-wise analysis revealed no significant difference between the two groups for durations below 200 ms. Significance becomes apparent from 200 ms and above. Therefore, if the differences between ART for BBN for younger and older subjects is related to hearing sensitivity, then the difference in the thresholds for the two groups would have been observed at shorter durations. These results support those of Silverman *et al.* (1983), who established that methodological differences between Jerger *et al.* (1978b) and others (Silman, 1979a, b; Gelfand and Piper, 1981; Handler and Margolis, 1977) accounted for the discrepant findings regarding the effect of age on the ART for the BBN activator. For example, the former used 5-dB intensity increments and visual monitoring of needle deflection to establish ART for the BBN activator whereas the latter used 1-dB intensity increments and a strip-chart recorder.

The truncated BBN temporal integration function for older adults with normal-hearing sensitivity as compared with that for younger adults with normal-hearing sensitivity reflects of saturation of temporal integration at activator durations greater than or equal to 200 ms in the older adult group. This finding is similar to the aging effect on the BBN acoustic-reflex growth function in that the growth function does not saturate in younger adults with normal-hearing sensitivity (Silman and Gelfand, 1981a; Thompson *et al.*, 1980; Wilson and McBride, 1978) but does so in older, normal-hearing adults (Silman *et al.*, 1978; Wilson and McBride, 1978; Thompson *et al.*, 1980). To explain saturation in the BBN acoustic-reflex growth function at high levels in the older adults, and the absence of such saturation in the younger adults, Silman *et al.* (1978) hypothesized that aging results in a reduced capacity of energy that can be integrated. This hypothesis may also explain saturation in the temporal integration function for the BBN ART in older adults. Perhaps aging causes alteration of the time frame during which there is an intensity/duration tradeoff. Beyond the time frame, an increase in stimulus duration will not improve threshold.

Jerger and Oliver (1987) investigated the interaction of age, ISI, and offset latency on reflex magnitude. They reported the interaction of age and ISI on the amplitude of the AR in the ipsilateral mode but not in the contralateral mode. In contrast, they reported an interaction between offset latency and age in the contralateral mode but not in the ipsilateral mode.

These authors also noted that some of the subjects in their older group had a mild to moderate sensorineural hearing loss while the younger subjects had normal hearing. Given the differential effect of the interaction between ISI and age on the ipsilateral and contralateral ARTs, these investigators suggested that if hearing loss rather than age had impacted their findings, both the contralateral and ipsilateral acoustic reflexes would have been similarly affected. Since the effects were different for different modalities, they concluded that changes in the acoustic-reflex arc in older adults reflect the effects of aging rather than the effects of hearing loss. Perhaps the reflex interneurons that control integration over time, and are located in the brainstem, have altered temporal integration for the ART in the older adult. In the current study the time frame for temporal integration for the AR is between 12.5 and 500 ms for the younger adults and between 12.5 and 250 ms in the older adults. The saturation of the BBN ART temporal integration function and BBN acoustic-reflex growth function in older adults suggests a

reduced energy capacity at threshold, as well as suprathreshold activator levels.

V. CONCLUSION

The results of this study indicate that temporal integration for BBN ART for older individuals is different from temporal integration for the young. This finding explains the differences in threshold between young and older adults at durations of 1000 ms and above reported in the literature. Future research investigating temporal integration for BBN ART should control for age. In addition, research is also suggested for temporal integration for a tonal activator.

ACKNOWLEDGMENTS

This study is based upon the work of the senior author in partial fulfillment of the thesis requirements for the Ph.D. in Speech and Hearing Sciences at the Graduate Center of the City University of New York.

ANSI (1987). ANSI S3.39-1987. "American national standard specifications for instruments to measure aural acoustic impedance and admittance (aural acoustic immittance)" (American National Standards Institute, New York).

ANSI (1991). ANSI S3.1-1991, "Maximum permissible ambient noise for audiometric test rooms" (American National Standards Institute, New York).

ANSI (1996). ANSI S3.6-1996, "Specification for audiometers" (American National Standards Institute, New York).

Barry, S. J., and Resnick, S. B. (**1976**). "Comparison of acoustic reflex and behavioral thresholds as a function of stimulus frequency and duration," J. Am. Aud Soc. **2**, 35–37.

Bazorov, V. G., and Moroz, B. (**1975**). "Dependence of the acoustic reflex threshold on the duration of the noise stimulus," Vestn. Otorinolaringol. **3**, 13–18.

Bosatra, A., Russolo, M., and Silverman, C. A. (**1984**). "Acoustic-reflex latency: State of the art," in *The Acoustic Reflex: Basic Principles and Clinical Applications*, edited by S. Silman (Academic, New York), pp. 301–328.

Cacace, A. T., Margolis, R. H., and Relkin, E. M. (**1991**). "Threshold and suprathreshold temporal integration effects in the crossed and uncrossed human acoustic stapedius reflex," J. Acoust. Soc. Am. **89**, 1255–1261.

Cohen, J. (**1988**). *Statistical Power Analysis for the Behavioral Science*, 2nd ed. (Erlbaum, Hillsdale, NJ).

Djupesland, G., and Zwislocki, J. J. (**1971**). "Effect of temporal integration on the human stapedius reflex," Acta Oto-Laryngol. **71**, 262–265.

Djupesland, G., Sundby, A., and Flottorp, G. (**1973**). "Temporal summation in the acoustic stapedius reflex mechanism," Acta Oto-Laryngol. **76**, 305–312.

Gelfand, S. A., and Piper, N. (**1981**). "Acoustic-reflex thresholds in young and elderly subjects with normal hearing," J. Acoust. Soc. Am. **69**, 295–297.

Handler, S. D., and Margolis, R. H. (**1977**). "Predicting hearing loss from stapedial reflex thresholds in patients with sensorineural impairment," Trans. Am. Acad. Ophthalmol. Otolaryngol. **84**, 425–431.

Jakimetz, J. J., Silman, S., Miller, M. H., and Silverman, C. A. (**1989**). "Some effects of signal bandwidth and spectral density on the acoustic reflex threshold in the elderly," J. Acoust. Soc. Am. **86**, 1783–1789.

Jerger, J. (**1970**). "Clinical experience with impedance audiometry," Arch. Otolaryngol. **92**, 311–324.

Jerger, J. (**1979**). "Comments on 'The effect of aging on the stapedius reflex thresholds' [J. Acoust. Soc. Am. **66**, 735–738 (1979)]," J. Acoust. Soc. Am. **66**, 908.

Jerger, J., and Oliver, T. (**1987**). "Interaction of age and intersignal interval on acoustic reflex amplitude," Ear Hear. **8**, 322–325.

Jerger, J., Hayes, D., and Anthony, L. (**1978a**). "Effect of age on prediction of sensorineural hearing level from the acoustic reflex," Arch. Otolaryngol. **104**, 393–394.

Jerger, J., Mauldin, L., and Lewis, N. (**1977**). "Temporal summation of the acoustic reflex," Audiology **16**, 177–200.

Jerger, J., Hayes, D., Anthony, L., and Mauldin, L. (**1978b**). "Factors influencing prediction of hearing level from the acoustic reflex," Maico Monogr. Contemp. Audiol. **1**, 1–20.

Morgan, D. E., Gilman, S., and Dirks, D. D. (**1977**). "Temporal integration at the 'threshold' of the acoustic reflex," J. Acoust. Soc. Am. **62**, 168–176.

Osterhammel, D., and Osterhammel, P. (**1979**). "Age and sex variations for the normal stapedial Reflex threshold and tympanometric compliance values," Scand. Audiol. **8**, 153–158.

Richards, A. M. (**1975**). "Threshold of the acoustic stapedius reflex for short-duration tone bursts," J. Aud Res. **15**, 87–94.

Silman, S. (**1979a**). "The effects of aging on the stapedius reflex thresholds," J. Acoust. Soc. Am. **66**, 735–738.

Silman, S. (**1979b**). "The acoustic reflex, aging, and the distortion product," J. Acoust. Soc. Am. **66**, 909–910.

Silman, S., and Gelfand, S. A. (**1981a**). "Effect of sensorineural hearing loss on the stapedius reflex growth function in the elderly," J. Acoust. Soc. Am. **69**, 1099–1106.

Silman, S., and Gelfand, S. A. (**1981b**). "The relationship between magnitude of hearing loss and acoustic reflex threshold levels," J. Speech Hear Disord. **46**, 312–316.

Silman, S., and Gelfand, S. A. (**1982**). "The acoustic reflex in diagnostic audiology-Part 2," Audiology **7**, 125–138.

Silman, S., and Silverman, C. A. (**1991**). *Auditory Diagnosis: Principles and Applications* (Academic, San Diego).

Silman, S., Gelfand, S. A., and Emmer, M. B. (**1987**). "Acoustic reflex in hearing loss identification and prediction," Semin. Hear. **3**, 379–390.

Silman, S., Popelka, G. R., and Gelfand, S. A. (**1978**). "Effect of sensorineural hearing loss on acoustic stapedius reflex growth functions," J. Acoust. Soc. Am. **64**, 1406–1411.

Silverman, C. A., Silman, S., and Miller, M. H. (**1983**). "The acoustic reflex threshold in aging ears," J. Acoust. Soc. Am. **73**, 248–255.

Stelmachowicz, P. G., and Seewald, R. C. (**1977**). "Threshold and suprathreshold temporal integration function in normal and cochlear-impaired subjects," Audiology **16**, 94–101.

Thompson, D. J., Sills, J. A., Recke, K. S., and Bui, D. M. (**1980**). "Acoustic reflex growth in the aging adult," J. Speech Hear. Res. **23**, 405–418.

Wilson, R. H. (**1981**). "The effects of aging on the magnitude of the acoustic reflex," J. Acoust. Soc. Am. **24**, 406–414.

Wilson, R. H., and McBride, L. M. (**1978**). "Threshold and growth of the acoustic reflex," J. Acoust. Soc. Am. **63**, 147–154.

Woodford, C., Henderson, R., Hamernik, R., and Feldman, A. (**1975**). "Threshold-duration function of acoustic reflex in man," Audiology **14**, 53–62.

Zwislocki, J. J. (**2003**). "A look at neural integration in the human auditory system through the stapedius muscle reflex," Proc. Natl. Acad. Sci. U.S.A. **100**, 9073–9078.

Comparison of the roex and gammachirp filters as representations of the auditory filter

Masashi Unoki[a)]
School of Information Science, Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan

Toshio Irino
Faculty of Systems Engineering, Wakayama University/930 Sakaedani, Wakayama, 640-8510, Japan

Brian Glasberg and Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge, CB2 3EB, United Kingdom

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

(Received 19 October 2005; revised 20 June 2006; accepted 23 June 2006)

Although the rounded-exponential (roex) filter has been successfully used to represent the magnitude response of the auditory filter, recent studies with the roex(p,w,t) filter reveal two serious problems: the fits to notched-noise masking data are somewhat unstable unless the filter is reduced to a physically unrealizable form, and there is no time-domain version of the roex(p,w,t) filter to support modeling of the perception of complex sounds. This paper describes a compressive gammachirp (cGC) filter with the same architecture as the roex(p,w,t) which can be implemented in the time domain. The gain and asymmetry of this parallel cGC filter are shown to be comparable to those of the roex(p,w,t) filter, but the fits to masking data are still somewhat unstable. The roex(p,w,t) and parallel cGC filters were also compared with the *cascade* cGC filter [Patterson *et al.*, J. Acoust. Soc. Am. **114**, 1529–1542 (2003)], which was found to provide an equivalent fit *with 25% fewer coefficients*. Moreover, the fits were stable. The advantage of the cascade cGC filter appears to derive from its parsimonious representation of the high-frequency side of the filter. It is concluded that cGC filters offer better prospects than roex filters for the representation of the auditory filter. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2228539]

PACS number(s): 43.66.Ba, 43.66.Dc, 43.64.Bt [AJO] Pages: 1474–1492

I. INTRODUCTION

The frequency selectivity of the auditory system is often conceptualized as a bank of bandpass auditory filters (Fletcher, 1940; see Patterson and Moore, 1986, for a review). In the case of the human auditory filter, the magnitude response, or shape, of the filter has frequently been derived from simultaneous masking experiments where a probe tone is masked by a notched noise that is either symmetrically or asymmetrically positioned with respect to the probe frequency (e.g., Patterson, 1976; Patterson *et al.*, 1982; Lutfi and Patterson, 1984; Glasberg *et al.*, 1984b; Moore *et al.*, 1990; Rosen *et al.*, 1998; Baker *et al.*, 1998). When the notch is centered on the frequency of the probe tone, the function that describes how probe threshold (in dB) decreases as the edges of the notch move away from the signal frequency is close to linear over a wide range of notch widths (Patterson, 1976), indicating that the threshold function is approximately exponential. At the narrowest notches, the threshold function bends down below the upward projection of the exponential approximation, and when the notch becomes relatively wide, the threshold function bends up from the downward projection of the exponential approximation. This led Patterson and Nimmo-Smith (1980) to suggest that the magnitude response of the auditory filter might well be represented by a pair of back-to-back exponential functions that were rounded in some way at the top and bottom, in accordance with the data. They described a series of rounded-exponential, or "roex," functions, to represent the magnitude characteristic of the auditory filter with a small number of filter parameters (Patterson and Nimmo-Smith, 1980; Patterson *et al.*, 1982; Lutfi and Patterson, 1984; Glasberg *et al.*, 1984a). The development of the roex filter family is described in Sec. I B–I D as it pertains to the current study.

The most successful version of the roex filter in recent times is the roex(p,w,t), originally described in Patterson *et al.* (1982); it is a combination of a narrow roex(p) filter and a broader roex(t) filter that operate in parallel to simulate the *p*assband and *t*ails, respectively, of the magnitude character-

[a)]Electronic mail: unoki@jaist.ac.jp. With regard to contributions: MU developed the parallel cGC filter used t separate the effects of filter shape and architecture, and performed all of the filter fitting on which the paper is based. TI developed the cascade cGC filter and the procedure for fitting the combined data sets. BG assisted in the analysis of the threshold data. BM assisted in the writing and revision of the paper. RP supervised the project and was largely responsible for the writing and revision of the paper.

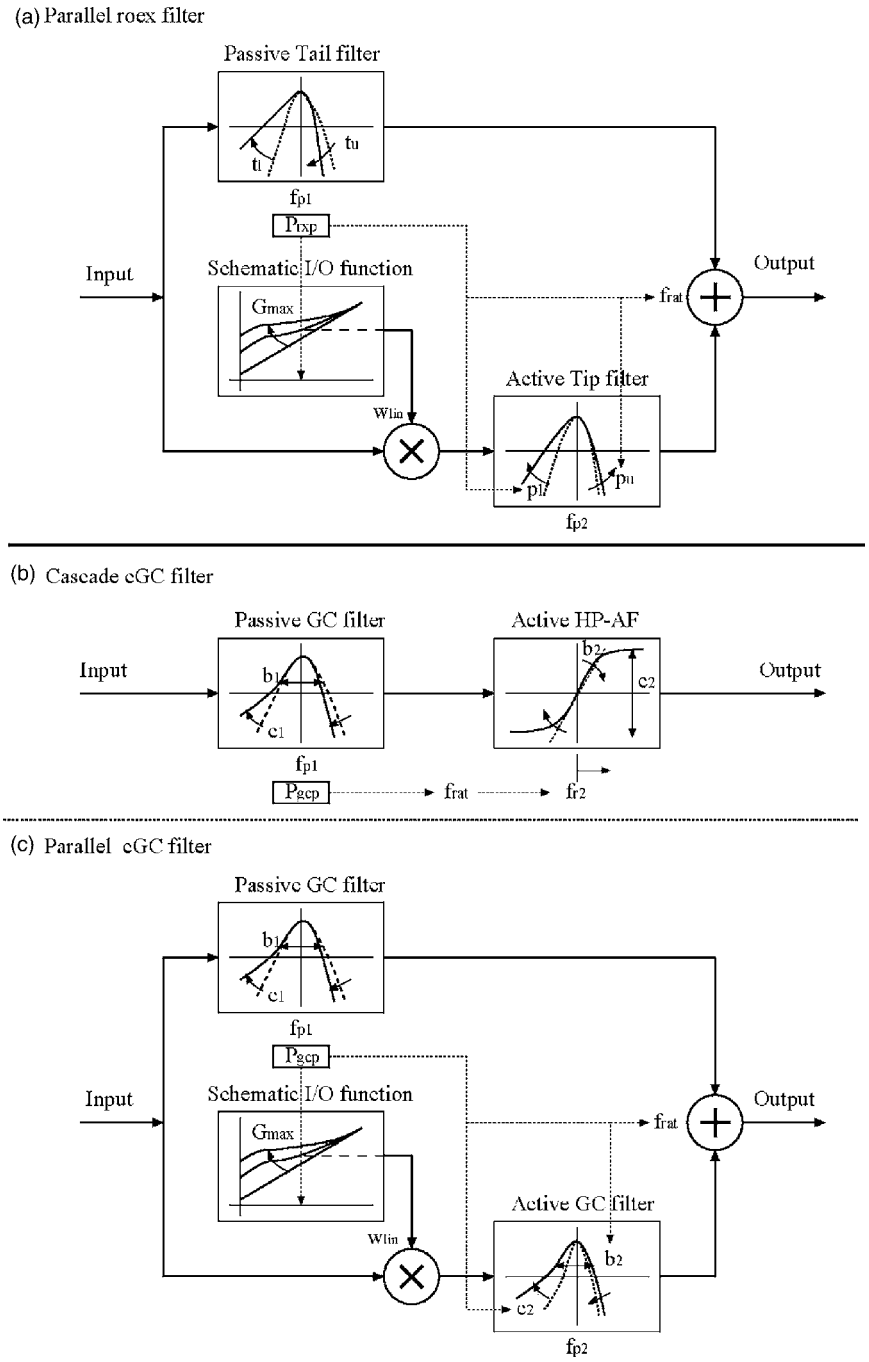


FIG. 1. The architecture of three auditory filters: (a) a roex filter with parallel tip and tail filters (parameters: t_l, t_u, p_l, p_u, $G_{\max}$, and f_{rat}); (b) a compressive gammachirp filter with cascaded tail and tip filters (parameters: b_1, c_1, b_2, c_2, and f_{rat}); and (c) a compressive gammachirp filter with parallel tip and tail filters (parameters: b_1, c_1, b_2, c_2, $G_{\max}$, and f_{rat}). In each case, the level at the output of the tail filter determines the characteristics of the level-dependent tip filter.

istic of the human auditory filter. The interaction of the roex(p) and roex(t) filters as a function of stimulus level has been likened to the interaction of the "tip" and "tail" components of the cochlear filter in response to changes in stimulus intensity (Glasberg and Moore, 2000), and this suggested that the roex filter could be used to derive properties of human cochlear filtering from psychophysical masking experiments (e.g., Glasberg and Moore, 2000; Oxenham and Shera, 2003; Baker and Rosen, 2006). The architecture of this parallel roex filter is illustrated in Fig. 1(a); the filter is described in Secs. I D and II B. The roex(p,w,t) filter does, however, have two serious problems that would appear to limit its usefulness: First, there are no time-domain versions of roex filters, and no roex(p,w,t) filterbank to support research into the perception of complex waveforms like those of speech and music. Second, the fits provided by the complete version of the roex(p,w,t) filter are commonly unstable; the high-frequency skirts of the tip and tail filters have proven difficult to differentiate. Attempts to solve the problem by reducing the number of parameters have led to a version of the filter that is discontinuous (Baker *et al.*, 1998), which means that it is not physically realizable and not physiologically plausible.

The basic problem with the roex filter was recognized some time ago (Patterson *et al.*, 1987), and an alternative family of gammatone (GT) and gammachirp (GC) filters was developed to provide time-domain simulations of auditory filtering with magnitude responses similar to those derived with the roex filter. The compressive gammachirp (cGC) filter (Irino and Patterson, 2001; Patterson *et al.*, 2003) can

accommodate much of the psychophysical data that the roex filter has previously been used to explain, in particular, the level-dependent asymmetry of the auditory filter observed in simultaneous masking (see Patterson *et al.*, 2003). The cGC filter has a cascade architecture as shown in Fig. 1(b); the filter is described in Secs. I E and II C. In this paper, we show how the cGC filter avoids the problems associated with the roex filter family, and we compare the fits to notched-noise data provided by the cGC and roex filters to determine whether there remain aspects of simultaneous masking that are still better explained by, or more compactly summarized by, a roex filter.

A. Time-domain simulation of the auditory filter

The most obvious limitation of the roex filter is the lack of a time-domain version. The phase responses and impulse responses of roex filters are not specified, so they cannot be used to filter a waveform. Roex "filters" are only used to filter sounds in the spectral domain by specifying the relative attenuation imposed on frequency components of stationary sounds. There is a standard filter design technique, based on the Fourier transform (Oppenheim and Schafer, 1975), which can be used to generate a finite-impulse-response (FIR) filter from a magnitude response with a continuous derivative. The filters are symmetric, and strictly speaking, noncausal, but they could be useful in certain circumstances, and Assmann and Summerfield (1990) showed how this technique could be used to produce a time-domain version of the simplest roex filter—the roex(p). However, the phase response of the filter derived in this way does not necessarily correspond to the phase response of the auditory filter. Furthermore, the most widely used version of the roex(p,w,t) filter is not continuous (e.g., Baker *et al.*, 1998; Glasberg and Moore, 2000; Bacon *et al.*, 2002; Oxenham and Shera, 2003; Baker and Rosen, 2006), and so there is no real prospect of creating a time-domain version of it. This suggests that there may be something fundamentally wrong with the parallel roex filter as a representation of the auditory filter—in particular, the way it represents the high-frequency side of the filter, and by analogy, how it represents the action of the passive basilar membrane. This issue is pursued in the latter part of Sec. I D.

The gammatone filter is defined in the time domain; it was originally developed by de Boer (1975) to represent the impulse response of the cat's auditory filter. The shape of the magnitude spectrum is similar to that of the roex at moderate sound levels, and so the gammatone filter was adapted to provide a time-domain simulation of the human auditory filter (Patterson *et al.*, 1987; 1992; Cooke, 1993), and a gammatone filterbank simulation of basilar-membrane motion (Patterson *et al.*, 1995). It is this representation of cochlear filtering, rather than the roex(p) filterbank of Assmann and Summerfield (1990), that is commonly used in auditory models that require a multichannel, time-domain simulation of cochlear filtering; examples include models of auditory perception (e.g., Patterson *et al.*, 1995; Cohen *et al.*, 1995; Hohmann, 2002), models of speech processing and speech recognition (e.g., Kubin and Kleijn, 1999; Cooke *et al.*, 2001; Gunawan and Ambikairajah, 2004), computational models of auditory scene analysis (e.g., Unoki and Akagi, 1999; Roman *et al.*, 2003; Divenyi, 2004; Irino *et al.*, 2006), and models of auditory brain activation (e.g., Patterson *et al.*, 2002; Krumbholz *et al.*, 2003).

The magnitude response of the gamma*tone* auditory filter is essentially symmetric, and so its use is limited to moderate stimulus levels where the auditory filter is not markedly asymmetric. To explain the level-dependent asymmetry observed in the auditory filter, Irino and Patterson (1997) used operator techniques to produce a more general version of the gammatone filter with variable asymmetry; it is referred to as the gamma*chirp* auditory filter because the instantaneous frequency of the impulse response glides up to its asymptotic frequency over the first few cycles. Gammachirp filters are, by their nature, physically realizable and causal, and the corresponding gammachirp filterbanks effectively perform a wavelet transform, which Reimann (2006) argues is the correct mathematical representation of cochlear filtering.[1] The compressive gammachirp (cGC) filter (Irino and Patterson, 2001) can explain much of the simultaneous masking data that the roex filter has previously been used to explain (see Patterson *et al.*, 2003). The cGC filter also has a chirp that does not vary with level, consistent with the impulse response data obtained physiologically from small mammals (Carney *et al.*, 1999; de Boer and Nuttall, 2000).[2]

It also appears that it may be possible to extend the gammachirp filter in a straightforward manner to explain two-tone suppression and forward masking (Irino and Patterson, 2005, 2006). The roex filter has been used to derive estimates of auditory filter width from forward masking data (Moore and Glasberg, 1981; Oxenham and Shera, 2003; Unoki and Tan, 2005), but without a time-domain representation, the roex filter is unlikely to provide a full account of the processing involved in forward masking and two-tone suppression. The dynamic version of the gammachirp filter (Irino and Patterson, 2005, 2006) also provides fast-acting compression that operates dynamically within the glottal cycle to compress glottal pulses while maintaining good frequency resolution for the analysis of vocal-tract resonances. It is argued that this dynamic adjustment of filter properties improves the robustness of speech recognition. There is no prospect of implementing fast-acting compression in a dynamic filter with discontinuous roex filters.

The success of the gammachirp family of filters has prompted us to develop a compressive, gammachirp filter system with parallel architecture like that of the parallel roex filter [Fig. 1(a)]. The architecture of this parallel cGC filter is shown in Fig. 1(c); the filter is described in Sec. II D. In essence, tip and tail gammachirp filters are substituted for the tip and tail roex filters; the compression mechanism is the same in the two systems. The behavior of this parallel cGC filter is quite similar to that of the parallel roex filter, but it has the distinct advantage of providing a basis for a time-domain implementation, just as the gammatone filter provided the basis for a time-domain implementation of the roex(p) filter. In this paper, we make a quantitative comparison of the parallel roex filter, the parallel cGC filter, and the cascade cGC filter to determine whether there remain aspects of simultaneous masking that are still better explained by, or

more compactly summarized by, a roex filter. The comparison is based on the simultaneous notched-noise masking data of Baker *et al.* (1998) and Glasberg and Moore (2000) Briefly, the results show that the parallel roex and parallel cGC filters provide comparable fits to simultaneous masking data, *and* the fits obtained are both somewhat unstable when both the tip and tail filters are complete. The cascade cGC filter provides as good a fit as either of the parallel filters to the simultaneous masking data, *with 25% fewer coefficients*, and the fits are stable. The advantage appears to be a direct result of the architecture of the gammachirp filter with its parsimonious representation of the high-frequency side of the filter.

B. The roex family of filters

In the original rounded-exponential filter of Patterson and Nimmo-Smith (1980), a cubic polynomial was used to do the rounding. The quadratic term was intended to broaden the sharp peak of the back-to-back exponentials and adjust the curvature in the region of the center frequency to match the curvature of the data at the top of the threshold function. The cubic term was intended to elevate the tails of the filter outside the passband, but the interaction of the cubic term with the exponential often led to an unrealistic oscillation in the filter characteristic where the passband gave way to the tails. Moreover, in the extremities, the tail had downward curvature which was also unrealistic.

Subsequently, Patterson *et al.* (1982) proposed a second form of roex filter which was the linear sum of two simple roex filters—one to represent the *p*assband of the filter [roex(p)] and the other to represent the *t*ails of the filter outside the passband [roex(t)]. In both cases, the exponentials were only rounded at the top, and the rounding was effected by a linear term for simplicity. The relative weight of the component filters was represented by a fixed coefficient, w, and the composite filter was referred to as the roex(p,w,t) auditory filter. Provided the tail roex has shallower slopes and less gain than the passband roex, the skirts of the passband will blend smoothly and monotonically into the tails of the filter, and this is a distinct advantage over the original roex filter with its cubic polynomial. In addition, Glasberg *et al.* (1984b) showed how the level-dependent asymmetry of the auditory filter could be implemented by allowing the parameters of the roex(p,w,t) filter to have different values on the *l*ower and *u*pper sides of the filter, producing a roex($p_l,p_u;w_l,t_l,w_u,t_u$) filter with six filter parameters. Different versions of this filter with roex(p) and roex(t) components operating in parallel proved useful as representations of the auditory filter, primarily because there are relatively few parameters in the system. For example, when the notch width is limited, the tail roex can be replaced by a fixed floor, to produce a roex(p,r) filter, or it can be omitted entirely, to produce a roex(p) filter (Patterson *et al.*, 1982; Patterson and Moore, 1986). Before proceeding to the details of the filter systems, we describe recent developments concerning the representation of the auditory filter by roex and gammachirp filters.

C. Roex and cGC fits to the tip of the auditory filter

It is important to note that the curvature of the parallel roex filters, in the region of the center frequency, is determined by the passband parameter, p; there is no separate parameter, like the quadratic term in the original roex (Patterson and Nimmo-Smith, 1980), to adjust the curvature around the tip. The value of the passband parameter, p, is largely determined by thresholds associated with intermediate and large notch widths (which are in turn determined by the slope of the skirt of the filter), rather than by thresholds associated with small notch widths (which are determined by the shape of the filter close to the tip). The most detailed measurement of the tip of the auditory filter is presented in Patterson (1976, Fig. 3), where threshold was measured in a symmetric notched noise for 17 notch widths (relative notch widths between 0.005 and 0.3), for probe frequencies of 0.5, 1.0, and 2.0 kHz. The data were fitted with a fifth-order polynomial to ensure that the shape was not unduly constrained by the threshold function; the polynomial values for the fit and the derived auditory filters are presented in Patterson (1976, Table A.I). We refer to the filter obtained in this way as the "unconstrained" filter.

We have compared the unconstrained auditory filter with the roex(p) and cascade cGC filters in the region of the passband (relative notch widths between 0.0 and 0.3). In the region around 0.1, the best fitting *roex* filter is consistently below the unconstrained filter by about 0.5 dB, on average, and in the region around 0.25, it is consistently above the unconstrained filter by a similar amount. This means that: (a) the auditory filter is flatter than the roex(p) filter at its center frequency; (b) the bandwidth of the auditory filter is slightly greater than that of the roex(p) filter; and (c) the maximum slope of the skirt that defines the passband of the auditory filter is slightly steeper than suggested by the roex(p) filter. The rms errors for the three roex fits were 0.59, 0.63, and 0.52 dB, for probe frequencies of 0.5, 1.0, and 2.0 kHz, respectively. The fit of the cascade cGC filter is noticeably better; the rms errors for the three probe frequencies are 0.06, 0.05, and 0.1 dB, respectively. The consistent under- and over-estimates that arise with the roex filter do not exist in the fits of the cascade cGC filter to the data. The discrepancy between the auditory filter and the roex filter is not sufficiently large to be concerned about tip filter results reported in previous studies. The point to note, however, is that the parallel roex filter does not provide a better fit than the cascade cGC filter in the region of the passband, and the fit of the cascade cGC to the data defining the passband of the auditory filter is extremely good.

D. Using the parallel roex filter to represent cochlear filtering

During the latter half of the 1980's, research on basilar membrane motion suggested that the cochlear filter has two components; a tail filter associated with the passive motion of the basilar membrane which is observed at high stimulus levels, and a tip filter that emerges out of the tail filter as stimulus level *decreases* (Evans *et al.*, 1989; Goldstein, 1990, 1995; Allen, 1997). The parallel roex filter with its

passband and tail components seemed naturally suited to simulate the operation of such a system, and the parallel roex filter was used in a series of studies to fit human masking data, with the implication that the results would reflect the properties of human cochlear filtering (e.g., Glasberg and Moore, 2000; Oxenham and Shera, 2003; Baker and Rosen, 2006).

The initial studies using this approach were performed by Rosen, Baker, and colleagues (Rosen and Baker, 1994, Rosen *et al.*, 1998; Baker *et al.*, 1998), who replicated the asymmetric notched-noise experiment of Patterson and Nimmo-Smith (1980) using a very wide range of probe frequencies, and a very wide range of probe levels, producing a massive database of notched-noise data. They showed how level-dependent asymmetry and compression could be explained with the parallel roex filter [Fig. 1(a)] by making the weighting parameter, w, a function of level (Rosen and Baker, 1994; Baker *et al.*, 1998). When all of the filter parameters are level dependent, the resulting roex(p_l,p_u;w_l,t_l,w_u,t_u) filter can require the specification of 18 or more filter coefficients, which typically leads to unstable fits. Accordingly, Rosen and Baker (1994) developed a POLYFIT procedure to fit the parallel roex filter to all of the data at one probe frequency simultaneously, and Baker *et al.* (1998) showed that the parameter values for a range of center frequencies could be summarized with linear regression functions when the parameter values are plotted as a function of the log of center frequency. This reduced the total number of filter coefficients from about 18 to 12. This combination of procedures allowed them to assess the effect of making certain parameters level independent, or eliminating certain parameters altogether, on the goodness of fit (the rms deviation between obtained and predicted thresholds).

Despite the reduction in the number of filter coefficients, however, the authors noted that the fits were still somewhat unstable because the value of the high-frequency tail parameter, t_u, was usually close to that of the high-frequency passband parameter, p_u (sometimes t_u was actually greater than p_u), indicating that one component of the high-frequency side was not really necessary to achieve a good fit. Accordingly, they removed the upper half of the tail filter and demonstrated that this incomplete, parallel roex(p_l,p_u;w_l,t_l) filter provided an excellent fit to the data. Subsequently, Glasberg and Moore (2000), Bacon *et al.* (2002), Oxenham and Shera (2003), and Baker and Rosen (2006) gathered simultaneous or forward notched-noise masking data for different purposes, and reported similar findings; that is, when they fitted the complete, parallel roex filter to the data, the value of t_u was not properly constrained, and so they too adopted the incomplete, parallel roex(p_l,p_u;w_l,t_l) filter of Baker *et al.* (1998), which provided excellent fits to the newer data with two parameters fewer than the complete parallel roex filter.

Collectively these studies show that notched-noise data and the incomplete parallel roex filter can be used to map out important features of auditory frequency selectivity, at least for stationary stimuli specified in the spectral domain. Specifically the studies show: (i) The auditory filter becomes progressively more asymmetric as level increases at all center frequencies, primarily because the low-frequency tail applies progressively less attenuation; (ii) The gain of the tip filter at its peak frequency decreases relative to the gain of the tail filter as stimulus level increases—a finding that is often interpreted as level-dependent gain, implying that there is compression; (iii) In the range 0.25 to 1.0 kHz, the amount of inferred compression increases with filter center frequency.

In the studies described above, the high-frequency side of the auditory filter was characterized by a single roex function, p_u, which was assumed to be part of the tip filter [roex(p_l,p_u;w_l,t_l)]. This assumption was made because, at low and medium sound levels, the output of the filter for frequencies close to the center frequency is dominated by the tip filter, which has to have both low- and high-frequency skirts to avoid a discontinuity at the center frequency. However, this means that at high levels, where the tail filter dominates, the filter characteristic would once again have a prominent discontinuity at its center frequency. Such a filter would not be physically realizable. *Moreover*, the tail component corresponds to the basic low-pass filtering action of the passive basilar membrane. In other words, it is the tail filter in the parallel roex, rather than the tip filter, that corresponds to the high-frequency side of the filter in the cochlea. Removing the tail component from the high-frequency side of the filter seems conceptually wrong, since it is generally assumed that the tail component of the filter remains when the active mechanism in the cochlea is damaged (Moore, 1998).

In more general terms, the authors of the studies that employ the incomplete, parallel roex filter seem to assume that some form of real filter could be found with a magnitude characteristic not unlike that of the incomplete roex filter, which would justify its continued use as a model of the auditory filter. It should, for example, be possible to smooth the discontinuity by completing the tail filter with a very steep upper side that is fixed in shape and so does not require extra free parameters in the fitting process. Then, an FIR filter could be developed using the filter design technique of Oppenheim and Schafer (1975) for this static, complete roex filter. However, such a filter would be very limited in its ability to account for peripheral auditory processing, because using FIR techniques it would be extremely difficult (if not impossible) to produce a time-varying version of the filter that applies fast-acting compression, resembling the nonlinear responses of the basilar membrane. This is because an FIR filter with the frequency resolution observed in the cochlea would require coefficients that represent durations which are long relative to the time-varying coefficients required to represent the fast action. In summary, the use of the incomplete, parallel roex filter to represent cochlear filtering seems rather unsatisfactory in its current form.

E. The gammachirp family of filters

The development of the gammat*one* and gamma*chirp* filters is described in Patterson *et al.* (2003, Appendix A). The compressive gammachirp (cGC) filter is composed of a *p*assive gammachirp (pGC) filter and a high-pass asymmetric function (HP-AF) arranged in cascade as shown in Fig. 1(b).

The *p*GC filter simulates the action of the passive basilar membrane and the output of the pGC filter is used to adjust the level dependency of the active part of the filter, which is the HP-AF. The HP-AF is intended to represent the interaction of the cochlear partition with the tectorial membrane as suggested by Allen (1997) and Allen and Sen (1998).[3] The effect is to sharpen the low-frequency side of the combined filter, which produces a tip in the cGC filter shape at low to medium stimulus levels. Note, however, that there is no high-frequency side to this tip filter; it only produces high-pass filtering and level-dependent gain in the region of the peak frequency. The fact that there is no high-frequency side to the tip filter keeps the number of parameters to a minimum and avoids the instabilities encountered with the parallel filter systems where the high-frequency sides of the tip and tail filters interact.

In Patterson *et al.* (2003), the cascade cGC filter was fitted to the combined notched-noise masking data of Glasberg and Moore (2000) and Baker *et al.* (1998). It was found that most of the effect of center frequency could be explained by the function that describes the change in filter bandwidth with center frequency. Patterson *et al.* expressed the parameters describing the filter as a function of ERB_N-rate (Glasberg and Moore, 1990), where ERB_N stands for the average value of the equivalent rectangular bandwidth of the auditory filter as determined for young, normally hearing listeners at moderate sound levels (Moore, 2003).[4] Once the parameters were written in this way, the shape of the cGC filter could be specified for the entire range of center frequencies (0.25–6.0 kHz) and levels (30–80 dB SPL) using just six fixed coefficients. The cascade cGC filter has several advantages: (a) The compression it applies is largely limited to the frequencies close to the center frequency of the filter, as happens in the cochlea (Robles *et al.*, 1986; Recio *et al.*, 1998); (b) The form of the chirp in the impulse response is largely independent of level, as in the cochlea (Recio *et al.*, 1998; Carney *et al.*, 1999; de Boer and Nuttall, 2000); (c) The impulse response can be used with an adaptive control circuit to produce a dynamic, compressive gammachirp filter (Irino and Patterson, 2005, 2006) to enable auditory modeling in which fast-acting compression is applied as part of the filtering process.

The parallel cGC filter [Fig. 1(c)] was developed to provide a basis for a time-domain filter with parallel architecture and a level-dependent magnitude response like that derived with the complete, parallel roex filter [Fig. 1(b)], and also to determine whether differences between the fits provided by the parallel roex and cascade cGC filters were due to differences in filter architecture (parallel versus cascade) or differences in component filter shape (roex versus gammachirp). It is also the case that the parallel cGC filter has a similar architecture to the MBPNL filter of Goldstein (1990; 1995) and the DRNL filter of Meddis *et al.* (2001), both of which are used to model a range of physiological phenomena including compression and suppression. So, including the parallel cGC in the comparison provides a basis for future research comparing, for example, the compression and suppression observed psychophysically with that observed physiologically.

All three filters (parallel roex, parallel cGC, and cascade cGC) were fitted to the combined human masking data of Glasberg and Moore (2000) and Baker *et al.* (1998) to determine whether the gammachirp filters could fit simultaneous masking data as well as the complete roex filter, and to support discussion of which might provide the best representation of cochlear filtering, and the best time-domain, level-dependent filterbank for auditory modeling and speech processing. Although the cascade cGC filter has been implemented as a time-domain filter, the analysis presented in this paper is based only on the power spectra of the stimuli; the three filters were all implemented and evaluated in the power-spectrum domain, using the power-spectrum model of masking, which is outlined before the filters themselves are described in detail.

II. MASKING MODEL AND FILTER ARCHITECTURE

A. The power-spectrum model of masking

The most common method for estimating the shape of the auditory filter is based on the power-spectrum model of masking (Fletcher, 1940), and it involves data gathered with the notched-noise, simultaneous masking technique (Patterson, 1976). The listener is required to detect a sinusoid, referred to as the "probe," in the presence of a noise with a spectral notch in the region of the probe. If the edges of the noise band are steep, it is possible to write a function that relates the probe level at masked threshold to the integral of the auditory filter. The details of the latest version of the procedure are presented in Patterson *et al.* (2003). With regard to notation: f_c denotes the filter center frequency (in Hz), P_s denotes the probe level (in dB SPL), and N_0 denotes the masker spectrum level (in dB *re*:20 μPa) in the band below the probe frequency between $f_{l_{\min}}$ and $f_{l_{\max}}$ *and* in the band above the probe frequency between $f_{u_{\min}}$ and $f_{u_{\min}}$. If the auditory filter shape is represented as a weighting function, $W(f)$, then the probe level at threshold is given by

$$P_s = K + N_0 + 10\log_{10}\left\{\int_{f_{l_{\min}}}^{f_{l_{\max}}} W(f)df + \int_{f_{u_{\min}}}^{f_{u_{\max}}} W(f)df\right\}, \quad (1)$$

where K is a constant which is related to the efficiency of the detection mechanism following the auditory filter.

B. The parallel roex filter

In the case of the parallel roex filter (Patterson *et al.*, 1982, 2005),

$$W(f) = W_{\text{tail}}(f) + w_{\text{lin}}W_{\text{tip}}(f), \quad (2)$$

where $W_{\text{tail}}(f)$ and $W_{\text{tip}}(f)$ are the weighting functions corresponding to the roex tail filter and the roex tip filter, respectively; w_{lin} specifies the gain of the tip filter relative to the gain of the tail filter, in linear power units. The parallel roex auditory filter was originally written as $W(f)=w_{\text{lin}}W_{\text{tail}}(f)+(1-w_{\text{lin}})W_{\text{tip}}(f)$, so as to give 0 dB gain at the tip (Patterson *et al.*, 1982). We have rewritten it with a fixed tail filter [Eq. (2)] to be compatible with the

current conception of cochlear filtering, in which the passive tail filter has a fixed gain and the gain of the tip filter varies (Patterson, *et al.* 2005).

The relative gain of the tip filter depends on the input/output (I/O) function of the basilar membrane and, thus, the active mechanism of the cochlea. We use here the gain function proposed by the Glasberg and Moore (2000), which, in dB, is

$$w_{\mathrm{dB}} = 0.9L + A + B\left[1 - \frac{1}{1+\exp(-0.05(L-50))}\right] - L, \tag{3}$$

where L is the input level, $A=-0.0894G_{\max}+10.89$, $B=1.1789G_{\max}-11.789$, and $G_{\max}$ is the maximum gain applied by the cochlear amplifier in dB. The value of w_{dB} is related to the value of w_{lin} by: $w_{\mathrm{lin}}=10^{w_{\mathrm{dB}}/10}$. This I/O function imposes strong compression over the range 20 to 80 dB; outside this range the I/O function is almost linear. Baker and Rosen (2006) have pointed out that "w_{lin}" can be approximated by a linear function over the range 30 to 70 dB, as in Baker *et al.* (1998). The I/O function above is only really required if the intensity range needs to be extended beyond 30 to 70 dB.

The roex version of the tip filter can be characterized as a function of frequency f and three parameters,

$$W_{\mathrm{tip}}(f) = R_{\mathrm{x}}(f;f_c,p_l,p_u). \tag{4}$$

The expressions for the *l*ower and *u*pper sides of the tip filter are

$$R_{\mathrm{x}}(f;f_c,p_l,p_u) = \begin{cases} (1+d_c p_l g)\exp(-d_c p_l g) & f < f_c, \\ (1+d_c p_u g)\exp(-d_c p_u g) & f \geq f_c, \end{cases} \tag{5}$$

where $g=|f-f_c|/f_c$, and $d_c=[\mathrm{ERB_N}(1000)]/[\mathrm{ERB_N}(f_c)]\times f_c/1000$. The normalized frequency variable, g, describes the distance in frequency from the center frequency of the filter to the edge of the noise, relative to the center frequency. By using the variable d_c, the deviation from the center frequency is expressed relative to the value of $\mathrm{ERB_N}$ at that center frequency (Glasberg and Moore, 1990), and the value of d_c is normalized so as to have a value of unity when $f_c=1000$ Hz. So, the bandwidth of the tip filter is proportional to the value of $\mathrm{ERB_N}$ for all center frequencies (Patterson *et al.*, 2003). Parameters p_l and p_u determine the sharpness of the *l*ower and *u*pper skirts of the *p*assband of the roex filter (that is, the tip filter).

The roex version of the tail filter can be expressed in a similar way,

$$W_{\mathrm{tail}}(f) = R_{\mathrm{x}}(f;f_c/f_{\mathrm{rat}},t_l,t_u), \tag{6}$$

$$R_{\mathrm{x}}(f;f_c,t_l,t_u) = \begin{cases} (1+d_c t_l g)\exp(-d_c t_l g) & f < f_c \\ (1+d_c t_u g)\exp(-d_c t_u g) & f \geq f_c \end{cases}. \tag{7}$$

Parameters t_l and t_u describe the sharpness of the tails of the filter. The parallel roex filter is a weighted sum of the filters described by Eqs. (5) and (7), the weighting being determined by Eq. (3). This parallel roex filter differs from that used by Baker *et al.* (1998) and by Glasberg and Moore (2000) in two ways: (1) both the tip and tail filters have a lower skirt and an upper skirt; (2) the center frequency of the tip filter is allowed to shift with level. The shift is defined by parameter f_{rat}, which is the ratio of the center frequencies of the tip and the tail filters, and is given by

$$f_{\mathrm{rat}} = f_{\mathrm{rat}}^{(0)} + f_{\mathrm{rat}}^{(1)} P_{\mathrm{rxp}}, \tag{8}$$

where the superscripts 0 and 1 designate the intercept and slope of the line defining f_{rat}, and P_{rxp} is the level of the probe-plus-masker at the output of the tail filter whose center frequency is equal to the probe frequency. This is the same level as that derived from the passive gammachirp, P_{gcp}, in Patterson *et al.* (2003).

We distinguish here between *parameters* and *coefficients*. The parallel roex filter is characterized by six parameters, t_l, t_u, p_l, p_u, $G_{\max}$, and f_{rat}. However, some of these parameters are functions of P_{rxp}, so the number of coefficients required to characterize the filter is greater than six; details are provided as the issue arises in later sections. Parameters t_l and t_u are level independent and determine the shape of the tail filter (bandwidth and asymmetry) as shown in the top block in Fig. 1(a). The value of P_{rxp} determines the values of the parameters of the tip filter: f_{rat}, p_l, p_u, and w_{dB}. The relative gain of the tip filter, w_{dB}, is determined by Eq. (3) and depends on $G_{\max}$. Parameters p_l and p_u control the shape of the tip filter (bandwidth and asymmetry), as illustrated in the "active tip filter" block of Fig. 1(a). These parameters are also level dependent: $p_l=p_l^{(0)}+p_l^{(1)}P_{\mathrm{rxp}}$ and $p_u=p_u^{(0)}+p_u^{(1)}P_{\mathrm{rxp}}$. The tip filter broadens as P_{rxp} increases.

C. The cascade cGC filter

We will use the terminology developed by Patterson *et al.* (2003) for the compressive gammachirp (cGC) filter. The weighting function in the power-spectrum domain is

$$W(f) = |G_{CC}(f)|^2, \tag{9}$$

and it is a cascade of a passive gammachirp (pGC), $|G_{CP}(f)|$, and a high-pass asymmetric function (HP-AF), $\exp(c_2\theta_2(f))$, as illustrated in Fig. 1(b). Thus,

$$|G_{CC}(f)| = |G_{CP}(f)|\exp[c_2\theta_2(f)]. \tag{10}$$

The pGC filter is itself a cascade of a gamma*tone* (GT) filter and a low-pass asymmetric function (LP-AF); that is,

$$|G_{CP}(f)| = a_\Gamma |G_T(f)|\exp[c_1\theta_1(f)]. \tag{11}$$

$|G_T(f)|$ is the Fourier magnitude spectrum of the GT filter and a_Γ is the amplitude of the Fourier magnitude of the GT. The antisymmetric functions $\theta_1(f)$ and $\theta_2(f)$ are

$$\theta_1(f) = \arctan\left(\frac{f-f_{r1}}{b_1\mathrm{ERB_N}(f_{r1})}\right), \tag{12}$$

and

$$\theta_2(f) = \arctan\left(\frac{f-f_{r2}}{b_2\mathrm{ERB_N}(f_{r2})}\right), \tag{13}$$

where $\mathrm{ERB_N}(f)$ is the value of $\mathrm{ERB_N}$ at frequency f (Glasberg and Moore, 1990), f_{r1} is the asymptotic center fre-

quency of the chirp, and f_{r2} is the center frequency of the HP-AF. For further details see Irino and Patterson (1997, 2001) and Patterson *et al.* (2003).

The cGC filter has five parameters: b_1, c_1, b_2, c_2, and f_{rat}. Parameters b_1 and c_1 control the bandwidth and asymmetry of the pGC filter, respectively, as shown in the first block of Fig. 1(b). The peak frequency of the pGC is $f_{p1}=f_{r1}+c_1 b_1 \text{ERB}_N(f_{r1})/n_1$, where n_1 is the order of the gamma function which is fixed at 4 as in previous studies. Parameters b_2 and c_2 control the slope and dynamic range of the HP-AF, respectively, as shown in the second block of Fig. 1(b). The dynamic range determines the amount of compression; f_{rat} is the ratio of the peak frequency of the passive GC (f_{p1}) to the center frequency of the HP-AF (f_{r2}); so $f_{\text{rat}}=f_{r2}/f_{p1}$. It describes the frequency shift of the HP-AF *relative* to the pGC as a function of level,

$$f_{\text{rat}} = f_{\text{rat}}^{(0)} + f_{\text{rat}}^{(1)} P_{\text{gcp}}. \tag{14}$$

Here, P_{gcp} is the level that controls the position of the HP-AF and produces the compression and gain of the compressive GC filter. It is the level of the probe plus masker at the output of the passive GC filter (Patterson *et al.*, 2003).

D. The parallel cGC filter

The parallel, compressive GC filter consists of a level-independent pGC filter that represents the passive basilar membrane [top block in Fig. 1(c)] and a level-dependent GC filter with an I/O function similar to that of the cochlea [bottom blocks in Fig. 1(c)]. So, the tail filter is

$$W_{\text{tail}}(f) = |G_C(f; f_c/f_{\text{rat}}, b_1, c_1)|^2, \tag{15}$$

and the tip filter is

$$W_{\text{tip}}(f) = |G_C(f; f_c, b_2, c_2)|^2. \tag{16}$$

Here, f_c is the center frequency of the composite filter with parallel architecture. The basic gammachirp filter, $|G_C(f)|$, is

$$|G_C(f; f_p, b, c)| = a_\Gamma \cdot |G_T(f)| \exp[c\theta(f)], \tag{17}$$

where $\theta(f)=\arctan((f-f_r)/[b\text{ERB}_N(f_r)])$ and $f_p=f_r+cb\,\text{ERB}_N(f_r)/n$. Here, f_r is the asymptotic frequency of the chirp, f_p is the peak frequency of the gammachirp filter, and n is the order of the gamma function ($n=4$). The ratio of the peak frequency of the tail and tip filters is $f_{\text{rat}}=f_{p2}/f_{p1}$, and it has the same form as in Eq. (14).

This parallel cGC filter has six parameters: b_1, c_1, b_2, c_2, $G_{\max}$, and f_{rat}. Parameters b_1 and c_1 control the bandwidth and asymmetry of the tail filter, respectively, as shown in the top block of Fig. 1(c). The level, P_{gcp}, is determined by the output of the pGC and it determines f_{rat} and the parameters of the tip filter, b_2 and c_2, and w_{dB}. Parameters b_2 and c_2 control the bandwidth and asymmetry of the tip filter. The filter gain, w_{dB}, is determined by Eq. (3) and depends on $G_{\max}$ and P_{gcp}. The parameters of the tip filter, b_2 and c_2, are also affected by level: $b_2=b_2^{(0)}+b_2^{(1)}P_{\text{gcp}}$ and $c_2=c_2^{(0)}+c_2^{(1)}P_{\text{gcp}}$. The tip filter broadens as P_{gcp} increases.

III. EVALUATION OF THE FILTER SYSTEMS

A. The notched-noise masking data

We used two large sets of notched-noise masking data: those of Baker *et al.* (1998) and Glasberg and Moore (2000). Baker *et al.* (1998) measured masked threshold using both fixed probe levels (P_s=30, 40, 50, 60, and 70 dB SPL) and fixed masker levels (N_0=20, 30, 40, and 50 dB), and threshold was measured at each of seven probe frequencies: 0.25, 0.5, 1.0, 2,0, 3.0, 4.0, and 6.0 kHz. The frequencies of the inner edges of the masker bands were varied to produce both symmetric and asymmetric notches about the probe frequency; there were 16 notch conditions with the outer edges of the masker bands fixed at $g=\pm 0.8$. The total number of masked thresholds was 973 per listener and there were two listeners. Glasberg and Moore (2000) measured masked threshold with fixed masker levels of 35, 50, 65, and 80 dB/ERB_N at their lowest probe frequency, 0.25 kHz, and with fixed masker levels of 40, 55, and 70 dB/ERB_N for probe frequencies of 0.5, 1.0, 2.0, and 4.0 kHz. They used "uniformly exciting noise," designed to produce constant cochlear excitation for center frequencies within the passbands of the noise. There were 19 notch conditions, with the notch positioned both symmetrically and asymmetrically about the probe frequency. In this study, the upper and lower noise bands that formed the notch had the same bandwidth, which was $0.4f_c$. The total number of masked thresholds was 304 per listener and there were three listeners.

B. The fitting procedure

The procedure for fitting each of the three auditory filter models to the combined data sets of Baker *et al.* (1998) and Glasberg and Moore (2000) is almost the same as that described by Patterson *et al.* (2003). Broadly speaking, it is the POLYFIT procedure of Baker *et al.* (1998) extended to include simultaneous fitting of all probe frequencies with the filter coefficients represented in terms of normalized ERB_N-rate, that is,

$$E_f = \text{ERB}_N\text{-rate}(f)/\text{ERB}_N\text{-rate}(1\text{ kHz}) - 1.$$

Note in particular that $G_{\max}$ is assumed to be a linear function of E_f, and therefore is characterized by two coefficients.

During the development of the cGC fitting procedure (Patterson *et al.*, 2003), it was noticed that the slope of the I/O function of the composite filter occasionally became slightly negative at higher stimulus levels. This unrealistic solution arises when the fitting procedure adopts a pGC filter whose peak frequency does not lie on the steep central section of the HP-AF. A smooth penalty function was introduced to exclude such fits. The resulting procedure effectively matches the amount of compression produced by the HP-AF to that observed in the data (Patterson *et al.*,). When we began fitting the parallel roex filter and the parallel cGC filter to the masking data, we found that a similar problem arises if the fit does not include a penalty function for unrealistic I/O functions. These "solutions" arise, as previously, at higher stimulus levels when the fitting procedure uses a tail filter with an overly narrow bandwidth, or a tip filter with an

TABLE I. Coefficients and rms error values for auditory filters with cascade and parallel architectures, fitted to the combined masking data of Baker *et al.* (1998) and Glasberg and Moore (2000). The first and second columns show the type of filter architecture and the total number of filter coefficients (not including nonfilter coefficients, K and P_0), respectively. Each filter parameter was represented either by a constant or by a linear regression line as a function of normalized ERB_N value, E_f. The central columns in the table show the filter coefficients on the left and the nonfilter coefficients on the right. The right-most column shows the rms error value for the fit in dB. In this table, "0" means that we set the corresponding value to zero.

Filter architecture	No. coeff	b_1	c_1	$f_{rat}^{(0)}$	$f_{rat}^{(1)}$	b_2	c_2	K (dB)	P_0 (dB)	rms (dB)
Cascade		1.85	−2.51	0.391	0.0124	2.01	1.85	−4.66	17.3	
cGC	12	$+0.111E_f$	$+0.03E_f$	$+0.15E_f$	$-0.003E_f$	$-0.012E_f$	$+0.359E_f$	$-2.74E_f$ $+8.22E_f^2$	$-0.220E_f$ $+4.54E_f^2$	3.59
							1.86	−4.66	17.3	
	7	1.85	−2.50	0.435	0.0115	2.11	$+0.381E_f$	$-2.75E_f$ $+8.22E_f^2$	$-0.230E_f$ $+4.54E_f^2$	3.64
								−3.73	16.8	
	6	1.81	−2.96	0.466	0.0109	2.17	2.20	$-4.89E_f$ $+8.30E_f^2$	$-1.27E_f$ $+5.74E_f^2$	3.71
Filter architecture	No. coeff	t_1	t_u	G_{max}		$p_l^{(0)}$	$p_l^{(1)}$	$p_u^{(0)}$	$p_u^{(1)}$	K (dB)	P_0 (dB)	rms (dB)
Parallel		10.9	67.9			35.1	−0.186	36.5	−0.230	−1.89	16.9	
roex	14	$-1.74E_f$	$+8.30E_f$	$35.4+17.1E_f$		$+9.69E_f$	$-0.083E_f$	$-9.27E_f$	$+0.044E_f$	$-3.10E_f$ $+5.95E_f^2$	$-2.76E_f$ $+4.95E_f^2$	3.58
										−2.26	16.9	
	8	10.3	68.0	$35.9+17.1E_f$		36.4	−0.176	35.5	−0.250	$-1.16E_f$ $+6.63E_f^2$	$-2.41E_f$ $+5.09E_f^2$	3.79
										−1.38	15.7	
	6	12.6	153.5	$25.4+3.44E_f$		33.0	0	23.5	0	$-2.04E_f$ $+7.29E_f^2$	$-1.23E_f$ $+4.89E_f^2$	4.15
Filter architecture	No. coeff	b_1	c_1	G_{max}		$b_2^{(0)}$	$b_2^{(1)}$	$c_2^{(0)}$	$c_2^{(1)}$	K (dB)	P_0 (dB)	rms (dB)
Parallel		0.636	−8.54			0.683	0.013	1.05	−0.028	−1.55	16.8	
cGC	14	$-0.002E_f$	$-1.74E_f$	$38.7+26.0E_f$		$-0.266E_f$	$+0.010E_f$	$+1.04E_f$	$-0.0001E_f$	$-4.38E_f$ $+7.15E_f^2$	$-2.75E_f$ $+6.15E_f^2$	3.53
										−1.23	17.2	
	8	0.611	−9.02	$36.9+24.0E_f$		0.531	0.016	−2.67	0.061	$-1.97E_f$ $6.87E_f^2$	$-1.80E_f$ $+6.25E_f^2$	3.75
										1.93	16.0	
	6	0.394	−10.6	$34.0+23.5E_f$		1.15	0	−0.338	0	$-1.37E_f$ $+7.20E_f^2$	$-1.93E_f$ $+6.06E_f^2$	4.87

overly wide bandwidth. Provided the procedure is restricted to a reasonable combination of tail filter and tip filter, the slope of the I/O function remains positive.

In this paper, the tip filter is intended to represent the active mechanism in the cochlea. We assume that: (a) it becomes broader as level increases; (b) the tip filter dominates the shape of the composite filter at low levels where the ERB of the tip filter is close to that of the composite filter; and (c) the tail filter dominates the shape at high levels where the ERB of the tail filter is close to that of the composite filter. A smooth penalty function was used with these constraints to restrict the ERBs of the component filters and so prevent the fitting procedure from adopting unrealistic solutions. This penalty function is "*the weighted sum of the difference between the ERB/f_c of the tip filter and the composite filter*" at low levels and "*the difference between ERB/f_c of the tail filter and the composite filter*" at higher levels. Normalizing the ERB values by center frequency, f_c, makes it possible to use the same penalty function across the entire region of center frequencies from 0.25 to 6.0 kHz.

In summary, the filter parameters for the compressive and parallel GC filters are b_1, c_1, b_2, c_2, f_{rat} and b_1, c_1, b_2, c_2, G_{max}, f_{rat}, respectively. The filter parameters for the parallel roex filter are t_l, t_u, p_l, p_u, G_{max}, f_{rat}. There are also two nonfilter parameters, K and P_0: K is the efficiency of the detection mechanism in Eq. (1); P_0 is related to the absolute threshold, and is used to predict the lower limit of threshold in the masking experiments. For the cGC filter, only one of the filter parameters is level dependent and that is f_{rat}. The ratio is linearly dependent on level and the intercept and slope are designated $f_{rat}^{(0)}$ and $f_{rat}^{(1)}$. So, the fit only requires six filter coefficients for the entire range of center frequencies and levels. The parallel cGC filter and the parallel roex filter are also quite parsimonious in the use of filter coefficients, but, in both cases, the tip filter has two slope parameters, and both have to be level dependent to produce a good fit. The central columns of Table I show that, if the linear term is set to 0 in either of the parallel filters, the rms error rises to values above 4 dB. For the parallel roex filter, the coefficients are t_l, t_u, $p_l^{(0)}$, $p_l^{(1)}$, $p_u^{(0)}$, $p_u^{(1)}$, G_{max}, $f_{rat}^{(0)}$, and $f_{rat}^{(1)}$; for the parallel cGC, they are b_1, c_1, $b_2^{(0)}$, $b_2^{(1)}$, $c_2^{(0)}$, $c_2^{(1)}$, G_{max}, $f_{rat}^{(0)}$,

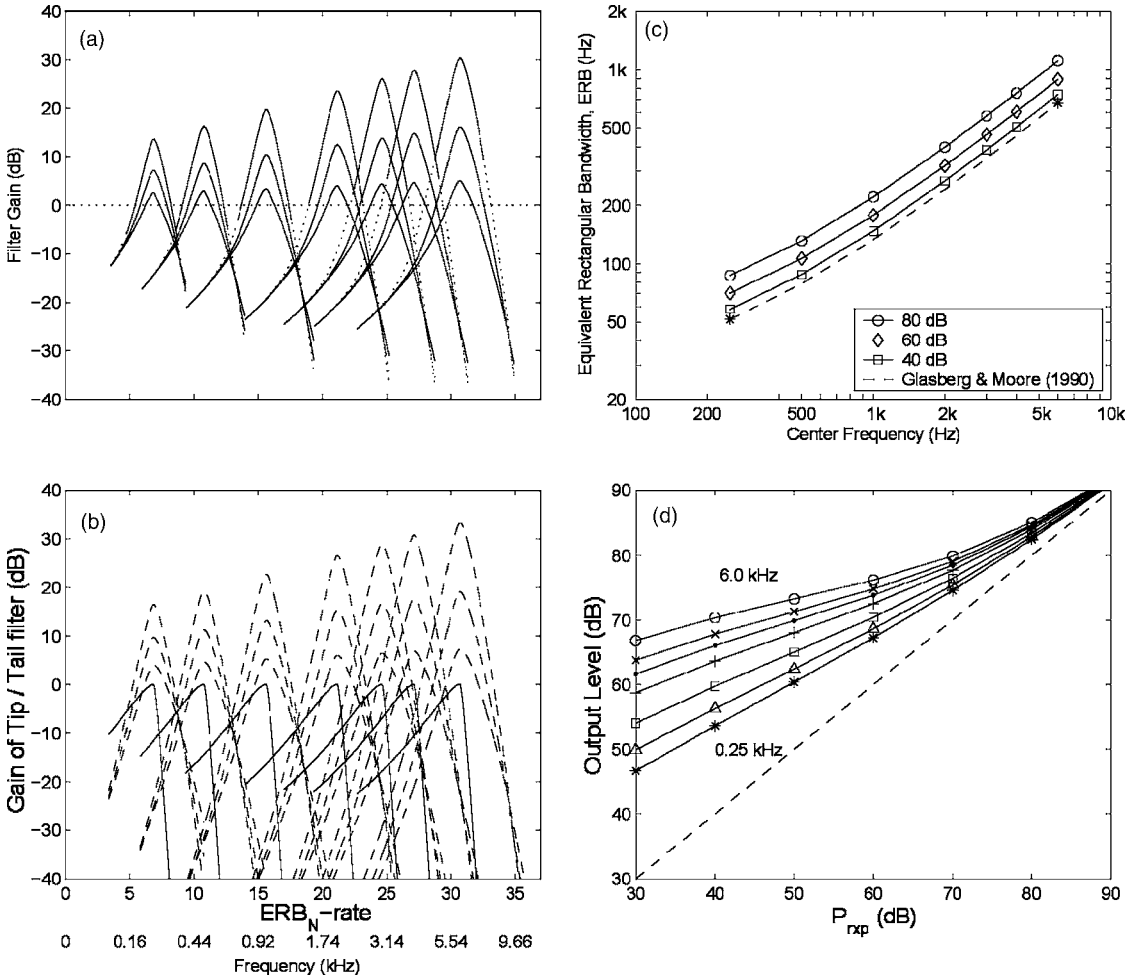


FIG. 2. Characteristics of the parallel roex auditory filter fitted to the combined data of Baker *et al.* (1998) and Glasberg and Moore (2000) for center frequencies from 0.25 to 6.0 kHz: (a) families of parallel roex filters and (b) their component (tip/tail) filters showing how the functions change with P_{rxp}, the output level of the tail filter; (c) Equivalent rectangular bandwidth, ERB, of the parallel roex filter (solid line) as a function of center frequency on a log-frequency scale. The parameter is the value of P_{rxp}; for a narrow-band signal at the center frequency of the filter this is equivalent to the input level in dB SPL. The dashed line shows the $\mathrm{ERB_N}$ values proposed by Glasberg and Moore (1990); (d) Input/output functions for the parallel roex filter; the parameter is center frequency. This version of the parallel roex filter had eight coefficients (see Table I).

and $f_{\mathrm{rat}}^{(1)}$. In both cases, the value of $G_{\max}$ is a linear function of E_f (defined by two coefficients), so ten filter coefficients are required.

Masked threshold was predicted for a range of filters with center frequencies around the probe frequency, using the power spectrum model of masking defined by Eq. (1); for each notched-noise condition, the center frequency was chosen so as to maximize the probe-to-masker ratio at the output of the filter. The Levenberg-Marquardt method (Press *et al.*, 1988) was used to minimize the root-mean-square (rms) difference between masked threshold and the predicted value of masked threshold (both in dB); this is hereafter called the "error."

C. Results

1. The parallel roex filter

As described earlier, for the parallel roex filter, the center frequency of the tail roex was fixed and that of the tip roex was allowed to vary with level, i.e., f_{rat} was allowed to vary. However, it was found that this had little effect on the fit, and produced negligible reduction in the rms error. Accordingly, for subsequent analyses, we set $f_{\mathrm{rat}}=1.0$, which means that the peak frequencies of the tail and tip filters were the same. This reduced the number of filter parameters by one and the number of fitting coefficients by two. With f_{rat} fixed, the total number of filter coefficients was eight; the coefficients and their values are in the eight-coefficient row of the central section of Table I. The rms error when fitting the combined data of Baker *et al.* (1998) and Glasberg and Moore (2000) was 3.79 dB. The rms error seems reasonable given that the masked thresholds come from two separate studies using different subjects, and they cover a very wide range of probe frequencies and levels.

The families of parallel roex filters fitted to the combined data are presented in Fig. 2, for seven probe frequencies, from 0.25 to 6.0 kHz, and for three values of P_{rxp}, the level at the output of the tail filter. In Fig. 2(a), the top filter in each group shows the filter shape for a P_{rxp} value of 40 dB; the two lower filters are for P_{rxp} values of 60 and 80 dB, respectively. The corresponding tail filters and tip filters are shown in Fig. 2(b) by solid and dashed lines, respectively. These component filters operate in parallel to produce

the composite filters in Fig. 2(a). The ordinate in Fig. 2(a) is the gain of the composite filter; the ordinate in Fig. 2(b) is the gain of the tail roex *or* the tip roex in isolation. The abscissas in Figs. 2(a) and 2(b) are both ERB_N-rate. Note that plotting the filters in terms of gain rather than absolute level gives the impression that we can specify the form of the *lower-level* filters over a range of as much as 80 dB. This is clearly not the case; there are no threshold values from the notched-noise experiment to specify the tails of the lower level filters, which correspond to levels below absolute threshold. Accordingly, the line used to represent each composite filter changes its form from solid to dotted at the point where the filter function would intersect the threshold limit, P_0.

Because of the parallel architecture, the gain of the composite filter is determined over most of its range by whichever of the two component filters has the greater gain. For signal frequencies near the peak of the composite filter, the tip filter dominates at most signal levels. For signal frequencies far below the center frequency, the tail filter dominates. When P_{rxp} is equal to 100 dB, the gain of the tip filter at its peak is the same as the gain of the tail filter at its peak; both have a gain of 0 dB. As a result the composite filter has a gain of 3 dB. To facilitate comparison with the cascade filter, the gains of the composite filters in Fig. 2(a) have all been scaled down by 3 dB, so that the gain is 0 dB when P_{rxp} is equal to 100 dB. It is noteworthy that the high-frequency side of the tip filter is relatively sharp, so that, for input frequencies above the filter center frequency, the output is determined by the tip filter, the tail filter having little influence. This explains why Baker *et al.* (1998) and Glasberg and Moore (2000) were able to model their data using a single roex function for the high-frequency side.

The families of parallel roex filters derived here all have the same general form [Fig. 2(a)]; the gain changes smoothly as center frequency increases. The gain of the composite filter decreases monotonically with increasing stimulus level and it increases monotonically with increasing center frequency. The maximum gain for the parallel roex filter (for P_{rxp}=40 dB) is about 30 dB at the highest center frequency, 6.0 kHz. The gain is largely restricted to a range of ±2 ERB_N around the center frequency of the filter, and the filter skirts for the different levels converge both on the low side and the high side, indicating that the gain is independent of level for frequencies far removed from the center frequency.

Figure 2(c) shows the ERB of the parallel roex filter as a function of center frequency for three values of P_{rxp}, 40, 60, and 80 dB, the same values as for the filters in Fig. 2(a). The bandwidth increases slowly with level and the bandwidth function has the same shape at all levels. The dashed line shows the ERB_N function for the roex filter suggested by Glasberg and Moore (1990); it is very similar to the function for the parallel roex filter at low levels, as would be expected.

Filter output level is plotted as a function of P_{rxp} in Fig. 2(d); the parameter is center frequency, which varies from 0.25 to 6.0 kHz. For a narrow-band signal at the center frequency of the parallel roex filter, the value of P_{rxp} is very close to input level in dB SPL. So, these I/O functions illustrate the form of the nonlinearity in the system. The dashed line shows the I/O function for a linear system with 0-dB gain. The varying vertical distance from the dashed line to one of the solid lines indicates how the gain varies with level for a given center frequency. The I/O functions all show compression and the amount of compression increases with increasing center frequency. The slope of the I/O function between 40 and 70 dB decreases as center frequency increases. These I/O functions are similar to those derived previously from simultaneous masking data (Baker *et al.*, 1998; Glasberg and Moore, 2000), but the amount of compression is somewhat less than that derived from forward masking data (Oxenham and Plack, 1997). A more detailed analysis of the compression is presented later.

2. The cascade cGC filter

The families of cascade cGC filters fitted to the data are shown in Fig. 3, in the same format as for Fig. 2. The composite cGC filters in Fig. 3(a) are produced by cascading the filters in Fig. 3(b); the pGC filters are shown by solid lines and the HP-AF filters are shown by dashed lines. The ordinate in Fig. 3(b) is component filter gain with the values for the pGCs on the left and those for the HP-AFs on the right. Since this is a cascade filter, the composite filter values are the sum of the component filter values in decibels. To facilitate comparison with the parallel roex and parallel cGC filters, the gains of the composite filters have been scaled up by a fixed factor at each center frequency, so that the peak gain becomes 0 dB for all center frequencies when P_{gcp} is 90 dB (this scaling is similar to that for the parallel roex filter). The scaling factors were 10.0, 13.6, 15.4, 16.3, 16.6, 16.7, and 16.9 dB, for center frequencies of 0.25, 0.5, 1.0, 2.0, 3.0, 4.0, and 6.0 kHz, respectively.

Six filter coefficients are required to characterize the cGC filter; they are listed in the six-coefficient row of the upper section of Table I. Although the cGC filter has two fewer filter coefficients than the parallel roex filter, the rms error was slightly less (3.71 dB) than for the latter.

The level dependence of the cGC filter arises from the frequency shift of the HP-AF relative to the pGC filter. These two filters appear as triplets of dashed and solid lines in Fig. 3(b); note that the steep center section of the HP-AF filter crosses the corresponding pGC filter on its *upper skirt*, and the shift of the HP-AF with level is large relative to the shift of the pGC filter. The asymmetry of the composite, cascade cGC filter in Fig. 3(a) is uniform across center frequency on the ERB_N-rate scale. The gain of the composite filter decreases as stimulus level increases at all center frequencies, and the change in gain is largely restricted to a range of about ±2 ERB_N around the center frequency. The filter skirts converge on both the lower and upper sides of the filter. The maximum gain at 0.25 kHz is about 18 dB; the gain increases with center frequency to about 27 dB at 2.0 kHz. Beyond 2.0 kHz, it remains at about 28 dB, and so the pattern of maximum gain across frequency is different for the parallel roex and the cascade cGC filters.

Figure 3(c) shows that the ERB of the cascade cGC filter is consistently wider than ERB_N and the ERB of the corresponding parallel roex filter. The difference is greatest at the

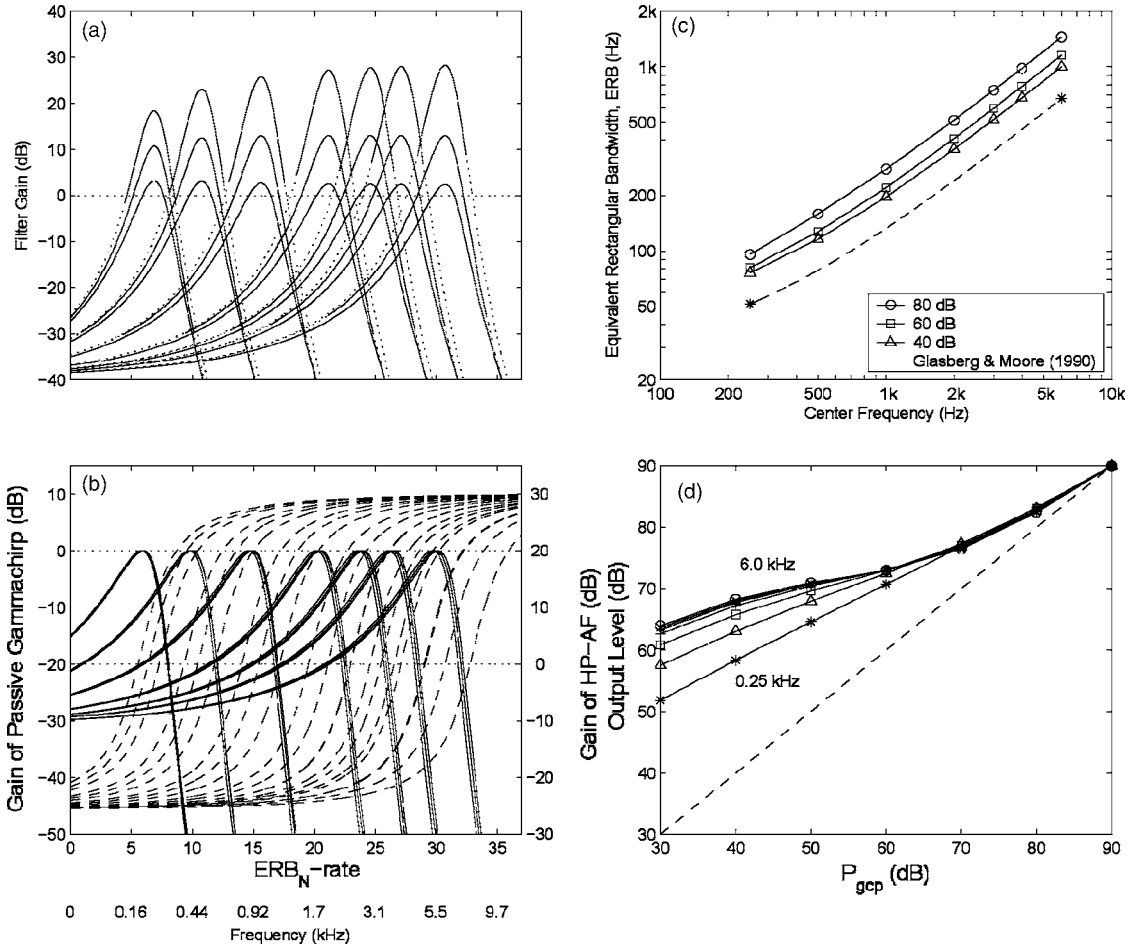


FIG. 3. Characteristics of the cascade cGC filter in the same format as Fig. 2. The three curves within each family show how the functions change with P_{gcp}, the output level of the tail filter. In panel (d), the compression functions for center frequencies from 2.0 to 6.0 kHz are superimposed. This version of the cascade cGC filter had six coefficients (see Table I).

lowest level, where the ERB of the cascade cGC is about 1.5 times that of the parallel roex. As the value of P_{gcp} increases from 40 to 60 dB, the bandwidth increase is relatively small; then, there is a relatively large increase as P_{gcp} increases from 60 to 80 dB. This nonuniform growth in bandwidth is the result of the interaction of the component filters. Until the steep section of the HP-AF is above the peak frequency of the pGC, increases in level have relatively little effect on bandwidth; thereafter they have a greater effect, as the shape of the cGC filter approaches that of the pGC filter.

Figure 3(d) shows that the I/O functions for the cascade cGC filter are similar in shape to those of the parallel roex filter, but the spread of the functions is reduced. At center frequencies below 2.0 kHz, the cascade cGC system produces a little more gain than the parallel roex system, whereas at center frequencies above 3.0 kHz, the cascade cGC system produces a little less gain than the parallel roex system.

3. The parallel cGC filter

The families of parallel cGC filters fitted to the data are shown, in the same format, in Fig. 4. The component filters are shown in Fig. 4(b); the tail and tip filters are represented by solid and dashed lines, respectively. They operate in parallel to produce the composite filters in Fig. 4(a). The gains of the composite filters have all been scaled down by 3 dB, in the same way as for the parallel roex filter, and for the same reason. The parameter f_{rat} was set to 1.0, as with the parallel roex, because varying the ratio produced virtually no improvement in the fit. The parallel cGC filter is characterized by eight filter coefficients; they are listed in the eight-coefficient row of the lower section of Table I. The rms error was 3.75 dB, which is similar to the values for the other two filters.

Figure 4 shows that the behavior of the parallel cGC filter system is more like that of the parallel roex filter than that of the cascade cGC filter. Although the tip filter has the shape of a gammachirp filter, it is narrower than the tip of the corresponding cascade cGC [Fig. 4(b)], and this narrower tip filter carries through to the tip of the composite parallel cGC [Fig. 4(a)]. The functions in Fig. 4(c) show that, at low levels, the ERB of the parallel cGC filter is similar to ERB_N and to the ERB of the parallel roex, and it increases uniformly as level increases. In all cases, the ERB is smaller than for the cascade cGC filter. Figure 4(d) shows that the I/O functions of the parallel cGC filter are also more similar to those of the parallel roex system than to those of the cascade cGC system. Moreover, the spread of the functions is even greater for the parallel cGC system. At the lowest center frequency, 0.25 kHz, the parallel cGC system produces a little less gain

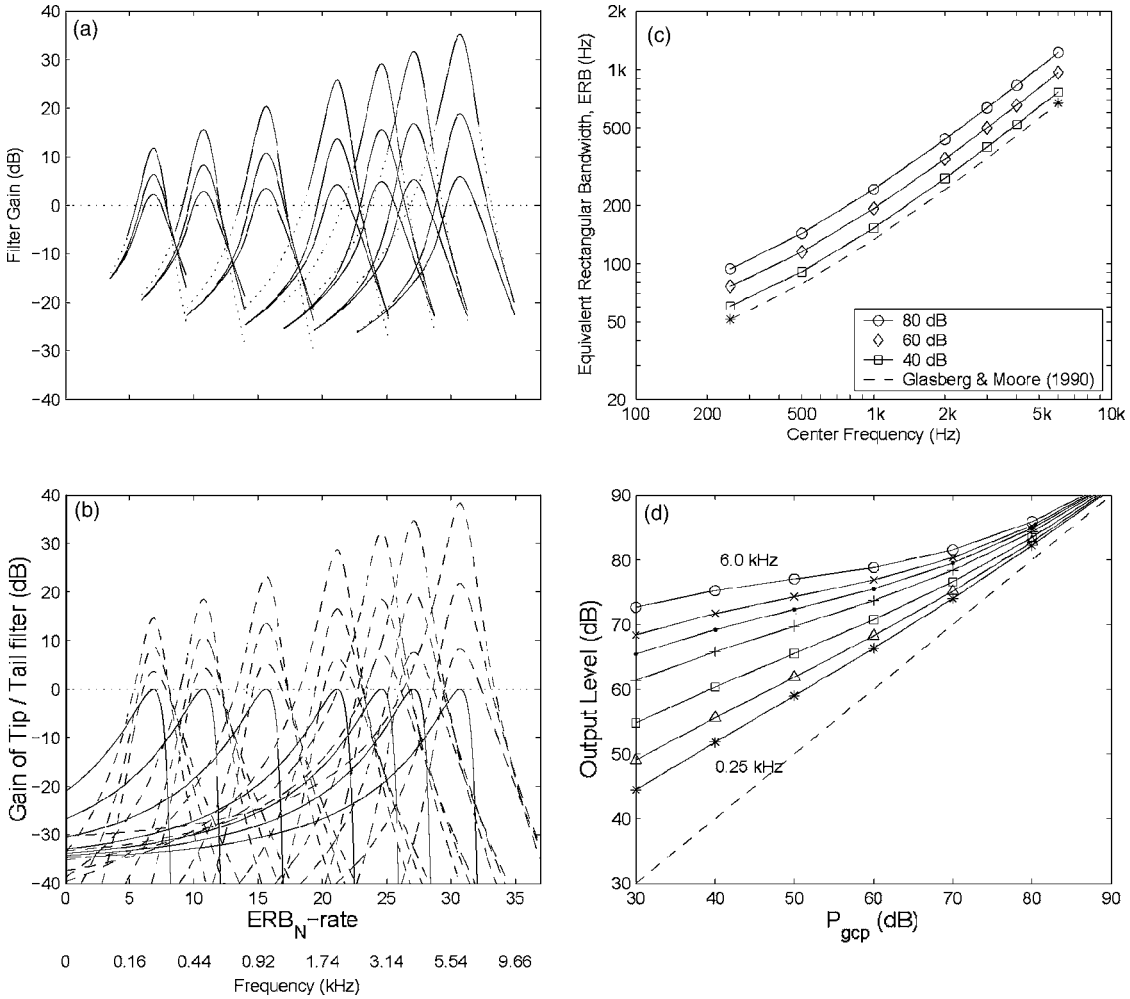


FIG. 4. Characteristics of the parallel cGC filter in the same format as Fig. 2. This version of the parallel cGC filter had eight coefficients (see Table I).

than the parallel roex system, and at the highest center frequency, 6.0 kHz, the parallel cGC system produces a little more gain than the parallel roex system.

There is one way in which the parallel cGC filter system differs from the other two; on the low-frequency side at the lowest level (40 dB), the tail of the composite filter does not converge with the tails of the higher-level filters for frequencies more than two ERB_Ns from the center frequency, which would seem to indicate that the gain is level dependent even for frequencies far removed from the center frequency. In practical terms, this is not a problem because the tails of low-level filters play essentially no role in determining threshold, and hence are not constrained by the threshold data. However, this finding suggests that the "rotation" of the tip filter at low levels may not be the best way to represent cochlear filtering.

4. Common characteristics of the three filter systems

Although the comparisons above highlighted a number of differences between the three filter systems, it should be emphasized that the composite filter systems have much in common: (1) The filters become broader with increasing level at all center frequencies; (2) The filters are asymmetric, with the high side sharper than the low side, and the asymmetry increases with level; (3) The tails of the filters converge on both sides at high stimulus levels, indicating that the gain is effectively level independent for frequencies far removed from the center frequency; (4) The gain of each filter at its peak frequency *increases* as level *decreases*, indicating the presence of relatively strong compression; (5) The amount of compression increases with center frequency, at least at the lower stimulus levels. These would appear to be the characteristics that a filter system must have to explain notched-noise masking. Both parallel and cascade systems can meet the requirements, but it appears that the system with the cascade architecture can produce a slightly better fit with *fewer* coefficients because of its more parsimonious representation of the high-frequency side of the tip filter.

IV. DISCUSSION

A. The trade-off between goodness of fit and number of coefficients

Table I shows the parameter values and goodness of fit values for three versions of each of the filter systems: the cascade cGC, the parallel roex, and the parallel cGC. In each case, the filter system was fitted to the combined data of Baker *et al.* (1998) and Glasberg and Moore (2000). The first and second columns show the type of filter architecture and the total number of filter coefficients, that is, the total exclud-

TABLE II. Three measures of bandwidth for the three types of filter, cascade cGC, parallel roex, and parallel cGC; the values are for the filter centered at 1.0 kHz and they are presented for two stimulus levels, 40 and 80 dB. The three bandwidth measures are the equivalent rectangular bandwidth (ERB), the 10-dB bandwidth, $BW_{10\ dB}$, and BW90, the 90% bandwidth suggested by Kollmeier and Holube (1992).

	ERB		BW90		$BW_{10\ dB}$		
	40 dB	80 dB	40 dB	80 dB	40 dB	80 dB	Average
Cascade cGC filter	196	278	275	416	340	492	333
Parallel roex filter	147	220	252	433	281	451	297
Parallel cGC filter	152	241	229	398	270	440	288
Average	165	246	252	416	297	461	306

ing the coefficients associated with K and P_0. The number of K and P_0 coefficients was the same for all versions of all three filter systems. The filter coefficients are all written as a function of normalized ERB_N-rate, E_f. The last column shows the rms error in decibels.

The upper section of Table I summarizes the results of fitting the cascade cGC filter to the combined data sets. The top row shows a 12-coefficient fit described by Patterson *et al.* (2003, Figs 6–10 and Table I); all five filter parameters were allowed to vary linearly with E_f, and one filter parameter, f_{rat}, was allowed to vary with level; f_{rat} is the center frequency of the pGC filter relative to that of the HP-AF. Patterson *et al.* (2003) noted that the slopes of the regression functions that show how the filter parameters vary with E_f were surprisingly shallow (their Fig. 7), and they showed that the number of regression coefficients could be progressively reduced from 12 to six with little increase in rms error, by successively setting the shallowest slope to 0 and then refitting the data (their Table I). The rms error rose from 3.59 to 3.71 dB. The coefficients for the resulting six-coefficient, cascade cGC filter are presented in the bottom row of the upper panel of Table I; these are the basis of the functions in Fig. 3. One attraction of the six-coefficient fit is that all of the changes associated with center frequency are directly related to the ERB_N function.

The remaining two sections of Table I show that the number of coefficients in the parallel filter systems can also be reduced by restricting the number of regression coefficients to one per filter parameter. In both cases, a reduction from 14 to eight coefficients is accompanied by a very modest increase in rms error. The fully level-dependent versions of the two parallel filter systems have 14 coefficients because both of the tip-filter parameters, p_l and p_u, have to vary with level to produce a good fit. Similarly the restricted versions have eight rather than six coefficients because these same parameters, p_l and p_u, still have to vary with level to produce a good fit. The eight-coefficient, parallel roex filter is the basis of the functions in Fig. 2, and the eight-coefficient, parallel cGC filter is the basis of the functions in Fig. 4. Both of these eight-coefficient filter systems provide excellent fits to the notched-noise data sets. However, if the tip-filter parameters are restricted to be level independent, giving six-coefficient fits, then the rms error rises markedly, as shown in Table I.

This analysis of the trade-off between number of coefficients and goodness of fit for the three filter systems indicates that a cascade system with an HP-AF function that shifts in frequency can explain the way the auditory filter changes with level better than a parallel system with a tip filter that narrows as level increases, and the cascade filter system requires fewer coefficients than either of the parallel filter systems.

B. Filter bandwidth

The width of the auditory filter is typically described in terms of its ERB (Patterson, 1974; Moore and Glasberg, 1983); this is the integral of the entire filter function divided by the filter response at its center frequency. However, Kollmeier and Holube (1992) have argued that, when comparing the bandwidths of highly asymmetric filters, where one skirt becomes much shallower than the other, it is better to use the 10-dB bandwidth, $BW_{10\ dB}$, or a measure they refer to as BW90. BW90 is "the bandwidth that encompasses 90% of the integrated area above, and 90% of the integrated area below, the maximum value of the filter characteristic" (Kollmeier and Holube, 1992). Table II shows the ERB, BW90, and $BW_{10\ dB}$ values for the three filter systems at two stimulus levels, 40 and 80 dB. The center frequency is 1.0 kHz in each case. The functions relating ERB to center frequency, shown in Figs. 2(c), 3(c), and 4(c), all have virtually the same shape, so the relative widths of the three filters will be essentially the same at all center frequencies.

Table II shows an orderly pattern of relationships. The ERB values are consistently smaller than the corresponding BW90 values, which are in turn consistently smaller than the corresponding $BW_{10\ dB}$ values. On average, BW90 is 1.52 times the ERB, and $BW_{10\ dB}$ is 1.84 times the ERB. The bandwidth at 80 dB SPL is considerably larger than at 40 dB; the ratio is 1.49 for the ERB, 1.65 for BW90, and 1.55 for $BW_{10\ dB}$. The filters with parallel architecture (the parallel roex and the parallel cGC) have comparable widths, no matter what the measure, and they are consistently narrower than the cascade cGC filter. The width of the cascade cGC filter is about 1.12 times the width of the parallel roex filter and 1.15 times the width of the parallel cGC filter. So, the bandwidth values provided by the three measures are much as would be expected from their definitions, and the wider values associated with the cascade architecture are not due to the use of the ERB as the bandwidth measure.

As noted in Sec. I C, detailed examination of the fit provided by the roex filter to the data associated with narrow

TABLE III. Slope values for the input/output functions of several auditory filters, for center frequencies from 0.25 to 6.0 kHz. In each case, the slope was calculated over the input range 30–80 dB SPL. (See the text for details.)

Probe frequency (kHz)	0.25	0.50	1.0	2.0	3.0	4.0	6.0
Roex (Table II of Glasberg and Moore, 2000)	0.73	0.70	0.39	0.56	⋯	0.57	⋯
Roex (Table IV of Glasberg and Moore, 2000)	0.51	0.50	0.45	0.44	0.37	0.39	0.36
Cascade cGC filter (six-coefficients) (Patterson *et al.*, 2003)	0.62	0.50	0.42	0.37	0.36	0.35	0.34
Cascade cGC filter (seven-coefficients) (Patterson *et al.*, 2003)	0.69	0.57	0.48	0.40	0.36	0.33	0.29
Parallel roex filter	0.71	0.65	0.58	0.49	0.44	0.40	0.35
Parallel cGC filter	0.75	0.67	0.56	0.45	0.37	0.32	0.25

notches in the masking noise typically reveals a systematic deviation from the data. In particular, the thresholds predicted by the roex filter drop more quickly than the observed thresholds as notch width increases from zero (Patterson *et al.*, 1982; Glasberg *et al.*, 1984b), and more quickly than the unconstrained auditory filter of Patterson (1976). This suggests that the underlying filter shape has a broader tip than the roex filter. Glasberg *et al.* (1984a) compared auditory filter shapes derived using three different types of masker, including notched noise and rippled noise (Houtgast, 1977); it has been argued that the latter may give a more accurate estimate of the shape of the tip of the filter. The tip of the filter derived with rippled noise was broader than that derived with notched noise. Houtgast (1977) also found relatively large filter bandwidths using rippled noise in simultaneous masking. Finally, Oxenham and Dau (2001a) found that they could not account for the masking produced by complex tone maskers with Schroeder-positive and Schroeder-negative phase using a gammatone filter with a relatively sharp tip (similar to that of the roex filter), but they could account for the data with a filter with a broader tip (resembling the cascade cGC). Overall, these results support the idea that filter shapes estimated using the roex filter model have tips that are slightly sharper than the "true" filters. The cascade cGC filter model may provide a more accurate representation of the sharpness of the tip of the auditory filter.

One thing that is clear from Table II is that the tradition of describing the auditory filter in terms of its ERB, or 3-dB, bandwidth *at low stimulus levels* leads to a characterization of auditory frequency selectivity that is not representative of everyday listening. People nowadays routinely listen to music at levels around 80 dB and they often have to listen to speech in noisy environments for which the overall level is rather high. The BW90 measure, which is sensitive to auditory filter asymmetry and the way the asymmetry increases with level, shows that the auditory filter will be relatively wide in listening conditions that are quite common. In these situations, it is unlikely that harmonics on the high-frequency side of speech formants will be resolved (Moore, 1998, page 211).

C. Compression and phase

The input/output function of a filter system reveals the degree of compression that it applies. Table III shows the slopes of the I/O functions for six roex and gammachirp filter systems, over the range of center frequencies from 0.25 to 6.0 kHz (top row). The slope is measured over the range from 30 to 80 dB SPL (assuming that the values of P_{rxp} and P_{gcp} are effectively equivalent to the input level in dB SPL for narrow-band signals centered close to the filter center frequency). The second and third rows show the slope values for the roex filters fitted by Glasberg and Moore (2000) and by Baker *et al.* (1998) to their own data. Recall that their version of the roex filter was incomplete, inasmuch as the upper side was defined by a single roex function. The fourth row shows the slope values for the cascade cGC filter with six coefficients; the I/O functions are presented in Fig. 3(d). The sixth and seventh rows show the slope values for the parallel roex and parallel cGC filters with eight coefficients; the I/O functions are shown in Figs. 2(d) and 4(d), respectively. The degree of compression increases with filter center frequency for all of these filter systems, that is, the slope of the I/O function decreases with increasing center frequency. In the current paper, the data from all of the probe frequencies are fitted simultaneously. The cross-frequency constraint forces the slope of the I/O function to vary monotonically with probe frequency. There was no such constraint in the fits of Glasberg and Moore (2000) (second row). In this case, the slope varies nonmonotonically with frequency; the largest compression occurs at 1.0 kHz where the slope is 0.39.

The parallel roex filter is the most similar to the incomplete roex filters fitted by Baker *et al.* (1998) and Glasberg and Moore (2000) to their data. The parallel roex was fitted to the combined data of Baker *et al.* (1998) and Glasberg and Moore (2000), and so it is not surprising that the slope values (sixth row) are comparable to those reported by Baker *et al.* (third row) and by Glasberg and Moore (second row). The parallel roex values are actually closer to those reported by Baker *et al.* than to those reported by Glasberg and Moore (2000) because the data set of Baker *et al.* has almost three times the number of thresholds as the data set of Glasberg and Moore, and the thresholds were all given equal weight in the parallel roex fit. The slope values for the parallel cGC

filter (seventh row) show a greater range than those for the parallel roex; the slopes for the lower center frequencies (0.25 and 0.5 kHz) are comparable to the largest values reported by Glasberg and Moore (second row), while those at the highest center frequencies (4.0 and 6.0 kHz) are smaller than the smallest values reported by Baker *et al.* (third row). In contrast, the slope values for the cGC filter decrease with increasing frequency only up to 2 kHz, and then remain roughly constant at about 0.35.

Assuming that these slope values are related to the "strength" of the cochlear active mechanism, it is instructive to compare the variation with frequency described above with the variation found by other researchers (Bacon *et al.* 2004), who have used a variety of measures thought to be related to the strength of the active mechanism in the cochlea. Hicks and Bacon (1999) used three measures: (1) the effects of level on frequency selectivity in simultaneous masking, measured using notched-noise maskers at spectrum levels of 30 and 50 dB; (2) two-tone suppression, measured using forward maskers at the signal frequency and suppressor tones above the signal frequency; and (3) growth of masking, measured using forward maskers well below the signal frequency. All three measures revealed a progressive increase in nonlinear behavior as the signal frequency increased from 0.375 to 3.0 kHz, although the change between 1.5 and 3.0 kHz was often rather small. Moore *et al.* (1999) measured the slopes of growth-of-masking functions in forward masking for a masker centered at the signal frequency and a masker centered well below the signal frequency. The ratio of the slopes for the two conditions was taken as a measure of cochlear compression (Oxenham and Plack, 1997). The ratio changed progressively over the range 2.0 to 6.0 kHz. Plack and Oxenham (2000) measured pulsation threshold for a sinusoidal signal alternated with a sinusoidal masker with frequency 0.6 times that of the signal. The slopes of the functions relating pulsation threshold to "masker" level decreased with increasing frequency from 0.25 up to 1.0 kHz, and then remained roughly constant as the signal frequency was increased up to 8.0 kHz. Lopez-Poveda *et al.* (2002) used temporal masking curves (Nelson *et al.*, 2001) to estimate cochlear compression for frequencies from 0.5 to 8.0 kHz. They interpreted their results as indicating that the amount of compression did not vary markedly with center frequency, but at low center frequencies compression extended over a wide range of frequencies relative to the center frequency. Finally, Rosengard *et al.* (2005) estimated cochlear compression using both growth-of-masking functions in forward masking and temporal masking curves. They pointed out that the interpretation of the results depended on what is assumed about the amount of compression that occurs when the masker frequency is well below the frequency/place at which the signal is detected. Depending on the assumption made, the compression could be interpreted as being relatively invariant with frequency over the range 1.0 to 6.0 kHz, or as increasing with increasing frequency. Overall, these results present a mixed picture. They do not allow a clear decision as to which of the filter models evaluated here gives a better representation of the way that cochlear compression varies with frequency.

It should also be noted that both the cascade and parallel filter systems could be modified to allow compression to vary with frequency and level in a different way, should a definitive set of data emerge in the future. Indeed, Patterson *et al.* (2003) report fitting a seven-coefficient version of the cGC filter, in which the compression-range parameter, c_2, was allowed to vary linearly with E_f, to the combined data set of Baker *et al.* (1998) and Glasberg and Moore (2000). The summary of this seven-coefficient fit is presented in the middle row of the top panel in Table I; the compression range increases with E_f, and allowing c_2 to vary reduces the rms error from 3.71 to 3.64 dB, which is just above the value for the 12-coefficient fit, 3.59 dB. The compression slopes for the seven-coefficient version of the cascade cGC filter are presented in the fifth row of the central panel of Table III. With regard to the parallel roex and the parallel cGC filters, the slopes of the I/O functions would probably asymptote in the region above 2.0 kHz if another coefficient were added to the function describing how $G_{\max}$ varies with frequency; this would allow the relationship to be parabolic. Whether this would improve the goodness of fit sufficiently to warrant the extra coefficient is not clear.

In this paper, we have focused on the magnitude response of the auditory filter, since the stimuli in the notched-noise experiments were specified only in terms of their power spectra. However, the phase response of the auditory filter model is also of interest, especially if the filter is to be used for characterizing the auditory analysis of stimuli such as Schroeder-phase complex tones (Kohlrausch and Sander, 1995; Carlyon and Datta, 1997). The phase response at a fixed place on the basilar membrane, corresponding to a single auditory filter, has negative curvature (Robles and Ruggero, 2001). Similarly, psychophysical estimates of the phase response of the human auditory filter suggest that it has negative curvature, and that the normalized (dimensionless) curvature increases with increasing center frequency (Lentz and Leek, 2001; Oxenham and Dau, 2001b). The phase response of the cGC filter has negative curvature of the appropriate form, for frequencies within the passband of the filter, and the normalized curvature increases with center frequency in the appropriate way. However, the curvature does not appear to be great enough to account for the data of Lentz and Leek (2001) or Oxenham and Dau (2001a). This may happen because the phase response of the cGC for frequencies well above the center frequency flattens off, whereas in the physiological data the slope remains negative for frequencies well above the center frequency. For most stimuli, the output of the filter would be dominated by frequency components within the passband, and for these frequencies the phase response of the gammachirp filter appears to be appropriate, at least qualitatively.

D. The nonfilter parameters *K* and P_0

In the fitting process, the relationship between the efficiency constant, K, and the frequency variable, E_f, was found to take the form of a downward-pointing parabola. In other words, detection efficiency was greatest for midrange frequencies. The parabolas for the three filter systems are

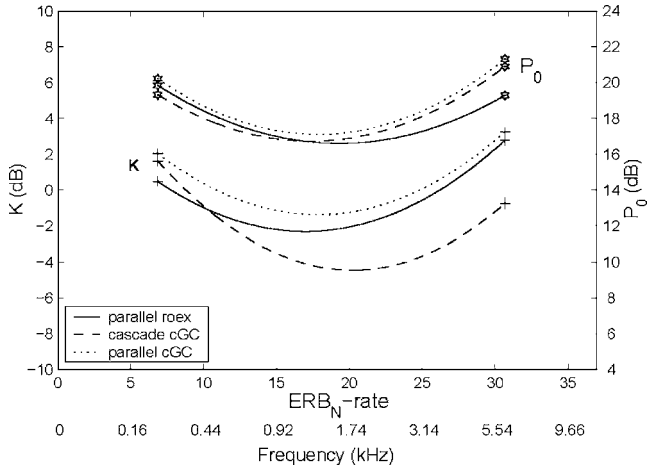


FIG. 5. Values of the nonfilter coefficients, K and P_0, plotted as a function of center frequency on the ERB_N-rate scale for the parallel roex filter (solid line), the cascade cGC filter (dashed line), and the parallel cGC filter (dotted line). The K and P_0 functions are identified by plus and star symbols, respectively, at the ends of the lines.

shown in Fig. 5; the ends of the functions are marked by plus signs. The parabolas for the cascade cGC, the parallel roex, and the parallel cGC filter systems are shown by dashed, solid, and dotted lines, respectively. The K function for the cGC filter falls below those for the parallel roex and parallel cGC filters in the region above 0.5 kHz. This follows directly from the fact that the bandwidth of the cascade cGC is greater than that of the parallel roex filter and the parallel cGC filter [see Figs. 2(c), 3(c), and 4(c)]. A broader filter passes more noise, so to predict the observed thresholds, the value of K has to be lower (Patterson *et al.*, 1982).

The relationship between P_0 and the frequency variable, E_f, was also found to take the form of a parabola, and the parabolas for the three filter systems are shown in Fig. 5 by lines that end in stars. Once again, the parabolas for the cascade cGC, the parallel roex, and the parallel cGC filter systems are shown by dashed, solid, and dotted lines, respectively. This parameter is related to the absolute threshold and it is used to predict the way threshold asymptotes in wide-notch conditions at low stimulus levels. The functions for the three filter systems are very similar and they are also in good agreement with the P_0 estimates reported by Glasberg and Moore (2000) and Baker *et al.* (1998) using the incomplete roex filter.

V. SUMMARY AND CONCLUSIONS

The roex filter has been successfully used for many years to represent the magnitude response of the auditory filter in studies of tone-in-noise masking, and recently the roex(p,w,t) version, with parallel roex(p) and roex(t) filters, has been used to simulate the interaction of the "tip" and "tail" components of the cochlear filter as a function of stimulus intensity (Glasberg and Moore, 2000; Oxenham and Shera, 2003; Baker and Rosen, 2006). Nevertheless, the roex(p,w,t) filter is limited in its application because there is no time-domain version of the filter and no roex(p,w,t) filterbank to simulate basilar-membrane motion in response to complex sounds like speech and music. It is also the case that the fits provided by the complete roex(p,w,t) filter to notched-noise masking data are somewhat unstable, unless one of the high-frequency skirts is removed, in which case the magnitude response is discontinuous and the filter is physically unrealizable.

We have developed a compressive gammachirp (cGC) filter system with parallel architecture like that of the parallel roex filter and compression in the tip filter branch, to provide a filter system that has a similar magnitude response to the parallel roex filter, and which provides a basis for producing a time-domain version of the filter. Basically, we replaced the roex filters with gammachirp filters [compare Figs. 1(a) and 1(c)]. A quantitative comparison of the parallel roex and parallel cGC filters was made by fitting both to the simultaneous notched-noise masking data of Baker *et al.* (1998) and Glasberg and Moore (2000). The behavior of the parallel cGC filter is quite similar to that of the parallel roex filter. In particular: (1) The filters become broader as stimulus level increases at all center frequencies; (2) The filters are asymmetric with the high side sharper than the low side, and the asymmetry increases with increasing level; (3) The tails of the filters converge on both sides at high stimulus levels, indicating that the gain is effectively level independent for frequencies far removed from the center frequency; (4) The gain of the filter at its peak frequency *increases* as stimulus level *decreases*, indicating the presence of relatively strong compression; (5) The amount of compression increases with increasing center frequency, at least in the region below 2.0 kHz. It remains the case, however, that fits to notched-noise data with the parallel cGC filter are still somewhat unstable because the high-frequency skirts of the component filters interact, as with the parallel roex filter.

We also compared the fit of the *cascade* cGC filter (Irino and Patterson, 2001; Patterson *et al.*, 2003) to the notched-noise data, to the fits provided by the parallel roex and parallel cGC filters. The cascade cGC filter provides at least as good a fit as either of the parallel filters to the simultaneous masking data, *with 25% fewer coefficients*, and the fits are stable. The advantage of the cGC filter appears to be a direct result of the architecture of the gammachirp filter with its parsimonious representation of the high-frequency side of the filter. The cascade cGC filter also provides a somewhat more precise fit to the tip of the auditory filter (as determined by the data for small notch widths) than the roex filter does. The effects of level and frequency on the cascade cGC filter system are quite similar to those described above for the parallel filter systems. Moreover, a dynamic version of the cascade cGC filter has been developed that can simulate fast-acting compression (Irino and Patterson, 2005, 2006), so the cGC filter could potentially facilitate future research on the temporal properties of auditory filtering.

In conclusion, it would appear that the compressive gammachirp filter, in either its parallel or cascade form, offers better prospects for the representation of the auditory filter than the roex filter.

ACKNOWLEDGMENTS

This research was supported by a Special Coordination Fund for Promoting the Science and Technology of young

researchers from the Japanese Ministry of Education. The research was also supported by the UK Medical Research Council (G9900369, G9901257, G0500221, G8717539), and a Grant-in-Aid for Science Research [(B)(2) 15300061] from JSPS.

[1]Reimann (2006) argues that the mathematics of the wavelet transform is more appropriate as a representation of cochlear filtering than the Fourier transform.

[2]The parallel architecture of the roex filter would appear to preclude a time-domain equivalent with a level-*in*dependent chirp; if the asymmetry of the tip filter varies with level, the chirp of the impulse response will also vary with level (Carney *et al.*, 1999).

[3]The idea of combining an analytic gammachirp with a high-pass asymmetric function was originally inspired by the NonLinear Resonant Tectorial Membrane (NL-RTM) model of Allen (1997). In this model, two filters interact as a function of level: the first filter has a bandpass characteristic and it represents the basilar-membrane (BM) traveling wave; the second filter has more of a high-pass characteristic and it represents the interaction between the basilar and tectorial membranes (BM-TM filter). The location of the peak of the BM motion moves relative to the BM-TM filter as a function of signal level. The magnitude of mechanical vibration is plotted as a function of cochlear location, as for an excitation pattern. These excitation patterns can be reinterpreted as the magnitude spectra of auditory filters using the conversion scheme proposed by Glasberg and Moore (1990), and in this case, the filter shape is found to be similar to the gammachirp. It still remains, however, to determine whether the conversion is applicable to physiological observations of basilar-membrane motion.

[4]The ERB_N is the average value of the equivalent rectangular bandwidth of the auditory filter for young, normal-hearing listeners at moderate sound levels (Moore, 2003). ERB_N-rate is the function that relates ERB_N to frequency. These are the same definitions as in Glasberg and Moore (1990), with the subscript added to qualify the measure. $ERB_N = 24.7(4.37 f_c + 1)$ and $ERB_N\text{-rate} = 21.4 \log_{10}(4.37 f_c + 1)$.

Allen, J. B. (**1997**). "OHCs shift the excitation pattern via BM tension," in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 167–175.

Allen, J. B., and Sen, D. (**1998**). "A bio-mechanical model of the ear to predict auditory masking," in *Proceedings Computational Hearing* (NATO Advanced Study Institute), Il Ciocco, Italy, 139–162.

Assmann, P. F., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Bacon, S. P., Fay, R. R., and Popper, A. N. (**2004**). *Compression: From Cochlea to Cochlear Implants* (Springer, New York).

Bacon, S. P., Repovsch-Duffey, J. L., and Liu, L. (**2002**). "Effects of signal delay on auditory filter shapes derived from psychophysical tuning curves and notched-noise data obtained in simultaneous masking," J. Acoust. Soc. Am. **112**, 227–237.

Baker, R. J., and Rosen, S. (**2006**). "Auditory filter nonlinearity across frequency using simultaneous notched-noise masking," J. Acoust. Soc. Am. **119**, 454–462.

Baker, R. J., Rosen, S., and Darling, A. M. (**1998**). "An efficient characterisation of human auditory filtering across level and frequency that is also physiologically reasonable," in *Psychophysical and Physiological Advances in Hearing: Proceedings of the 11th International Symposium on Hearing*, edited by A. Palmer, A. Rees, Q. Summerfield, and R. Meddis (Whurr, London), pp. 81–88.

Carlyon, R. P., and Datta, A. J. (**1997**). "Excitation produced by Schroeder phase complexes: Evidence for fast-acting compression in the auditory system," J. Acoust. Soc. Am. **101**, 3636–3647.

Carney, L. H., McDuffy, J. M., and Shekhter, I. (**1999**). "Frequency glides in the impulse responses of auditory-nerve fibers," J. Acoust. Soc. Am. **105**, 2384–2391.

Cohen, M. A., Grossberg, S., and Wyse, L. L. (**1995**). "A spectral network model of pitch perception," J. Acoust. Soc. Am. **98**, 862–879.

Cooke, M. P. (**1993**). *Modelling Auditory Processing and Organization* (Cambridge University Press, Cambridge).

Cooke, M., Green, P., Josifovski, L., and Vizinho, A. (**2001**). "Robust automatic speech recognition with missing and unreliable acoustic data," Speech Commun. **34**, 267–285.

de Boer, E. (**1975**). "Synthetic whole-nerve action potentials for the cat," J. Acoust. Soc. Am. **58**, 1030–1045.

de Boer, E., and Nuttall, A. L. (**2000**). "The mechanical waveform on the basilar membrane. III. Intensity effects," J. Acoust. Soc. Am. **107**, 1497–1507.

Divenyi, P. (Ed.) (**2004**). *Speech Separation by Human and Machines* (Kluwer Academic, Dordrecht, The Netherlands).

Evans, E. F., Pratt, S. R., and Cooper, N. P. (**1989**). "Correspondence between behavioural and physiological frequency selectivity in the guinea pig," Br. J. Audiol. **23**, 151–152.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–61.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Glasberg, B. R., and Moore, B. C. J. (**2000**). "Frequency selectivity as a function of level and frequency measured with uniformly exciting noise," J. Acoust. Soc. Am. **108**, 2318–2328.

Glasberg, B. R., Moore, B. C. J., and Nimmo-Smith, I. (**1984a**). "Comparison of auditory filter shapes derived with three different maskers," J. Acoust. Soc. Am. **75**, 536–544.

Glasberg, B. R., Moore, B. C. J., Patterson, R. D., and Nimmo-Smith, I. (**1984b**). "Dynamic range and asymmetry of the auditory filter," J. Acoust. Soc. Am. **76**, 419–427.

Goldstein, J. L. (**1990**). "Modeling rapid waveform compression on the basilar membrane as multiple-bandpass-nonlinearity filtering," Hear. Res. **49**, 39–60.

Goldstein, J. L. (**1995**). "Relations among compression, suppression, and combination tones in mechanical responses of the basilar membrane: Data and MBPNL model," Hear. Res. **89**, 52–68.

Gunawan, T. S., and Ambikairajah, E. (**2004**). "Speech enhancement using temporal masking and fractional Bark gammatone filters" in *Proceedings Aust. Int. Conference on Speech Science and Technology (SST'04)*, Sydney, 420–425.

Hicks, M. L., and Bacon, S. P. (**1999**). "Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing," J. Acoust. Soc. Am. **105**, 326–338.

Hohmann, V. (**2002**). "Frequency analysis and synthesis using a Gammatone filterbank," Acta. Acust. Acust. **88**, 433–442.

Houtgast, T. (**1977**). "Auditory-filter characteristics derived from direct-masking data and pulsation-threshold data with a rippled-noise masker," J. Acoust. Soc. Am. **62**, 409–415.

Irino, T., and Patterson, R. D. (**1997**). "A time-domain, level-dependent auditory filter: The gammachirp," J. Acoust. Soc. Am. **101**, 412–419.

Irino, T., and Patterson, R. D. (**2001**). "A compressive gammachirp auditory filter for both physiological and psychophysical data," J. Acoust. Soc. Am. **109**, 2008–2022.

Irino, T., and Patterson, R. D. (**2005**). "Explaining two-tone suppression and forward masking data using a compressive gammachirp auditory filterbank," J. Acoust. Soc. Am. **117**, 2598.

Irino, T., and Patterson, R. D. (**2006**). "A dynamic, compressive gammachirp auditory filterbank," IEEE Trans. Audio, Speech, and Language Processing (in press).

Irino, T., Patterson, R. D., and Kawahara, H. (**2006**). "Speech segregation using an auditory vocoder with event-synchronous enhancements," IEEE Trans. Audio, Speech, and Language Processing (in press).

Kohlrausch, A., and Sander, A. (**1995**). "Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets," J. Acoust. Soc. Am. **97**, 1817–1829.

Kollmeier, B., and Holube, I. (**1992**). "Auditory filter bandwidths in binaural and monaural listening configurations," J. Acoust. Soc. Am. **92**, 1889–1901.

Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., and Lütkenhöner, B. (**2003**). "Neuromagnetic evidence for a pitch processing centre in Heschl's gyrus," Cereb. Cortex **13**, 765–772.

Kubin, G., and Kleijn, W. B. (**1999**). "On speech coding in a perceptual domain," in *Proceedings of ICASSP99*, 2327–2330, Phoenix, AZ.

Lentz, J. J., and Leek, M. R. (**2001**). "Psychophysical estimates of cochlear phase response: Masking by harmonic complexes," J. Assoc. Res. Otolaryngol. **2**, 408–422.

Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (**2002**). "Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing," J. Acoust. Soc. Am. **113**, 951–960.

Lutfi, R. A., and Patterson, R. D. (**1984**). "On the growth of masking asymmetry with stimulus intensity," J. Acoust. Soc. Am. **76**, 739–745.

Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (**2001**). "A computational algorithm for computing nonlinear auditory frequency selectivity," J. Acoust. Soc. Am. **109**, 2852–2861.
Moore, B. C. J. (**1998**). *Cochlear Hearing Loss* (Whurr, London).
Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, London).
Moore, B. C. J., and Glasberg, B. R. (**1981**). "Auditory filter shapes derived in simultaneous and forward masking," J. Acoust. Soc. Am. **69**, 1003–1014.
Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.
Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1990**). "Auditory filter shapes at low center frequencies," J. Acoust. Soc. Am. **88**, 132–140.
Moore, B. C. J., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (**1999**). "Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism," J. Acoust. Soc. Am. **106**, 2761–2778.
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.
Oppenheim, A. V., and Schafer, R. V. (**1975**). *Digital Signal Processing* (Prentice-Hall International, London).
Oxenham, A. J., and Dau, T. (**2001a**). "Reconciling frequency selectivity and phase effects in masking," J. Acoust. Soc. Am. **110**, 1525–1538.
Oxenham, A. J., and Dau, T. (**2001b**). "Towards a measure of auditory filter phase response," J. Acoust. Soc. Am. **110**, 3169–3178.
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.
Oxenham, A. J., and Shera, C. A. (**2003**). "Estimates of human cochlear tuning at low levels using forward and simultaneous masking," J. Assoc. Res. Otolaryngol. **4**, 541–554.
Patterson, R. D. (**1974**). "Auditory filter shape," J. Acoust. Soc. Am. **55**, 802–809.
Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640–654.
Patterson, R. D., and Moore, B. C. J. (**1986**). "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London).
Patterson, R. D., and Nimmo-Smith, I. (**1980**). "Off-frequency listening and auditory-filter asymmetry," J. Acoust. Soc. Am. **67**, 229–245.
Patterson, R. D., Allerhand, M., and Giguére, C. (**1995**). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890–1894.
Patterson, R. D., Unoki, M., and Irino, T. (**2003**). "Extending the domain of center frequencies for the compressive gammachirp auditory filter," J. Acoust. Soc. Am. **114**, 1529–1542.
Patterson, R. D., Unoki, M., and Irino, T. (**2005**). "Comparison of the compressive-gammachirp and double-roex auditory filters," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D., Pressnitzer, A., de Cheveigne, S., McAdams, and L., Collet (Springer, New York).
Patterson, R. D., Holdsworth, J., Nimmo-Smith, I., and Rice, P. (**1987**). "SVOS Final Report: The Auditory Filterbank," APU Report 2341.
Patterson, R. D., Nimmo-Smith, I., Wever, D. L., and Milroy, R. (**1982**). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788–1803.
Patterson, R. D., Uppenkamp, S., Johnsrude, I., and Griffiths, T. D. (**2002**). "The processing of temporal pitch and melody information in auditory cortex," Neuron **36**, 767–776.
Patterson, R. D., Robinson, K., Holdsworth, J. W., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). "Complex sounds and auditory images," in *Auditory Physiology and Perception*, edited by Y., Cazals, L., Demany, K., Horner (Pergamon, Oxford), pp. 429–446.
Plack, C. J., and Oxenham, A. J. (**2000**). "Basilar-membrane nonlinearity estimated by pulsation threshold," J. Acoust. Soc. Am. **107**, 501–507.
Press, W. H., Flannery, B. P., Teukolsky, A. A., and Vetterling, W. T. (**1988**). *Numerical Recipies in C* (Cambridge University Press, Cambridge, UK).
Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). "Basilar-membrane responses to clicks at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **103**, 1972–1989.
Reimann, H. M. (**2006**). "Invariance principles for cochlea mechanics: Hearing phases," J. Acoust. Soc. Am. **119**, 997–1004.
Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). "Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases," J. Acoust. Soc. Am. **80**, 1364–1374.
Roman, N., Wang, D. L., and Brown, G. J. (**2003**). "Speech segregation based on sound localization," J. Acoust. Soc. Am. **114**, 2236–2252.
Rosen, S., and Baker, R. J. (**1994**). "Characterizing auditory filter nonlinearity," Hear. Res. **73**, 231–243.
Rosen, S., Baker, R. J., and Darling, A. (**1998**). "Auditory filter nonlinearity at 2 kHz in normal hearing listeners," J. Acoust. Soc. Am. **103**, 2539–2550.
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.
Unoki, M., and Akagi, M. (**1999**). "A method of signal extraction from noisy signal based on auditory scene analysis," Speech Commun. **27**, 261–279.
Unoki, M., and Tan, C. T. (**2005**). "Estimates of auditory filter shape using simultaneous and forward notched-noise masking," in *Proceedings of Forum Acusticum 2005*, Budapest, Hungary, 1497–1502.

Detection and *F*0 discrimination of harmonic complex tones in the presence of competing tones or noise

Christophe Micheyl,[a] Joshua G. W. Bernstein, and Andrew J. Oxenham[b]
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307

(Received 5 October 2005; revised 26 May 2006; accepted 3 June 2006)

Normal-hearing listeners' ability to "hear out" the pitch of a target harmonic complex tone (HCT) was tested with simultaneous HCT or noise maskers, all bandpass-filtered into the same spectral region (1200–3600 Hz). Target-to-masker ratios (TMRs) necessary to discriminate fixed fundamental-frequency (*F*0) differences were measured for target *F*0s between 100 and 400 Hz. At high *F*0s (400 Hz), asynchronous gating of masker and signal, presenting the masker in a different *F*0 range, and reducing the *F*0 rove of the masker, all resulted in improved performance. At the low *F*0s (100 Hz), none of these manipulations improved performance significantly. The findings are generally consistent with the idea that the ability to segregate sounds based on cues such as *F*0 differences and onset/offset asynchronies can be strongly limited by peripheral harmonic resolvability. However, some cases were observed where perceptual segregation appeared possible, even when no peripherally resolved harmonics were present in the mixture of target and masker. A final experiment, comparing TMRs necessary for detection and *F*0 discrimination, showed that *F*0 discrimination of the target was possible with noise maskers at only a few decibels above detection threshold, whereas similar performance with HCT maskers was only possible 15–25 dB above detection threshold. [DOI: 10.1121/1.2221396]

PACS number(s): 43.66.Dc, 43.66.Fe, 43.66.Hg [JHG] Pages: 1493–1505

I. INTRODUCTION

In everyday life, several sound sources are often simultaneously present in the environment. An essential task for the auditory system is to analyze these complex acoustic mixtures in order to detect, identify, and track sounds of interest amid other sounds (Bregman, 1990). For humans and many animal species, the target sounds often fall under the category of harmonic complex tones (HCTs). This encompasses voiced speech sounds and animal vocalizations, the sounds produced by most musical instruments, and many artificially produced alarm signals. Thus, determining how HCTs are "heard out" in the presence of various other types of interfering sounds is an important step toward a better understanding of auditory perception in everyday situations. Besides its theoretical importance, this type of research could have interesting applications in the design of artificial auditory scene analysis systems for automatic speech recognition, automated musical transcription, or perceptual coders for audio compression. Similarly, understanding how the normal auditory system processes HCTs in the presence of other sounds may improve our relatively limited understanding of the listening difficulties experienced by hearing-impaired listeners and cochlear-implant users in environments where multiple sound sources are simultaneously present.

Our understanding of listeners' abilities to hear out and identify HCTs in the presence of competing sounds stems primarily from studies involving either vowel identification or fundamental-frequency (*F*0) identification or discrimination. In so-called double- or concurrent-vowel experiments, two (usually synthetic) vowels are presented simultaneously to a listener, whose task is to identify them. The results of such experiments have demonstrated an important role of differences in fundamental frequency ($\Delta F0$) between the two vowels in promoting correct identification (e.g., Scheffers, 1983; Summerfield and Assmann, 1991; Culling and Darwin, 1993; de Cheveigné *et al.*, 1995). Various models have been proposed to explain this effect (Parsons, 1976; Weintraub, 1987; Stubbs and Summerfield, 1988; Assmann and Summerfield, 1990; Meddis and Hewitt, 1992; de Cheveigné *et al.*, 1995; de Cheveigné, 1997; Cariani, 2001). These models, which operate in the spectral or temporal domain, usually involve as a first stage the estimation of at least one of the two *F*0s present, and as a second stage the use of that *F*0 information to either select or suppress groups of harmonics. An assumption of this class of model is that the improvement in concurrent-vowel identification with increasing $\Delta F0$ is related to *F*0-based perceptual segregation. However, it has been suggested (Culling and Darwin, 1994) that *F*0-based segregation may only work for $\Delta F0$s larger than those at which the improvement in concurrent-vowel identification performance typically plateaus (i.e., about 1 semitone, or 6%). At smaller $\Delta F0$s, identification could be mediated by beats between adjacent harmonics in the experimental steady-state sounds, which may allow "glimpsing" of the individual vowels at different times [Assmann and Summerfield, 1994; Culling and Darwin, 1994; however, see de

[a]Author to whom correspondence should be addressed. Electronic mail: cmicheyl@mit.edu
[b]Present address: Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455. Electronic mail: oxenham@umn.edu

Cheveigné (1999b) for a critique of this interpretation]. Furthermore, it has been shown that listeners can identify above chance even pairs of vowels that have the same *F*0 (e.g., Scheffers, 1983; Zwicker, 1984; Assmann and Summerfield, 1989; Qin and Oxenham, 2005). Thus, good performance in a concurrent-vowel experiment does not necessarily imply *F*0-based perceptual segregation.

Segregation of concurrent HCTs has also been studied using *F*0 identification (Beerends and Houtsma, 1989) and *F*0 discrimination (Carlyon 1996a; 1997). Beerends and Houtsma (1989) found that the ability of listeners to correctly identify the pitches of two simultaneous two-component complexes was strongly impaired only when none of the four components was sufficiently separated in frequency from the others to be individually resolved by the peripheral auditory system. This finding suggests that the ability to exploit $\Delta F0$s in order to hear out concurrent HCTs may be constrained by the interference of spectral components within the auditory periphery, due to limited cochlear frequency resolution. Using rather different techniques, Carlyon (1996a; 1997) also concluded that peripheral resolvability played an important role in determining listeners' ability to hear out one HCT in the presence of another. The first study (Carlyon, 1996a) measured listeners' performance in a sequential *F*0 discrimination task between two consecutive target HCTs in the absence and presence of a simultaneous masker HCT, bandpass-filtered into the same spectral region, with the same *F*0 in the two observation intervals. In the condition where the target and masker both consisted primarily of resolved harmonics, performance was only moderately affected by the masker. In the condition where the target and masker only contained unresolved harmonics, listeners were still able to perform the task, but reported not being able to hear two sounds; instead, they heard the target and masker as a single "crackly" sound. Carlyon concluded that performance in that condition was probably based on the discrimination of global changes in the pitch evoked by the target-plus-masker mixture, rather than the pitch of the target alone. Specifically, the rate of envelope peaks of the combined masker and target increased with increases in the target *F*0, giving a so-called "mean-rate cue." This conclusion was supported by a later study (Carlyon, 1997), which showed chance performance in a similar task when the mean-rate cue was neutralized by using pseudorandom pulse trains generated in such a way that, when mixed together, they contained the same number of pulses.

The present study was designed to gain additional information regarding the perceptual segregation of concurrent harmonic complexes occupying overlapping spectral regions, and the role of peripheral frequency resolution in this ability. Similar to Carlyon (1996a, 1997), we used an *F*0-discrimination paradigm, with two important differences. First, instead of using a fixed target-to-masker ratio (TMR) of 0 dB [as in Carlyon (1996a)], we measured the TMR necessary to detect a fixed $\Delta F0$ between the targets. A fixed $\Delta F0$, which was suprathreshold in the absence of the masker, provided us with an operational definition for listeners' ability to "hear out" the target and enabled us to test this ability over a range of TMRs. Second, the *F*0 of the masker was roved across intervals, to discourage listeners from using any composite pitch of the target and masker, deriving from a "mean-rate" cue. Varying degrees of peripheral resolvability of the harmonics were obtained by using *F*0s ranging from below 100 Hz to above 400 Hz and bandpass filtering the stimuli between 1200 and 3600 Hz. For example, with the 100-Hz target *F*0, the spectral components of the target were unresolved from each other, and thus remained unresolved after the addition of the masker; with the 200-Hz target, some components were resolved but may have been unresolved after being added to the masker (depending on the TMR and masker *F*0); at the 400-Hz *F*0, the target components were well resolved from each other and some of them could still remain resolved after the masker was added. Finally, in addition to using harmonic maskers, we also used noise maskers. Besides the fact that noise constitutes another common source of interference in real-life environments, the results of the noise-masker conditions provide control or reference data, against which to compare the effects of HCT maskers.

A preliminary experiment measured *F*0 discrimination thresholds for isolated HCTs. The main experiment (experiment 2) measured the TMR required to discriminate the *F*0 of a HCT in the presence of harmonic or noise maskers. Experiment 3 repeated experiment 2, but without roving the masker *F*0 across the two intervals of each trial, to investigate the possible roles of masker distraction and of composite (signal-plus-masker) percepts when they provide reliable information. Experiment 4 investigated the effect of introducing additional segregation cues between the masker and target, based on onset and offset asynchronies. The final experiment measured detection thresholds for the target HCTs in both harmonic and noise maskers, in order to determine the relationship between the TMR necessary for the detection of a target and the TMR necessary for pitch discrimination judgments.

II. GENERAL METHODS

A. Stimuli

In all experiments, the target stimuli were 500-ms (total duration) HCTs gated on and off with 20-ms raised-cosine ramps. They were bandpass filtered between 1200 and 3600 Hz using digital trapezoidal filters (on a log frequency scale) with slopes of 48 dB/oct. The filtering was done in the spectral domain by individually adjusting the amplitude of each sinusoidal component and then summing the components together. Depending on the experiment, the target HCTs were either presented alone (experiment 1) or accompanied by simultaneous maskers (experiments 2–5). When present, the maskers were usually 500 ms in duration (including 20-ms raised-cosine ramps) and were usually gated on and off together with the targets (with the exception of experiment 4). The maskers were either HCTs or Gaussian noise, bandpass filtered in the same way as the targets. When the masker was a HCT, depending on the condition being tested, its nominal (i.e., average) *F*0 was either equal to (*C*0 condition), 7 semitones lower than ($C-7$ condition), or 7 semitones higher than ($C+7$ condition) the nominal target

*F*0. In most experiments, the masker *F*0 was roved across trials and intervals. The details of the roving are provided within the descriptions of each experiment.

The target HCTs had nominal *F*0s of 100, 200, or 400 Hz; actual target *F*0s were roved across trials over a 6-semitone range (±3 semitones), independent of the masker *F*0. Within the selected passband (1200–3600 Hz), these nominal *F*0s result in different degrees of peripheral resolvability of the harmonics. While the exact limit between resolved and unresolved harmonics can vary somewhat depending on how the measurements are made and how resolvability is defined, it is generally believed that the first six to ten harmonics are peripherally resolved (Plomp, 1964; Bernstein and Oxenham, 2003). With the stimulus passband used here (1200–3600 Hz), target HCTs having a nominal *F*0 of 100 Hz contained only harmonics higher than the 10th, i.e., all peripherally unresolved; target HCTs with a nominal *F*0 of 200 Hz contained harmonics between the 5th and the 21st, thus including some resolved and some unresolved components; and target HCTs with a nominal *F*0 of 400 Hz contained only harmonics less than about the 10th, i.e., all resolved. Of course, this analysis only holds for each HCT in isolation; the addition of a masker HCT reduced the degree to which individual harmonics were resolved in ways that depended on the TMR as well as on the specific masker and target *F*0s; we will return to this question when discussing the results.

The starting phases of the target and masker components were drawn randomly and independently from a uniform distribution spanning 0°–360° on each presentation. This was done in order to avoid providing listeners with consistent cues associated with a specific choice of phase relationship for the stimulus components (e.g., de Cheveigné, 1999b).

All stimuli were presented in lowpass-filtered (1200-Hz cutoff) pink noise. The purpose of this background noise was to prevent listeners from detecting distortion products generated at lower harmonic frequencies, which could have confounded the interpretation of the results. To ensure that distortion products would always be masked, the 1/3rd-octave band level of the lowpass noise was adjusted on each trial to be equal to the highest level per component of the target or masker. The background noise was turned on 500 ms before the onset of the first target on a trial and off 500 ms after the offset of the second target on the same trial. It had a total duration of 2.5 s, including 20-ms raised-cosine ramps.

B. Procedure

In all experiments, thresholds were measured using a two-interval, two-alternative forced-choice (2I-2AFC) procedure with an adaptive three-down one-up rule, which tracked the 79.4%-correct point on the psychometric function (Levitt, 1971). Each trial consisted of two 500-ms observation intervals separated by 500 ms. The intervals were marked by "lights" on a virtual response box, displayed on a computer screen. In all experiments except the last, the two intervals contained target HCTs that differed in *F*0 by an amount, $\Delta F0$, which was either varied adaptively (experiment 1) or kept constant (experiments 2–4) within each experimental run. The interval containing the target with the higher *F*0 was selected randomly with a 0.5 probability, and the listener's task was to indicate whether the higher-*F*0 target occurred in the first or second interval. Responses were entered via a computer keyboard (key "1" for interval 1, key "2" for interval 2). Visual feedback was provided following each trial. The procedure and task used in the final experiment are described separately under Sec. VII.

When a masker was present, its overall [root-mean-square (rms)] level was either kept fixed at 56 dB SPL (experiments 2–4) or roved across observation intervals between 51 and 61 dB SPL (experiment 5). The rms level of the target was always set relative to the (actual) rms level of the masker. In all experiments except the first, the TMR was the variable in the adaptive tracking procedure. After three consecutive correct responses, the TMR was decreased; after each incorrect response, it was increased. At the beginning of a block of trials, the step size by which the level of the target was increased or decreased was set to 6 dB; it was reduced to 4 dB after the first reversal in the direction of tracking from increasing to decreasing, and to 2 dB after the second such reversal. Thereafter, the step size was kept fixed at 2 dB. The procedure stopped after a total of six reversals with the 2-dB step size. The threshold was defined as the average TMR at the last six reversal points. Each condition was repeated at least three times for each listener.

C. Apparatus

The stimuli were generated digitally and played out via a soundcard (LynxStudio LynxOne) with 24-bit resolution and a sampling frequency of 32 kHz. The stimuli were then passed to a headphone buffer (TDT HB6) before being presented to the listener via the left earpiece of Sennheiser HD 580 headphones. Subjects were seated in a double-walled sound-attenuating chamber.

D. Listeners

Four listeners (one female, ages between 22 and 28 yr) took part in the study. They all had pure-tone hearing thresholds less than 20 dB HL at octave frequencies between 250 and 8000 Hz. All had some musical education and had several years of experience playing an instrument.

III. EXPERIMENT 1. *F*0 DISCRIMINATION THRESHOLDS

A. Rationale

The aim of this preliminary experiment was to measure *F*0 difference limens (DL*F*0s) for the target HCTs at each of the three nominal target *F*0s tested in the absence of any interference. This enabled the individual adjustment of the (fixed) $\Delta F0$ between the two targets in the other experiments, such that it corresponded to a constant proportion of each listener's DL*F*0.

B. Methods

DL*F*0s for target complexes with a nominal *F*0 of 100, 200, or 400 Hz were measured using a 2I-2AFC paradigm

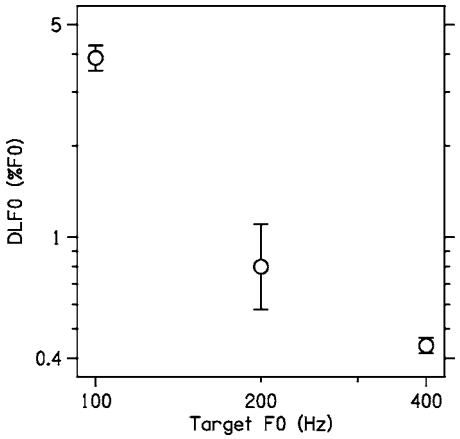


FIG. 1. Mean $F0$ discrimination thresholds or DL$F0$s for HCTs with nominal $F0$s of 100, 200, and 400 Hz (experiment 1). Thresholds are expressed in percent of the nominal $F0$, and plotted on a logarithmic scale. The error bars represent standard errors of the mean across listeners.

with an adaptive three-down one-up procedure. At the start of each trial block, the $\Delta F0$ between the two targets was set to 20% of the actual reference $F0$, which was drawn randomly on each trial from a 6-semitone range centered on the nominal $F0$ as explained in Sec. II. The $\Delta F0$ was then adaptively varied by a factor of 4 for the first two reversals, 2 for the next two reversals, and $\sqrt{2}$ for the last six reversals. The DL$F0$ was computed as the geometric mean of $\Delta F0$ at the last six reversal points. Each listener completed at least three such runs at each nominal target $F0$. The resulting DL$F0$s were geometrically averaged to produce a single DL$F0$ estimate per listener per condition. The overall level of the target was kept constant at 56 dB SPL. As in all other experiments, the background lowpass noise was present to mask distortion products but, unlike in subsequent experiments, there was no masker in the same spectral region as the target.

C. Results

The mean DL$F0$s, averaged (geometrically) across listeners are shown in Fig. 1. They were largest at the 100-Hz $F0$ (around 4%), and smallest at the 400-Hz $F0$ (around 0.45%). This is consistent with data from other studies using comparable stimulus conditions. In particular, the finding of substantially larger DL$F0$s in the 100-Hz $F0$ condition (for which the passband of the target HCT contained only unresolved harmonics) than in the 200- and 400-Hz $F0$ conditions (for which the target contained resolved harmonics), is consistent with earlier results (e.g., Hoekstra, 1979; Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003; 2005).

IV. EXPERIMENT 2. *F0* DISCRIMINATION OF HARMONIC TARGETS IN SIMULTANEOUS HARMONIC OR NOISE MASKERS

A. Methods

The target stimuli were similar to those used in experiment 1. However, in contrast to experiment 1, the $\Delta F0$ between the two consecutive targets was kept constant throughout each run. Depending on the condition, the $\Delta F0$ for each listener was set to 2 or 4 times the DL$F0$ in quiet, as measured individually in experiment 1. The target was always accompanied by a HCT or noise masker, filtered into the same spectral region, and the TMR was varied adaptively. The overall masker level was fixed at 56 dB SPL. The masker $F0$ was roved over a 6-semitone range across trials. Further roving was introduced between the two intervals within each trial, to discourage listeners from basing their judgments on any global percepts evoked by the target-plus-masker mixtures. The within-trial roving was set to 3 times the $\Delta F0$ between the two targets, corresponding to 6 or 12 times each listener's DL$F0$.

B. Results and discussion

Figure 2 shows the threshold TMRs averaged across all four listeners for the different testing conditions. These data were analyzed using a repeated-measures analysis of variance with target $F0$ (100, 200, or 400 Hz), masker type ($C-7$, $C0$, $C+7$, or noise), and $\Delta F0$ (2 or 4 times the listener's DL$F0$) as within-listener factors. A general finding, which was observed in this and all subsequent experiments of the same type, is that TMRs were lower at the larger $\Delta F0$ ($4\times$DL$F0$) than at the smaller $\Delta F0$ ($2\times$DL$F0$) [$F(1,3)=41.31$, $p=0.008$]. This can be explained rather simply by considering that increasing the $\Delta F0$ between the two targets made it easier for listeners to detect which was the higher in pitch, so that listeners could tolerate lower TMRs while still

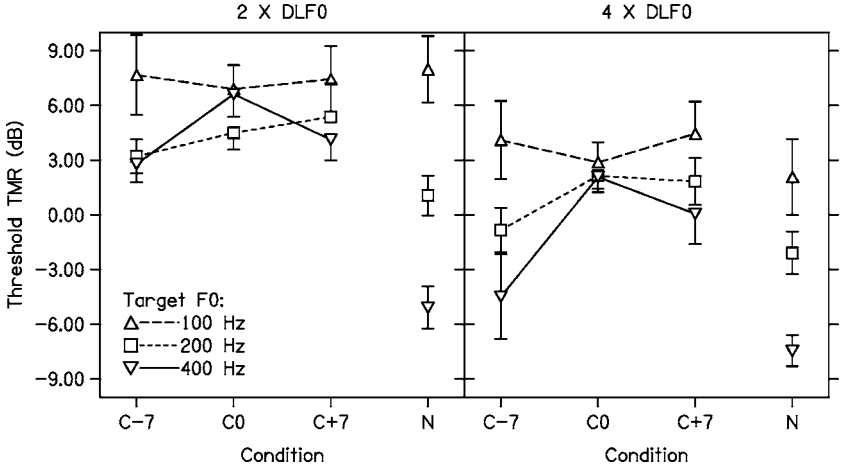


FIG. 2. TMRs required for 79.4%-correct target $F0$ discrimination in the presence of harmonic or noise maskers (experiment 2). Left and right panels show data using $\Delta F0$s between the targets in each trial corresponding to two and four times each listener's DL$F0$, respectively. In this and subsequent figures, the masker type is indicated under the horizontal axis: "*N*" stands for noise and "*C*-7," "*C*0," and "+*C*7" stand for a complex tone where the $F0$ range of the masker was centered 7 semitones below, equal to, or 7 semitones above that of the target, respectively. Symbols represent target complex nominal $F0$s: downward-pointing triangles for 100 Hz, circles for 200 Hz, and upward-pointing triangles for 400 Hz. In this and subsequent figures, the error bars represent standard errors of the mean across listeners. In cases of overlap, error bars on one side of the symbol are omitted for clarity.

achieving about 79% correct responses in the discrimination task. The same general trends were usually observed at both values of $\Delta F0$ in all experiments, so that there were no significant interactions between this and the other experimental factors. The effects of target $F0$ and masker type are addressed below.

1. Effect of average target F0

Overall, the threshold TMRs improved significantly as the target's nominal $F0$ increased from 100 to 400 Hz [main effect: $F(1.20,3.60)=10.55$, $p=0.035$; linear contrast: $F(1,3)=1.887$, $p=0.041$].[1] This effect was particularly large in the noise-masker conditions (plotted on the right in each panel of Fig. 2), where the mean threshold TMR decreased by more than 10 dB as the nominal target $F0$ increased from 100 to 400 Hz [$F(1.40,4.21)=20.04$, $p=0.0008$]. With harmonic maskers the reduction in threshold TMR with increasing target $F0$ was generally not as dramatic, and was only present when the average masker $F0$ was different from that of the target $F0$ [$F(1.31,3.92)=8.54$, $p=0.018$].

The reason why TMRs in noise vary so dramatically with target $F0$ is not completely clear, although it is consistent with earlier studies (Bilsen, 1973; Hoekstra, 1979). Bilsen (1973) showed that a 6% change in the $F0$ of a two- or three-component HCT was more easily masked by broadband noise for high than for low harmonic numbers. Hoekstra (1979, pp. 42–43) measured rate discrimination thresholds for bandpass-filtered periodic pulse trains in a noise masker as a function of TMR, with pulse rate (or $F0$) as a parameter. His results indicate that as TMR decreases, thresholds increase much more dramatically for low-$F0$ than for high-$F0$ pulse trains in a given spectral region. Envelope pitch information carried by unresolved harmonics may be more susceptible to noise than the spectral or fine-structure information carried by resolved harmonics, and this might explain the very high thresholds in the 100-Hz $F0$ condition. For resolved harmonics, such as those in the 200- and 400-Hz $F0$ conditions, if discrimination could be based on the frequencies of individual harmonics, then the improvement with increasing $F0$ might be due to the higher level per component in the higher $F0$ conditions: every doubling in $F0$ leads to a halving of the number of components within the passband and, for a constant overall level, a concomitant doubling (3-dB increase) in the intensity of each component. The extent to which these different explanations can account for the results remains to be explored.

In the presence of a HCT masker whose $F0$ was different from that of the target, thresholds decreased with increasing $F0$. This is probably due to the concomitant increase in average spacing between adjacent spectral components of the composite stimulus. This presumably resulted in greater peripheral separation of the target and masker components on average, so that the likelihood of there being some peripheral auditory filters with a relatively large TMR at the output increased. In this case, therefore, the improvement in TMR with increasing masker and target $F0$ may relate to an increase in the audibility of the target. The question of target audibility is addressed further in experiment 5.

2. Effect of masker type

The effect of masker type was dependent on the target $F0$ range. For the lowest average target $F0$ (100 Hz), there was no effect of masker type: TMRs at threshold were approximately the same whether the masker was a noise or a HCT, and were roughly constant across the three different relative masker $F0$ ranges. At the 200-Hz target $F0$, some emerging trends for TMRs to vary with masker type and masker $F0$ were apparent; harmonic maskers appeared on average to produce higher threshold TMRs than noise maskers, and harmonic maskers with an $F0$ below that of the target appeared to produce lower threshold TMRs than maskers with an $F0$ equal to, or a higher than, the target $F0$. However, these trends failed to reach significance. In contrast, for the highest average target $F0$ (400 Hz), mean TMRs at threshold were strongly dependent on masker type, varying over a 14-dB range. Threshold TMRs for the 400-Hz targets were generally lower in noise than in HCT maskers [$F(1,3)=284.77, p<0.001$]. This may reflect a pitch-interference effect, whereby the pitch of the masker interferes with the ability of listeners to discriminate the pitch of the target (Gockel *et al.*, 2004). Consistent with this interpretation, threshold TMRs were highest when the average target and masker $F0$s were the same, possibly reflecting the difficulty in distinguishing the masker from the target when they are both drawn from the same range of $F0$s.

3. Implications of positive and negative TMRs with HCT maskers

Threshold TMRs of 0 dB or less imply that the pitch of both the target and HCT masker may have been heard by the listeners. This is because the masker level was either the same as or higher than the target level and therefore was likely to have had at least as clear a pitch as the target (assuming that the pitch salience of the masker and target are roughly equal when played in isolation). The threshold TMR was negative for the 400-Hz target $F0$ in the $4\times$DL$F0$ condition when the masker and target $F0$s were different, and approached zero for one of the 200-Hz target $F0$ conditions.

Positive TMRs are less straightforward to interpret. It may be that the target essentially masks the pitch of the masker at high TMRs. This seems particularly likely for the 100-Hz $F0$ target, where threshold TMRs were rarely below +3 dB. If the masker $F0$ was not perceived in the 100-Hz $F0$ target conditions, this would explain why changes in the masker $F0$ range had no effect, and why the TMRs were roughly the same for both noise and HCT maskers. The inaudibility of the masker $F0$ for the 100-Hz $F0$ target condition would be consistent with Carlyon's (1996a) observation that when listeners were presented with a target and masker containing in their passband only unresolved harmonics having approximately the same level, listeners could not perceptually separate the two, and heard the mixture as a fused sound.

Because our technique involved an adaptive tracking threshold procedure, and always started with a high TMR, it is also possible that a positive TMR threshold reflects the minimum level difference at which it was possible to distin-

guish the target and masker based on loudness differences. In other words, the psychometric function may be nonmonotonic, with a minimum reached when the target and masker are at similar levels. In that case, it is possible that performance might again improve at negative TMRs, when listeners could focus on the quieter of the two sounds present (for an example from the speech perception literature, see Brungart, 2001). However, the most likely situation for a level difference to be effective would be in cases where the *F*0 range of the masker and target are the same. In the other cases, it appears unlikely that a level difference would provide a more salient cue than the large pitch-range differences already present between the masker and target.

In summary, the positive TMRs observed in most of the cases tested in experiment 2 suggest that it is usually not possible to hear out the pitch of a target in the presence of a synchronously gated, higher-level harmonic masker in the same spectral region, even when substantial target-masker *F*0 differences exist and when the target and masker themselves contain mostly resolved harmonics when presented in isolation.

V. EXPERIMENT 3. INFLUENCE OF WITHIN-TRIAL RANDOMIZATION OF THE MASKER *F*0

A. Rationale

In the previous experiment, the masker *F*0 was randomized between the two observation intervals on each trial. This was done to encourage listeners to base their judgments on the *F*0 of the target, rather than on some overall sensation derived from the mixture of the masker plus the target, such as Carlyon's (1996a) mean-rate cue. Although this approach helps to ensure that responses were based on the pitch of the target alone, it has the disadvantage that performance might be limited by the potentially distracting variations in the masker *F*0. In other words, performance in experiment 2 may have been limited by confusions between the target and masker, even when the listeners were able to perceptually separate the two simultaneous sounds. The observed tendency for TMRs to improve with increased target-masker *F*0 differences is consistent with this idea.

In this experiment, we kept the masker *F*0 constant across the two observation intervals to test two hypotheses related to the observed threshold TMR measurements of experiment 2. The first hypothesis was that the threshold TMRs were limited by masker distraction or target-masker confusion in the 400-Hz (and possibly 200-Hz) target *F*0 conditions, where the threshold TMRs measured in experiment 2 were usually the lowest, indicating that the masker *F*0 may have been heard and was sometimes louder than the target. If so, then a constant masker *F*0 should lead to an improvement (a decrease) in the threshold TMRs by decreasing the possible influence of masker distraction or target-masker confusions. The second hypothesis was that mean-rate cues could play a role in determining thresholds with complex consisting of only unresolved harmonics (Carlyon, 1996a; 1997). If so, then introducing a useful mean-rate cue by keeping the masker *F*0 constant should lead to an improvement in performance with the 100-Hz *F*0 target.

B. Methods

The stimuli and procedure for this experiment were the same as those for experiment 2, except that the *F*0 of the masker was the same in both intervals of a given trial. The masker *F*0 was still roved across trials, as in the previous experiment. Four threshold estimates were obtained in each condition for each listener. The mean threshold TMRs shown below were computed as the mean of these threshold estimates across repetitions and listeners.

C. Results and discussion

The average threshold TMRs for this experiment are shown in the top two panels of Fig. 3. The results displayed the same general trends as those of experiment 2. In particular, thresholds improved with increasing target *F*0 [$F(1.77,5.29)=28.66, p=0.002$], and at the highest nominal target *F*0 tested (400 Hz), there was some advantage to having a difference in average *F*0 between target and masker [$F(1,3)=16.20, p=0.028$]. To facilitate comparisons between these results and those of experiment 2, threshold TMR differences (obtained by subtracting the thresholds measured in experiment 2 from those measured in corresponding conditions of the current experiment) are plotted in the lower two panels of Fig. 3. Negative values indicate an improvement in TMRs with the elimination of the within-trial rove of the masker *F*0.

The comparison reveals that keeping the masker *F*0 constant across observation intervals resulted in substantial improvements in thresholds at 400 Hz [$F(10,3)=296.09, p<0.001$], but not at 100 or 200 Hz ($p>0.05$). This outcome is not predicted under the hypothesis that the lack of within-trial randomization of the masker *F*0 would allow listeners to take advantage of Carlyon's (1996a) mean-rate cue, because it is an envelope-based cue that should have an effect mainly at low *F*0s, where the harmonics of the target were not peripherally resolved.[2] It may be that our use of random-rather than constant sine- or cosine-phase harmonics, resulting on average in a weaker representation of the fluctuations in the stimulus temporal envelope, reduced the usefulness of the mean-rate cue. Instead, the outcome is consistent with our earlier observation that the threshold TMRs measured at the 100-Hz *F*0 in experiment 2 were usually above 0 dB, indicating that the masker was less intense than the target. Under such conditions, the target was presumably more salient than the masker, making it perhaps less necessary for the listeners to actively ignore the irrelevant changes in the masker *F*0. These results suggest that the masker *F*0 was playing little or no role for both the 100- and 200-Hz target *F*0 conditions, making it even less likely that the positive TMRs measured in experiment 2 reflect segregation based on simple loudness differences and nonmonotic psychometric functions as discussed in Sec. IV B 3, or target-masker confusions. Instead, they confirm our interpretation that, in that experiment, the 100- and 200-Hz target pitch could not usually be heard out when the target was less intense than the masker.

In contrast, in the 400-Hz nominal target *F*0 conditions, the threshold TMRs were usually lower, and sometimes negative, making it necessary for listeners to actively ignore

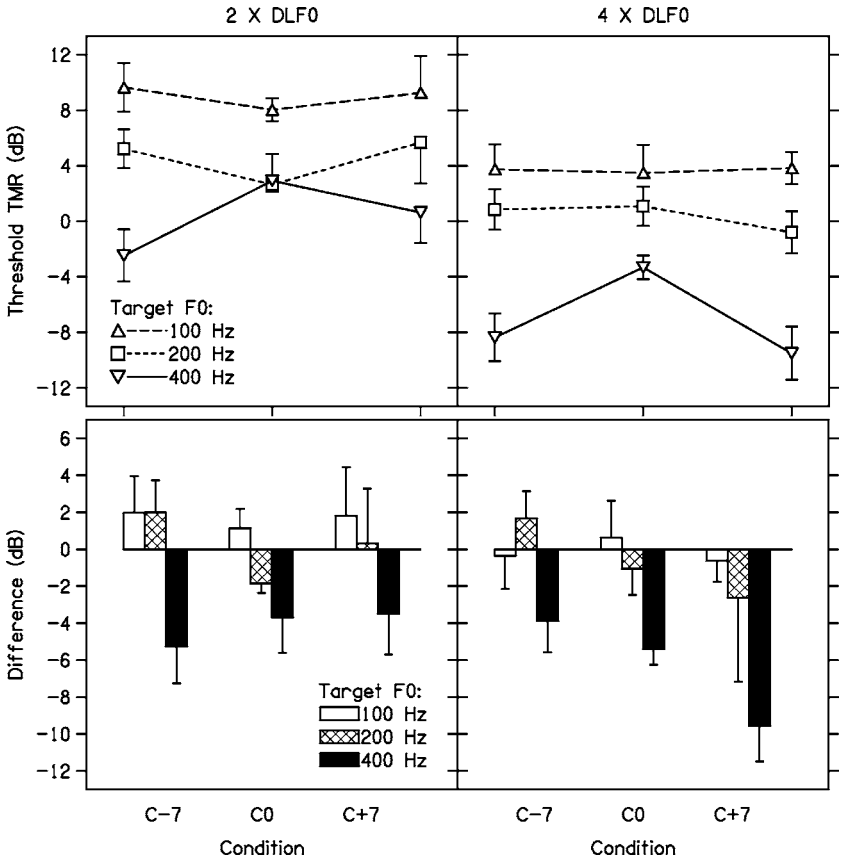


FIG. 3. Upper panels: Threshold TMRs required for 79.4%-correct *F*0 discrimination of harmonic targets in the presence of harmonic maskers having the same *F*0 in the two observation intervals within a trial. Lower panels: Differences in TMRs between experiment 3 (masker *F*0 roved across intervals) and experiment 2 (masker *F*0 held constant across intervals). Negative values indicate an improvement in TMR with the elimination of across-interval *F*0 roving in experiment 3. Left and right panels show data using Δ*F*0s between the targets in each trial corresponding to two and four times each listener's DL*F*0 in quiet, respectively.

the irrelevant variations in masker *F*0. From that point of view, the finding that removing the within-trial variation in masker *F*0 improved thresholds at the 400-Hz *F*0 suggests that, at that nominal target *F*0, performance in experiment 2 was limited by target-masker confusions and an inability to ignore the irrelevant variation in the pitch of the masker, even though listeners may have heard the masker and target as two separate objects.

VI. EXPERIMENT 4: INFLUENCE OF ONSET-OFFSET ASYNCHRONIES

A. Rationale

Onset and offset asynchronies are usually regarded as a powerful cue for concurrent sound segregation. Components that start and end at the same time tend to be grouped together by the auditory system, while components that start or end at different times tend to be perceived as separate objects. There are many illustrations of this in the psychoacoustical literature (for reviews, see Bregman, 1990; Darwin and Carlyon, 1995). An important question is whether onset-offset asynchronies can help listeners to segregate concurrent HCTs.

Some evidence that they do can be found in two earlier studies in which listeners had to identify a synthetic vowel masked either by another vowel (Akeroyd and Summerfield, 2000) or by a flat-spectrum HCT (Demany and Semal, 1990). In the asynchronous condition, the masker started approximately 500 ms before and ended either simultaneously with (Demany and Semal, 1990) or 100 ms after (Akeroyd and Summerfield, 2000) the target. The asynchrony was found to have a beneficial effect on identification accuracy (Akeroyd and Summerfield, 2000) or masked thresholds (Demany and Semal, 1990).

On the other hand, Carlyon (1996a, 1996b) found no benefit of target-masker asynchronies in experiments which, like the present ones, required listeners to discriminate the *F*0 of two consecutive target HCTs in the presence of another HCT. In fact, when the target contained only unresolved harmonics in its passband, an asynchronous masker not only failed to help listeners, but even had a detrimental influence on performance. Carlyon (1996b) interpreted this as a sequential interference effect in pitch discrimination, whereby the leading (or trailing) portion of the masker had a negative influence on the processing of the *F*0 of the subsequent (or preceding) target.

The present experiment further tested the influence of onset-offset asynchronies on the segregation of concurrent HCTs but with two important methodological differences from the Carlyon (1996a, 1996b) studies. First, measuring threshold TMR enabled us to investigate the hypothesis that asynchronies could provide a useful segregation cue for unresolved HCTs for sufficiently large TMRs where the target *F*0 can be heard out. Second, this experiment used a target-masker asynchrony of 500 ms (instead of the 150-ms asynchrony in the Carlyon studies), testing the hypothesis that the lack of a benefit observed by Carlyon was due to an insufficiently long onset cue, as might be predicted from the results of Darwin and Ciocca (1992), who found that for one component to cease contributing to the overall pitch of a complex, the asynchrony had to exceed 300 ms.

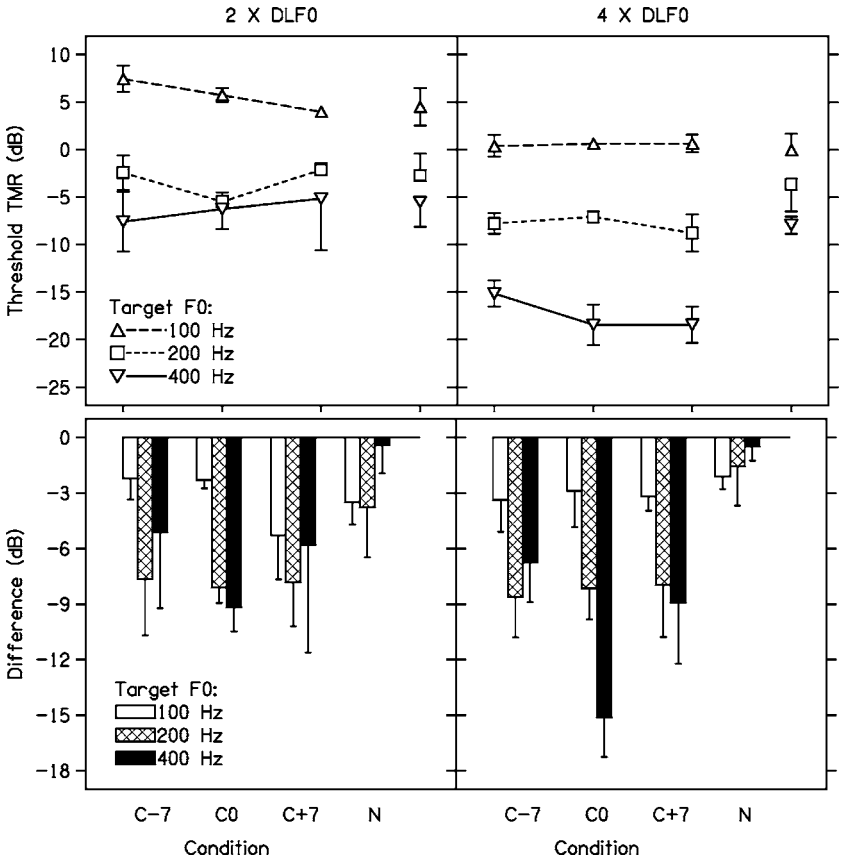


FIG. 4. Upper panels: TMRs corresponding to 79.4%-correct *F*0 discrimination of harmonic targets in the presence of asynchronous maskers (experiment 4). Lower panels: Differences in TMRs between experiment 4 (asynchronous maskers) and those measured in experiment 3 (synchronous harmonic maskers) or 2 (synchronous noise maskers). The average differences shown in these panels were computed by subtracting the synchronous TMRs measured in experiments 2 or 3 from the asynchronous TMRs measured in experiment 4. Negative TMRs indicate the benefit provided by target-masker asynchronies. Left and right panels show data using $\Delta F0$s between the targets in each trial corresponding to two and four times each listener's DL*F*0 in quiet, respectively.

B. Methods

This experiment was similar to experiment 3, except that the masker was turned on 500 ms before the onset of the first target and it continued uninterrupted until 500 ms after the offset of the second target. Thus, the masker duration was 2.5 s. The masker *F*0 remained constant throughout the 2.5 s but, as in the previous two experiments, it was roved over 6 semitones across trials. Because the masker *F*0 remained constant within each trial, the threshold TMRs were compared to the results of experiment 3, and not to the results of experiment 2, in which the masker *F*0 was also roved across intervals.

C. Results and discussion

The upper panels in Fig. 4 show the threshold TMRs measured in the current experiment. The lower panels show the difference between these TMRs and those measured in experiment 3 (synchronous maskers, fixed masker *F*0 across observation intervals) or, for the noise maskers, those measured in experiment 2. The differences were computed by subtracting the threshold TMRs measured in experiment 3 (or 2, for the noise maskers) from those measured in the current experiment, so that negative values reflect an improvement (decrease) in thresholds with the introduction of the onset-offset asynchrony. The differences were on average all negative, suggesting that when target-masker asynchronies had an effect, it was beneficial. For the noise maskers, the difference between the TMRs measured in the synchronous condition (experiment 2) and the asynchronous condition (this experiment) failed to reach statistical significance. The lack of effect of asynchronous gating with the noise maskers is probably due to the fact that these maskers differed markedly in timbre from the targets, which already provided a strong cue for perceptual segregation between the target and the masker, and made it unlikely that listeners would confuse target and masker.

For HCT maskers, the difference in TMR between the synchronous condition (experiment 3) and the asynchronous one (this experiment) proved significant at both 200 Hz $[F(1,3)=19.50, p=0.022]$ and 400 Hz $[F(1,3)=11.88, p=0.041]$, but just failed to reach significance at 100 Hz $[F(1,3)=8.15, p=0.065]$. These results are in contrast with those of Carlyon (1996a, 1996b), who found no benefit of target-masker asynchronies on target-*F*0 discrimination performance, even when the target and masker contained resolved harmonics in their passbands, at least prior to being mixed. This suggests that the relatively short (150-ms) asynchrony in the Carlyon studies may have been insufficient to promote the perceptual segregation of concurrent HCTs, at least for the purpose of subsequently discriminating their pitch. The studies of Demany and Semal (1990) and Akeroyd and Summerfield (2000), which found a beneficial effect of target-masker asynchrony on the identification of target vowels, also used onset asynchronies of about 500 ms. The current finding that target-masker asynchronies helped significantly at the 200 and 400 Hz nominal *F*0s indicates that two HCTs containing resolved components in their passband can be perceptually segregated, and that the segregation can be enhanced through asynchronous gating.

Our results also differ from those of Carlyon (1996a, 1996b) in that we never found the asynchronous gating to

have a detrimental effect in the case where the target and masker harmonics contained only unresolved harmonics. This apparent discrepancy in outcomes could be related to the fact that the TMR in Carlyon's study was fixed at 0 dB, whereas in ours the TMR could be positive. At positive TMRs, the masker is lower in level than the target, and so may interfere less with the target, irrespective of whether it is gated synchronously or asynchronously with the target.

The finding that target-masker asynchronies did not significantly help listeners when the nominal target *F*0 was low (100 Hz) agrees with Carlyon's (1996a) observations in suggesting that it may not be possible to perceptually segregate two simultaneous complexes containing only unresolved components, even when large gating asynchronies exist between them.

VII. EXPERIMENT 5. DETECTION OF HARMONIC TARGETS IN NOISE AND HARMONIC MASKERS

A. Rationale

Experiments 2–4 measured the TMR necessary to discriminate a given difference in the *F*0 of the target. Some of the differences between conditions in this study may be related to differences in the audibility of the target. This final experiment was devoted to measuring the detectability of the target in the conditions of experiment 2. Thus, whereas experiment 2 required listeners to hear out the pitch of the target, here listeners were only required to detect its presence. An interesting question is whether and how the detection thresholds relate to the TMR levels required for *F*0 discrimination. One possibility is that the target pitch can be discriminated with the required level of accuracy as soon as the target is detected, in which case the target detection thresholds should be similar to the TMRs measured in the previous experiments. A more likely scenario is that the target must be set some level above its detection threshold in order for its pitch to be accurately perceived. In this case, the interesting questions are how far above its detection threshold must a target HCT be for its pitch to be correctly discriminated, and whether that level is the same or different as the masker or target *F*0 is varied, or as the masker changes from noise to a HCT.

B. Methods

Detection thresholds were measured using a 2I-2AFC procedure. In contrast to all the earlier experiments, the target was presented in only one interval, which listeners had to identify. The interval containing the target was selected randomly with a 0.5 probability prior to each trial. Another difference with the previous experiments is that the level of the masker was roved over a 10 dB range (−5 to +5 dB) around the nominal level across observation intervals as well as across trials, in order to discourage listeners from basing their judgments on differences in global loudness between the stimuli. In all other respects, the targets and maskers in this experiment had identical characteristics to those used in experiment 2.

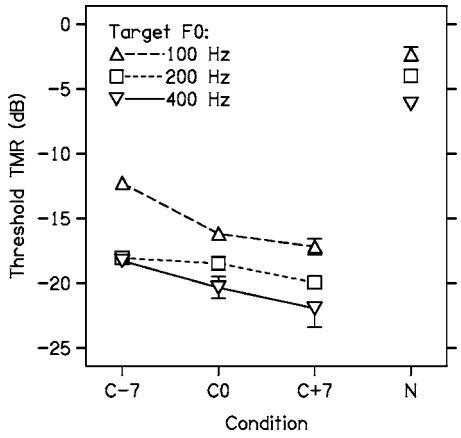


FIG. 5. Masked detection thresholds for target HCTs with nominal *F*0s of 100, 200, and 400 Hz, in the presence of harmonic or noise maskers (experiment 5).

C. Results and discussion

1. Masked detection thresholds

The masked target detection thresholds are shown in Fig. 5. Detection thresholds were considerably higher (i.e., poorer) in the presence of the noise masker than in the presence of a harmonic masker [planned comparison: $F(1,3)=2618.24$, $p<0.0005$]. This result can be explained by considering that when a masker is periodic, listeners can easily detect the disruption in temporal (envelope or fine structure) regularity caused by the addition of a signal with a different periodicity, in what is typically heard as fluctuations or as roughness. In contrast, when the masker is aperiodic, this cue is generally not available (Hellman, 1972; Moore *et al.*, 1998; Treurniet and Boucher, 2001b; 2001a; Gockel *et al.*, 2002; 2003). This effect may be captured by de Cheveigné's (1993, 1999a) harmonic cancellation model, which can be understood as a mechanism for detecting a disruption in temporal regularity.

Thresholds usually decreased (i.e., improved) as the nominal *F*0 of the target increased from 100 to 200 and then 400 Hz. This effect was found for both harmonic maskers [$F(1,3)=48.37, p=0.006$] and noise maskers [$F(1,3)=117.56, p=0.002$]. The decrease in thresholds with increasing *F*0 may reflect the increasing level per component of the target HCT for a given overall stimulus level. To the extent that listeners are not able to integrate spectral information efficiently, thresholds may be dependent on the level of each component within a complex, rather than the overall level of the complex (e.g., van den Brink and Houtgast, 1990). For the *C*0, *C*+7, and *N* masker conditions, the decrease in threshold TMR is roughly equal to the 3 dB per *F*0 doubling that would be expected if this were the case. Another possibility is that the complexes sound more tone-like and less noise-like at higher *F*0s. Gockel *et al.* (2002) found a similar pattern of results to those reported here when the HCTs were in random phase, but not when they were in sine phase, which generally produces a more salient pitch.

The improvement in target detection thresholds with increasing *F*0 in the HCT masker conditions may be more an effect of masker *F*0 than of target *F*0. This is suggested by

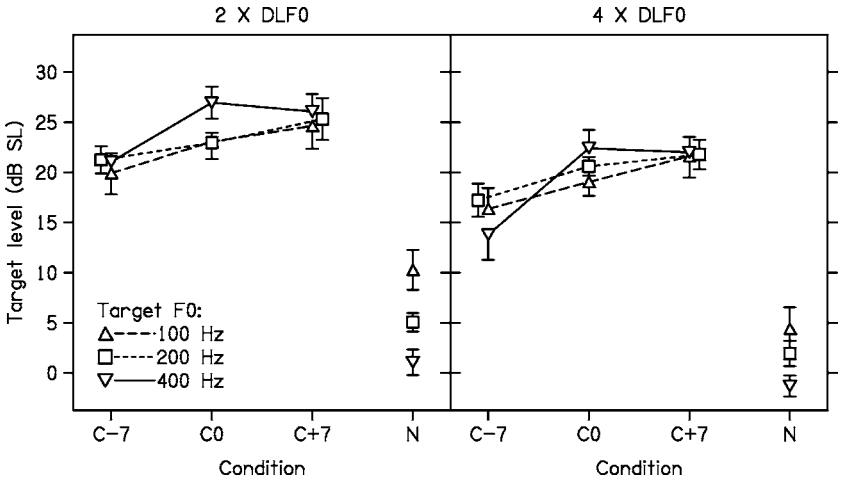


FIG. 6. Target levels above detection threshold (dB SL) required for 79.4%-correct target $F0$ discrimination in the presence of harmonic or noise maskers, showing the level required, relative to the detection threshold, to discriminate the changes in target $F0$ with an accuracy of 79.4% correct. Each "SL" value was computed by subtracting the detection threshold measured in experiment 2 (Fig. 2) from the TMR measured in experiment 5 (Fig. 5).

the following two observations: First, in the HCT masker conditions, thresholds decreased (improved) consistently with increases in the nominal $F0$ of the masker, relative to that of the target; thresholds in the $C-7$ condition were higher than in the $C0$ condition [$F(1,3)=13.93, p=0.034$], which in turn were higher than in the $C+7$ condition [$F(1,3)=20.54, p=0.020$]. Second, a comparison of conditions with very similar masker $F0$s but different target $F0$s suggests only a small and inconsistent effect of target $F0$ in the presence of a harmonic masker.[3] Thus, in contrast to thresholds in noise, thresholds in the presence of harmonic maskers do not seem to be strongly dependent on the target $F0$ (at least within the range tested here), but instead depend more on the masker $F0$. The reason for the dependence of thresholds on masker $F0$ may relate to the spectral spacing of masker components. As the $F0$ increases, the spectral gaps between masker components also increase, thereby increasing the possibility for detecting the target in spectral regions between successive masker components.

2. Levels above masked threshold required for accurate target F0 discrimination

In order to investigate how the TMRs required for correct target $F0$ discrimination were related to the TMRs required for correct target detection, we subtracted the threshold TMRs for detection measured in this experiment from the TMRs measured for target $F0$ discrimination in experiment 2. The resulting values are referred to as target sensation levels (SL). The SL indicates the level required, relative to the detection threshold, to discriminate the given changes in target $F0$ with an accuracy of 79.4% correct. Presenting the results in this way allows a determination of the role of audibility in determining listeners' abilities to hear out the $F0$ of the target complex. If differences in threshold TMR across conditions resulted from differences in audibility alone, then these conditions would yield similar threshold TMRs when expressed in SL terms.

Target SLs are shown in Fig. 6. The format of this figure is similar to that of Fig. 2, with the left- and right-hand panels showing the results obtained with target $\Delta F0$s set to two and four times the listener's DL$F0$, respectively. The important difference between the current figure and Fig. 2 is that here, the threshold TMRs in ordinate are expressed in dB SL, rather than relative to the masker level.

Several observations are apparent in the SL data plotted in Fig. 6. First, with the noise masker, the target SLs decreased significantly as the target $F0$ increased from 100 to 200–400 Hz [linear contrast: $F(1,3)=12.81$, $p=0.037$]. This confirms our earlier statement that, in the noise masking conditions of experiment 2, the improvement in threshold TMRs with increasing target $F0$ (shown in Fig. 2) could not be entirely explained in terms of target audibility. For the harmonic maskers, the threshold TMRs in dB SL (shown in Fig. 6) were not found to vary significantly with the target $F0$. Thus, the improvements in threshold TMRs with increasing target and masker $F0$ observed in the harmonic-masker conditions of experiment 2 (Fig. 2) were paralleled by similar improvements in detection thresholds (Fig. 5). The improvements in discrimination and detection in the presence of HCT maskers may depend on the separation of the spectral components of the target and masker, which increased with $F0$.

The target SLs measured using harmonic maskers were generally much higher than those measured with noise maskers [contrast analysis: $F(1,3)=656.35$, $p<0.001$]. Specifically, with harmonic maskers, listeners typically needed target levels between 15 and 25 dB above the detection threshold in order to correctly discriminate the changes in target $F0$. This suggests that, in HCT-masker conditions, the ($F0$) discriminability of the targets was not determined primarily by their detectability. In contrast, with noise maskers, listeners could correctly discriminate changes in the $F0$s of targets whose level was only a few decibels above their detection threshold. This is consistent with the idea that the detection of a HCT in noise may be mediated by the detection of pitch (Hafter and Saberi, 2001). A similar conclusion, for the detection of frequency modulation on a harmonic complex, was reached by Carlyon and Stubbs (1989).

VIII. GENERAL DISCUSSION

A. On the possible influence of target-masker confusion

We have generally discussed the results in terms of the ability of listeners to perceptually segregate the target from

the masker. However, another logical possibility is that the target and masker are perceptually segregated (i.e., produce two separate percepts with different pitches), but that thresholds reflect a confusion on the part of the subject as to which pitch belongs to the target and which to the masker. This appears to be unlikely in the case of the 100-Hz *F*0 targets: in all cases, manipulations that should have reduced confusion, as well as improved segregation, such as keeping the masker *F*0 constant or introducing asynchronous gating, had no significant effect on threshold.

Confusion was most likely to play a role when the target *F*0 was 400 Hz and masker and target levels and *F*0s were most similar. For instance, in Fig. 2, it can be seen that the highest thresholds (poorest performance) in the 400-Hz target conditions were observed when the target and masker occupied the same *F*0 range (*C*0 conditions). On the other hand, keeping the *F*0 constant across trials within each interval (Fig. 3) did not selectively improve performance in the *C*0 condition, as might have been expected if the results in Fig. 2 were dominated by confusion effects. Thus, it appears that confusion did not play a major role in determining thresholds.

B. The role of peripheral resolvability in concurrent sound segregation

When presented in isolation, the targets in the present study comprised resolved harmonics in the *F*0 ranges of 200 and 400 Hz, but only unresolved harmonics in the 100-Hz *F*0 range. The experiments showed that manipulations designed to enhance the perceptual segregation of the masker and target improved performance for the 200- and 400-Hz targets, but not for the 100-Hz targets. This is probably because, in the case of the 100-Hz target, only one pitch was heard at a time, and the target was only discriminable when it was perceptually more salient than the masker. It might therefore be tempting to conclude that resolved harmonics are crucial for hearing out one harmonic sound in the presence of another. However, it is important to consider not only the resolvability of the target components in isolation, but also their resolvability once they are added to the masker.

The resolvability of components within a mixture depends on several factors, including the actual *F*0s of the target and masker, their relative and absolute levels and what criterion is used in defining resolvability. To provide some initial guidance, we assume the case of equal-amplitude harmonics in both the target and masker. Under such circumstances, resolvability is traditionally estimated either in terms of the minimum spacing between components (e.g., Plomp, 1964) or in terms of the number of components falling within the bandwidth of auditory filters with center frequencies within the stimulus passband (e.g., Shackleton and Carlyon, 1994). Here, we used a technique consistent with both of these approaches. We assume that a component is resolved if no other component falls within the 10-dB bandwidth of the auditory filter centered at the component's frequency. Note that this is equivalent to requiring a minimum spacing between a component and its nearest neighbor in order for that component to be deemed resolved; in the present case, that minimum distance is equal to half the 10-dB bandwidth of a model auditory filter. The 10-dB bandwidth equals 1.8 times the equivalent rectangular bandwidth (ERB) defined as: $21.4 \log_{10}(4.37\mathrm{CF}+1)$, where CF is the filter center frequency, in kilohertz (Glasberg and Moore, 1990). Thus, we consider that a target component is resolved if its nearest neighbor is at least 0.9 ERB away from it one either side. Since 0.9 ERB is almost equivalent to a critical band, our approach is broadly consistent with the results of Plomp (1964), who concluded that "the ear is able to distinguish a simple tone in a complex sound if the frequency distance to the adjacent tones exceeds the critical bandwidth." Our approach for estimating resolvability also agrees with the common view, inspired by early as well as recent findings (e.g., Bernstein and Oxenham, 2003), that harmonics below the 10th are usually resolved. Indeed, 0.9 ERB is close to about 10%, which is roughly the distance between the 10th harmonic in a single harmonic complex and its nearest neighbors, and is also similar to the minimum distance between audible partials, as established by Plomp (1964).

Using these criteria, it is easy to show that in the 100-Hz target *F*0 condition, no target component was resolved within the stimulus passband (1200–3000 Hz) prior to the addition of the masker, from which it obviously follows that no target component was resolved after the masker was added. For the 200- and 400-Hz target *F*0 conditions, the situation is less straightforward, and computer simulations were performed in order to determine on what proportion of the trials, on average, at least one component of the target was resolved. The simulations used the same stimulus parameters as in the experiments, including the relative positions and roving range of the target and masker *F*0s. The results revealed that in conditions where the nominal *F*0 of the target was 200 Hz, the addition of the masker resulted in none of the target components being resolved. For the 400-Hz nominal target *F*0, the proportion of trials on which at least one target component was resolved in the mixture was less than 1% for the $C-7$ condition, around 6% for the *C*0 condition, and around 30% for the $C+7$ condition. Thus, for the case in which the harmonics of the target and masker are of equal amplitude, only the components of the 400-Hz *F*0 target were partially resolved when added to the masker.

This simple analysis, in combination with our results showing good performance in many conditions with a 200-Hz *F*0 target suggest the possibility that a sound may be successfully segregated from another even if none of its components is resolved within the mixture. This seems to contradict the findings of Beerends and Houtsma (1986; 1989), who found that at least one component from each two-tone complex had to be resolved for the two pitches to be successfully identified. However, their complexes consisted of only two components, which generally produce a relatively weak pitch percept, even in isolation. Nevertheless, it is premature to draw strong conclusions based on this simple analysis. More insight might be gained from comprehensive computational modeling, incorporating both relative and absolute level effects; this is beyond the scope of the current study.

C. Implications for hearing impairment and artificial auditory-scene analysis

Multiple harmonic sources are common in everyday listening environments, such as multitalker conversations and music, and present particular challenges to hearing-impaired listeners and cochlear-implant users. The current results, showing that higher TMRs are required for the successful separation of concurrent HCTs when the spectral components are not well separated at the auditory periphery, contribute to further establishing the link between the loss of peripheral frequency resolution that is often associated with cochlear damage and the difficulties in complex acoustic environments that are often reported by hearing-impaired listeners. At the same time, our finding of negative threshold TMRs in conditions where no resolved harmonic were present (e.g., the 400-Hz nominal target *F*0, $C-7$ condition in experiment 2, or the 200-Hz nominal target *F*0 conditions in experiment 4) suggest that hearing out sounds may not rely solely on their spectral resolution within a mixture.

A second area where the present findings may have interesting applications is the design of artificial auditory scene analysis systems. In recent years, mechanisms have been proposed for the separation of simultaneous HCTs, which operate based on cues present in the temporal pattern of activity from individual peripheral channels (Cariani, 2001; de Cheveigné, 2003), leading to the possibility that segregation can be achieved without spectrally resolved components. While the present results provide some evidence that concurrent HCTs can indeed be separated with some degree of success even when no resolved harmonics are present, they also indicate that factors that promote increased peripheral separation between spectral components usually promote perceptual separation. Furthermore, the observation that in conditions in which the peripheral separation of the spectral components is poorest (e.g., at the 100-Hz nominal target *F*0), positive TMRs are usually required for successful extraction of the target *F*0, suggests that some degree of spectral analysis may be required for the successful separation of simultaneous HCTs by human listeners. This is consistent with de Cheveigné's (2001, 2005) proposal that peripheral filtering may enhance the target-to-masker ratio in some channels, up to a point where temporal source-segregation mechanisms can operate effectively.

IX. SUMMARY

The ability to detect and discriminate *F*0 differences in HCT targets was measured in normal-hearing listeners in the presence of spectrally overlapping HCT or noise maskers. The following conclusions were drawn:

(1) The TMR required for listeners to be able to accurately discriminate changes in the *F*0 of a target HCT varied depending on the nature of the masker. With noise maskers, the TMR for *F*0 discrimination was generally within 10 dB of the TMR required for detection. In contrast, with a harmonic masker, the threshold TMRs for *F*0 discrimination were typically between 15 and 25 dB higher than the threshold TMR for detection. Thus, detection thresholds are probably not a valid measure of listeners' ability to perceptually segregate concurrent HCTs.

(2) When the target comprised only unresolved harmonics (100-Hz *F*0 range), TMRs were always above 0 dB for the discrimination tasks, suggesting that listeners were only able to perform the task when the target *F*0 dominated the percept, and were not able to hear the pitches of both the target and the masker simultaneously, in line with earlier studies (Beerends and Houtsma, 1986; Carlyon, 1996a).

(3) When the target contained resolved harmonics (200- and 400-Hz *F*0 range), TMRs were reduced (improved) by manipulations designed to assist in perceptually segregating the masker and target, such as difference in *F*0 range and onset/offset asynchronies. No such changes were observed with the 100-Hz *F*0 target, again suggesting that listeners' ability to hear out the pitches of harmonic complexes presented simultaneously in the same spectral region is limited by the peripheral resolvability of their components.

(4) At low or negative TMRs, even the 200- and 400-Hz targets probably contained only unresolved harmonics when combined with the masker. Therefore, harmonics may not need to remain spectrally resolved after being mixed with another source in order for the pitches of two simultaneous HCTs to be heard. However, further empirical and computational modeling studies will be required to provide a more stringent test of this preliminary conclusion.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (NIDCD Grant No. R01 DC 05216). The authors are grateful to R. P. Carlyon, A. de Cheveigné, and J. H. Grose, for helpful suggestions on an earlier version of this manuscript.

[1]Noninteger degrees of freedom in the reported *F* statistics reflect the use of the Greenhouse-Geisser correction whenever the sphericity assumption was not met.

[2]Carlyon *et al.* (2002) showed that the pitch evoked by a mixture of two HCTs comprised of unresolved harmonics corresponded to that of the HCT with the higher *F*0, and appeared to be related to the mean first-order interval between pitch pulses in the mixture. If listeners' judgments in the present experiment were based on this mean-rate pitch, thresholds should have been lower in the condition where the target *F*0 was higher than the masker's, since in that situation, according to Carlyon *et al.*'s (2002) results, the perceived *F*0 probably corresponded to that of the target. The fact that this prediction was not borne out in the present data suggests that the listeners did not rely on the mean-rate cue here.

[3]The relevant comparisons in Fig. 5 are the 100-Hz/$C+7$-semitones condition with the 200-Hz/C-7-semitones condition, and the 200-Hz/$C+7$-semitones condition with the 400-Hz/C-7-semitones condition. Both pairs of conditions have masker *F*0 ranges within a semitone of each other, but target *F*0s one octave apart. The differences between the conditions in each pair are small (2 dB or less) and in opposite directions.

Akeroyd, M. A., and Summerfield, Q. (**2000**). "Integration of monaural and binaural evidence of vowel formants," J. Acoust. Soc. Am. **107**, 3394–3406.

Assmann, P. F., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Assmann, P. F., and Summerfield, Q. (**1989**). "Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency," J. Acoust. Soc. Am. **85**, 327–338.

Assmann, P. F., and Summerfield, Q. (**1994**). "The contribution of waveform interactions to the perception of concurrent vowels," J. Acoust. Soc. Am. **95**, 471–484.
Beerends, J. G., and Houtsma, A. J. M. (**1986**). "Pitch identification of simultaneous dichotic two-tone complexes," J. Acoust. Soc. Am. **80**, 1048–1055.
Beerends, J. G., and Houtsma, A. J. M. (**1989**). "Pitch identification of simultaneous diotic and dichotic two-tone complexes," J. Acoust. Soc. Am. **85**, 813–819.
Bernstein, J. G., and Oxenham, A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?," J. Acoust. Soc. Am. **113**, 3323–3334.
Bernstein, J. G. W., and Oxenham, A. J. (**2005**). "An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination," J. Acoust. Soc. Am. **117**, 3816–3831.
Bilsen, F. A. (**1973**). "On the influence of the number and phase of harmonics on the perceptibility of the pitch of complex signals," Acustica **28**, 60–65.
Bregman, A. S. (**1990**). *Auditory Scene Analysis; The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).
Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.
Cariani, P. (**2001**). "Neural timing nets," Neural Networks **14**, 737–753.
Carlyon, R. P., van Wieringen, A., Long, C. J., Deeks, J. M., and Wouters, J. (**2002**). "Temporal pitch mechanisms in acoustic and electric hearing," J. Acoust. Soc. Am. **112**, 621–633.
Carlyon, R. P. (**1996a**). "Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker," J. Acoust. Soc. Am. **99**, 517–524.
Carlyon, R. P. (**1996b**). "Masker asynchrony impairs the fundamental-frequency discrimination of unresolved harmonics," J. Acoust. Soc. Am. **99**, 525–533.
Carlyon, R. P. (**1997**). "The effect of two temporal cues on pitch judgments," J. Acoust. Soc. Am. **102**, 1097–1105.
Carlyon, R. P., and Shackleton, T. M. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.
Carlyon, R. P., and Stubbs, R. J. (**1989**). "Detecting single-cycle frequency modulation imposed on sinusoidal, harmonic, and inharmonic carriers," J. Acoust. Soc. Am. **85**, 2563–2574.
Culling, J. F., and Darwin, C. J. (**1993**). "Perceptual separation of simultaneous vowels: Within and across-formant grouping by *F*0," J. Acoust. Soc. Am. **93**, 3454–3467.
Culling, J. F., and Darwin, C. J. (**1994**). "Perceptual and computational separation of simultaneous vowels: Cues arising from low-frequency beating," J. Acoust. Soc. Am. **95**, 1559–1569.
Darwin, C. J., and Carlyon, R. P. (**1995**). "Auditory organization and the formation of perceptual streams," in *Handbook of Perception and Cognition, Volume 6. Hearing*, edited by B. C. J. Moore (Academic, San Diego).
Darwin, C. J., and Ciocca, V. (**1992**). "Grouping in pitch perception: effects of onset asynchrony and ear of presentation of a mistuned component," J. Acoust. Soc. Am. **91**, 3381–3390.
de Cheveigné, A. (**1993**). "Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing," J. Acoust. Soc. Am. **93**, 3271–3290.
de Cheveigné, A. (**1997**). "Concurrent vowel identification. III. A neural model of harmonic interference cancellation," J. Acoust. Soc. Am. **101**, 2857–2865.
de Cheveigné, A. (**1999a**). "Pitch shifts of mistuned partials: a time-domain model," J. Acoust. Soc. Am. **106**, 887–897.
de Cheveigné, A. (**1999b**). "Waveform interactions and the segregation of concurrent vowels," J. Acoust. Soc. Am. **106**, 2959–2972.
de Cheveigné, A. (**2001**). "The auditory system as a separation machine," in *Physiological and Psychophysical Bases of Auditory Function*, edited by J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven Shaker Publishing BV, Maastricht, The Netherlands, pp. 453–460.
de Cheveigné, A. (**2003**). "Scene analysis without spectral analysis?," J. Acoust. Soc. Am. **113**, 2231(A).
de Cheveigné, A. (**2005**). "Separable representations for cocktail party processing," Forum Acusticum (FA2005), Budapest, pp. 1527–1531.
de Cheveigné, A., McAdams, S., Laroche, J., and Rosenberg, M. (**1995**). "Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement," J. Acoust. Soc. Am. **97**, 3736–3748.
Demany, L., and Semal, C. (**1990**). "The effect of vibrato on the recognition of masked vowels," Percept. Psychophys. **48**, 436–444.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Gockel, H., Carlyon, R. P., and Plack, C. J. (**2004**). "Across-frequency interference effects in fundamental frequency discrimination: Questioning evidence for two pitch mechanisms," J. Acoust. Soc. Am. **116**, 1092–1104.
Gockel, H., Moore, B. C., and Patterson, R. D. (**2002**). "Asymmetry of masking between complex tones and noise: The role of temporal structure and peripheral compression," J. Acoust. Soc. Am. **111**, 2759–2770.
Gockel, H., Moore, B. C., and Patterson, R. D. (**2003**). "Asymmetry of masking between complex tones and noise: Partial loudness," J. Acoust. Soc. Am. **114**, 349–360.
Hafter, E. R., and Saberi, K. (**2001**). "A level of stimulus representation model for auditory detection and attention," J. Acoust. Soc. Am. **110**, 1489–1497.
Hellman, R. P. (**1972**). "Asymmetry of masking between noise and tone," Percept. Psychophys. **11**, 241–246.
Hoekstra, A. (**1979**). "Frequency discrimination and frequency analysis in hearing," Ph.D. thesis, Institute of Audiology, University Hospital, Groningen, Netherlands.
Houtsma, A. J. M., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am. **87**, 304–310.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics, "J. Acoust. Soc. Am. **49**, 467–477.
Meddis, R., and Hewitt, M. (**1992**). "Modeling the identification of concurrent vowels with different fundamental frequencies," J. Acoust. Soc. Am. **91**, 233–245.
Moore, B. C., Alcántara, J. I., and Dau, T. (**1998**). "Masking patterns for sinusoidal and narrow-band noise maskers," J. Acoust. Soc. Am. **104**, 1023–1038.
Parsons, T. W. (**1976**). "Separation of speech from interfering speech by means of harmonic selection," J. Acoust. Soc. Am. **60**, 911–918.
Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.
Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on *F*0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.
Scheffers, M. T. M. (**1983**). "Sifting vowels: Auditory pitch analysis and sound segregation," Ph.D. thesis, Groningen University, The Netherlands.
Stubbs, R. J., and Summerfield, Q. (**1988**). "Evaluation of two voice-separation algorithms using normally-hearing and hearing impaired listeners," J. Acoust. Soc. Am. **84**, 1236–1249.
Summerfield, Q., and Assmann, P. F. (**1991**). "Perception of concurrent vowels: Effects of pitch-pulse asynchrony and harmonic misalignment," J. Acoust. Soc. Am. **89**, 1364–1377.
Treurniet, W. C., and Boucher, D. R. (**2001a**). "Masked threshold difference due to masker harmonicity: Uniform modulation rates of auditory filter outputs versus periodicity in waveform fine structure," J. Acoust. Soc. Am. **110**, 1267–1270.
Treurniet, W. C., and Boucher, D. R. (**2001b**). "A masking level difference due to harmonicity," J. Acoust. Soc. Am. **109**, 306–320.
van den Brink, W. A. C., and Houtgast, T. (**1990**). "Spectro-temporal integration in signal detection," J. Acoust. Soc. Am. **88**, 1703–1711.
Weintraub, M. (**1987**). "Sound separation and auditory perceptual organisation," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Martinus Nijhoff, Dordrecht).
Zwicker, U. T. (**1984**). "Auditory recognition of diotic and dichotic vowel pairs," Speech Commun. **3**, 265–277.

The influence of spatial separation on divided listening[a)]

Virginia Best, Frederick J. Gallun, Antje Ihlefeld, and Barbara G. Shinn-Cunningham[b)]
Hearing Research Center, Boston University, Boston, Massachusetts 02215

(Received 12 January 2006; revised 21 June 2006; accepted 29 June 2006)

If spatial attention acts like a "spotlight," focusing on one location and excluding others, it may be advantageous to have all targets of interest within the same spatial region. This hypothesis was explored using a task where listeners reported keywords from two simultaneous talkers. In Experiment 1, the two talkers were placed symmetrically about the frontal midline with various angular separations. While there was a small performance improvement for moderate separations, the improvement decreased for larger separations. However, the dependency of the relative talker intensities on spatial configuration accounted for these effects. Experiment 2 tested whether spatial separation improved the intelligibility of each source, an effect that could counteract any degradation in performance as sources fell outside the spatial spotlight of attention. In this experiment, intelligibility of individual sources was equalized across configurations by adding masking noise. Under these conditions, the cost of divided listening (the drop in performance when reporting both messages compared to reporting just one) was smaller when the spatial separation was small. These results suggest that spatial separation enhances the intelligibility of individual sources in a competing pair but increases the cost associated with having to process both sources simultaneously, consistent with the attentional spotlight hypothesis. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2234849]

PACS number(s): 43.66.Dc, 43.66.Pn, 43.66.Qp [AJO] Pages: 1506–1516

I. INTRODUCTION

In the 1950s, Broadbent described auditory attention using the Filter Theory, in which some stimuli are filtered out and others are admitted on the basis of basic characteristics such as frequency and spatial location (Broadbent, 1958). This idea has since been developed most extensively for spatial attention in the visual perception literature, where it is known as the "spotlight" model. In the visual system a representation of space is available directly at the periphery, and is largely maintained at higher levels of the system. The spatial spotlight of attention is thought to operate directly on this representation to modulate competition between simultaneous objects. However, full development of the spotlight model in auditory spatial attention has proven to be a challenge for researchers. Although there is evidence that listeners can orient their attention spatially to enhance the detection and identification of simple targets (Spence and Driver, 1994; Mondor and Zatorre, 1995; Quinlan and Bailey, 1995), the role of spatial attention in the case of *simultaneous* sound sources is less clear. One difficulty lies in the fact that auditory location is computed in the auditory system rather than being represented in the sensory epithelium. As the peripheral representation is a frequency map, sounds coming from different locations often share the same set of peripheral receptors, and auditory source location must be computed from the mixture of signals reaching the left and right ears. This means that the ability to admit an acoustic source at one location and filter out a source at another location will be limited by the ability to separate the acoustic energy coming from different locations in addition to any constraints on the ability to distribute spatial attention.

When a listener must extract the content of one source (a "target") in the presence of competing sources ("maskers"), spatial separation of the target and masker is generally beneficial to performance (Bronkhorst, 2000). When the masker reduces the audibility of components of the target ("energetic masking"), there are two ways in which spatial separation offers an advantage. First, the relative energy of the target and masker at the ears changes with target and masker location, such that spatial separation increases the target audibility in each frequency band at one of the ears (the "better ear"). Second, binaural processing allows listeners to detect the presence of target energy at a particular time and frequency band if the target and masker contain different interaural time and/or level differences (Zurek, 1993; Bronkhorst, 2000). When competing sources do not have significant frequency overlap and reduced audibility is not the primary source of interference, a masker with similar spectrotemporal characteristics can still interfere with the perception of the target (so-called "informational masking"). One important source of informational masking is trial-to-trial variability in the target and/or masker, which leads to listener uncertainty as to how to classify a given spectrotemporal pattern. This kind of interference is reduced when the target and masker are distinguished in a way that reduces confusions between them. For example, differences in perceived spatial location have been shown to reduce informational masking by allowing listeners to selectively attend to the target at the location of interest (Freyman *et al.*, 1999; Freyman *et al.*, 2001; Best

[a)]Portions of this work were presented at the 2005 International Conference on Auditory Display and the 2006 mid-winter meeting of the Association for Research in Otolaryngology.

[b)]Author to whom all correspondence should be addressed. Electronic mail: shinn@cns.bu.edu

et al., 2005; Kidd *et al.*, 2005; Shinn-Cunningham *et al.*, 2005). In these situations, the spotlight of attention may play a role.

This study examines the effect of spatial separation when a listener must attend to *two* sustained sound sources simultaneously. In the example used here, keywords had to be extracted from each of two talkers in a competing pair. Previous studies of divided listening with speech have typically used dichotic signals in which each ear receives only one of the two competing sources (Broadbent, 1954; Massaro, 1976) and have not considered spatial factors in detail. One recent study (Shinn-Cunningham and Ihlefeld, 2004) examined the effect of spatially separating two competing talkers (by 90°) on the ability of listeners to report both messages. In that study, in which the talkers were presented at different relative intensities, the louder talker could always be recalled with relative ease. As a result, listeners appeared to allocate attention primarily to the quieter talker, a strategy similar to that employed in a selective attention task. Spatial separation improved performance, probably for the same reasons it improves performance in a true selective listening task (discussed above). In contrast, in the present study the two talkers were presented with equal intensity and were separated symmetrically about the midline. Thus, the two talkers are equally difficult to hear, and processing resources should be more equally allocated between the two competing talkers (i.e., the listening strategy is more likely to engage truly "divided" listening). In this case, it is not clear what the effect of spatial separation of the two targets might be. It is reasonable to expect that spatial separation would be advantageous in that it would enhance the audibility of the two sources as well as reducing confusion between them, as described above. However, if one considers the putative spotlight of spatial attention, spatial separation could be detrimental in a divided listening task. If the spotlight is focused at a given moment on one source, then the other is likely to be excluded if it is distant from the first, and simultaneous processing will be impaired.

In Experiment 1, the effect of spatial separation on the ability of listeners to report keywords from two simultaneous talkers was examined. Results suggest that there is little effect of spatial separation overall, except for some modulation of performance due to changes in energy at the two ears. In Experiment 2, an attempt was made to untangle two potentially opposing effects: (1) a benefit of spatial separation for segregating competing messages; and (2) a disadvantage of spatial separation when dividing spatial attention.

II. EXPERIMENT 1

A. Methods

1. Subjects

Eight paid subjects (ages 20–30) participated in the experiment. Four subjects had previous experience in psychophysical studies of a similar nature. All subjects participated in Experiment 1A, and six of the subjects went on to participate in Experiment 1B.

2. Speech materials

Speech materials were spoken sentences taken from the publicly available Coordinate Response Measure speech corpus (Bolia *et al.*, 2000). These sentences all contain seven words, three of which are keywords that vary from utterance to utterance. The form of the sentences is "Ready *call-sign* go to *color number* now," where the italicized words indicate keywords. In the corpus there are eight possible call-signs ("arrow," "baron," "charlie," "eagle," "hopper," "laker," "ringo," "tiger"), four possible colors ("blue," "green," "red," "white"), and eight possible numbers (1–8). All combinations of these words produce 256 phrases, which are each spoken by eight talkers (four male, four female), giving a total of 2048 sentences. The sentences are time aligned such that the word "ready" always starts at the same time, but some variations in overall rhythm occur between different sentences so that the keywords in different utterances are not exactly aligned.

3. Stimuli

For each trial, two sentences spoken by the same talker were chosen randomly from the corpus with the restriction that all keywords differed in the two sentences. In order to reduce the energetic interference between the two sentences, they were processed to produce intelligible speechlike signals that had little spectral overlap (Shinn-Cunningham *et al.*, 2005; see also Arbogast *et al.*, 2002; Brungart *et al.*, 2005, for similar approaches). The signals were bandpass filtered into eight nonoverlapping frequency bands of 1/3 octave width, with center frequencies spaced evenly on a logarithmic scale from 175 to 4400 Hz. Four bands were randomly chosen for the first sentence (two from the four lower bands and two from the four higher bands). The Hilbert envelope of each band was then used to modulate a sinusoidal carrier at the center frequency of that band, and the sentence was reconstructed by summing the four modulated sinusoids. For the second sentence, the remaining four frequency bands were chosen and the same procedure was followed. The two reconstructed sentences were root-mean-square (rms) normalized to result in a relative level of 0 dB (see Fig. 1 for example spectra).

The stimuli were processed to create binaural signals containing realistic spatial cues and presented over headphones. In Experiment 1A, a full set of spatial cues was used in the simulation. Binaural stimuli were created by convolving the speech signal with the appropriate anechoic left and right head-related transfer functions (HRTFs) measured on a KEMAR manikin at a distance of 1 m (Brungart and Rabinowitz, 1999). The two binaural stimuli were then added to simulate the two speech sources at their desired locations in external space. In Experiment 1B, level differences that were present in the HRTF simulation were removed in order to eliminate location-dependent changes in the relative level of the two sentences at the ears; thus only one spatial cue (the interaural time difference, ITD) was used in these simulations. A single, frequency-independent ITD was extracted from each HRTF pair by finding the time delay of the peak in the broadband interaural cross-correlation function. These

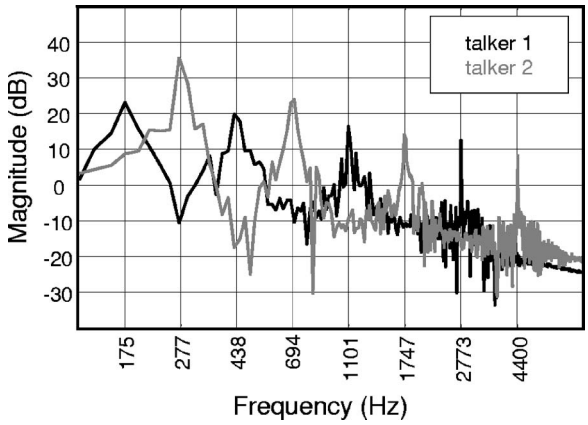


FIG. 1. Frequency spectra of two example sentences after processing. The two signals were processed to minimize their spectral overlap. Sentences were divided into eight 1/3 octave bands with center frequencies between 175 and 4400 Hz. Four different bands were chosen for each sentence and their envelopes used to modulate sinusoids at the center frequency of each band. Intelligible speech signals were reconstructed by summing the four modulated sinusoids.

ITDs were used to delay the left and right ear signals relative to one another to control the perceived lateral locations of the competing sources.

4. Procedure

Subjects were seated in a sound-treated booth in front of a personal computer (PC) terminal. Presentation of the stimuli was controlled by a PC, which selected the stimulus to play on a given trial. Digital stimuli were sent to Tucker-Davis Technologies hardware for digital-analog conversion and attenuation before presentation over insert headphones (Etymotic Research ER-2). Following each presentation, subjects indicated their responses by clicking on a graphical user interface displayed on the PC monitor. The interface consisted of an eight-by-four grid, with the color of the square and the number printed within it representing one of the 32 possible color/number pairs.

In each trial, two simultaneous sentences were presented and subjects were required to respond with *two* color/number pairs (in either order). No feedback was provided. A response was considered correct only if both color/number pairs were reported correctly. Note that chance performance, achieved by randomly guessing the two color/number pairs, is 0.3% for this task.

Stimulus locations were all on the horizontal plane passing through the ears (0° elevation). Performance was measured with sources separated symmetrically about the midline with separations of 0°, 30°, 60°, 90°, 120°, 150°, or 180°. The seven configurations were presented five times in a random order in each run. Each subject completed ten such runs for each experiment, for a total of 50 responses for each configuration. The 20 runs (ten each for Experiments 1A and 1B) were carried out over four to five sessions. Subjects did no more than one hour of testing per day.

5. Training

Before the start of the experiment, subjects participated in a short series of training runs designed to familiarize them with the stimuli and task. In a training test, subjects were presented with stimuli containing a single sentence in quiet, and were required to indicate the color/number pair they perceived. After each trial, correct-answer feedback was provided by a written message on the screen. A training run consisted of 130 trials. Subjects completed as many runs as required to bring their proportion of correct responses to at least 95%. All subjects reached this level within three training runs.

B. Results—Experiment 1A

Individual subjects differed in their absolute level of performance, but overall trends were similar. Mean percent correct scores across subjects (and standard errors) are shown in Fig. 2(a). Spatial separation had a modest effect on performance; for a given subject, performance did not vary by more than 30 percentage points across all spatial configurations. However, there were consistent patterns in the data: performance tended to first increase and then decrease with increasing source separation, peaking at 90°–120° separation.

In order to factor out overall differences in subject performance and concentrate on the effect of spatial separation, percent correct scores for each subject were normalized by subtracting the percent correct in the colocated (separation 0°) configuration. The resulting normalized values summarize how performance changed with source separation. Figure 2(b) shows the normalized data pooled across the eight subjects (means and standard errors). The trends described for the raw data are reinforced when individual differences are factored out in this way: increasing the spatial separation tended to first improve and then degrade performance. A repeated measures analysis of variance (ANOVA) confirmed that there was a significant effect of spatial separation on performance $[F(6,42)=7.1,\ p<0.001]$. Post-hoc analyses (pairwise comparisons with a Bonferroni correction) indicated that separations of 60°, 90°, 120°, and 150° were all significantly different from the colocated configuration (no other comparisons reached significance).

Although the two targets were nominally presented with equal intensity, variations in the HRTFs with source location caused variations in the level of each target at each ear. This is especially true for a target placed to the side, where the acoustic shadow cast by the head greatly attenuates the level received at the far ear, particularly at high frequencies (above about 2 kHz). Indeed, for a given spatial configuration, each of the two sources has a different ear in which their level (relative to that of the other source) is greater. Moreover, the magnitude of this better ear "level ratio" varies as a function of the spatial configuration. Note, however, that as the stimuli were composed of nonoverlapping frequency bands, the level ratio does not correspond to signal-to-noise ratio in the traditional sense (i.e., it is not the "within frequency band" signal-to-noise ratio and thus is not a direct measure of energetic masking). It may be better described as representing an overall loudness ratio of one target relative to the other.

An acoustic analysis was performed to examine whether

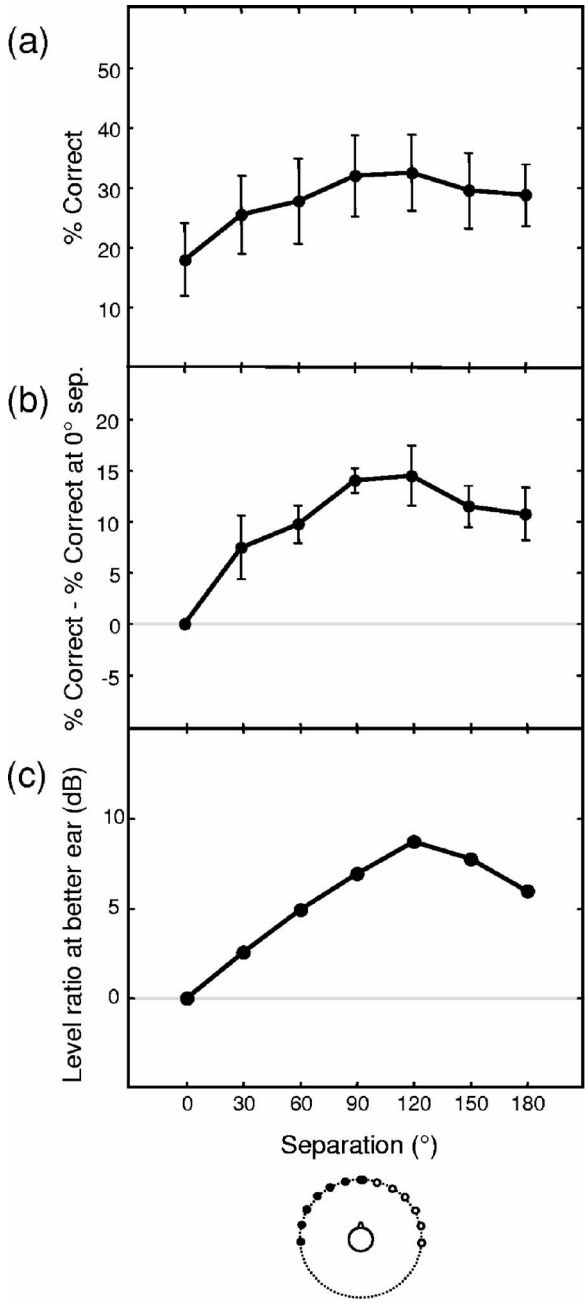


FIG. 2. (a) Mean percent correct scores for the different spatial configurations in Experiment 1A. Results are pooled across the eight subjects and error bars indicate standard error of the mean. (b) Normalized percent correct scores for Experiment 1A. Normalization was carried out for each individual by subtracting the score for the colocated configuration. Results are pooled across the eight subjects and error bars indicate standard error of the mean. (c) Level ratios for the different spatial configurations. Level ratios describe the level of a source in its "better ear" relative to the level of the other source. These ratios were calculated for 50 example stimuli and the means across the two sources (at their respective better ears) are shown.

the relative level of the competing sources at the ears might help to explain the trends seen in the behavioral data. For each spatial configuration, 50 speech pairs were generated and the level ratio (LR) for each source was calculated using the broadband rms level of each source after HRTF filtering for each ear. The changes in better-ear LR as a function of spatial separation (averaged across the two sources and their respective better ears) are shown in Fig. 2(c). Note that by

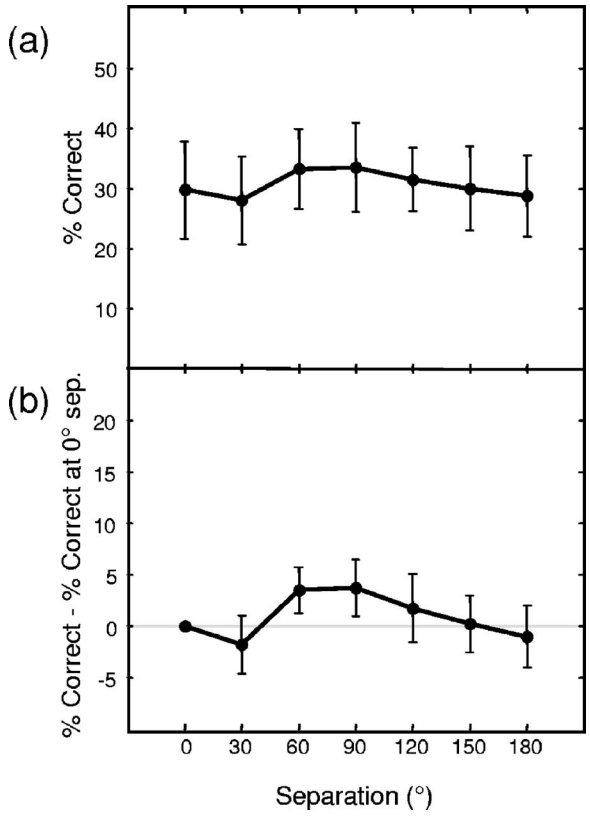


FIG. 3. (a) Mean percent correct scores for the different spatial configurations in Experiment 1B. Results are pooled across the six subjects and error bars indicate standard error of the mean. (b) Normalized percent correct scores for Experiment 1B. Normalization was carried out for each individual by subtracting the score for the colocated configuration. Results are pooled across the six subjects and error bars indicate standard error of the mean.

definition, the LR for the colocated pair is 0 dB. It can be seen that the LR increases as separation grows to 120°, but then decreases with further separation. This analysis suggests that the relative overall loudness of the two talkers at each ear can at least partially account for the behavioral results. Performance was positively correlated with the mean of the computed LRs across the two better ears ($r^2=0.90$, $p=0.001$).

Experiment 1B was designed to eliminate energy effects in order to confirm their role in the results of Experiment 1A and to determine whether there is any residual influence of perceived spatial separation of the two sources in a divided attention task. By using only ITDs in the spatial simulation, the level variations induced by the HRTF processing in Experiment 1A were removed (in essence, the LRs for these stimuli are fixed at 0 dB).

C. Results—Experiment 1B

The mean percent correct scores (and standard errors) across the six subjects for the different configurations are shown in Fig. 3(a). The curve is noticeably flatter than that obtained in Experiment 1A. Interestingly, overall performance is better (by approximately five percentage points) in Experiment 1B than in Experiment 1A. However, it is important to keep in mind that all subjects in Experiment 1B

TABLE I. Distribution of error types for incorrect trials in Experiment 1A (left column) and Experiment 1B (right column). Results are pooled across subjects and across the seven spatial separations.

Error type	Experiment 1A (% trials)	Experiment 1B (% trials)
Drop 1	46.1	43.8
Drop 2 (same)	11.0	8.0
Drop 2 (different)	8.4	5.3
Switch	5.7	6.7
Drop 3	1.2	0.7
Drop 4	0.2	0.0

completed Experiment 1A first. Thus, overall improvements in performance may reflect the fact that subjects were more experienced in the task in Experiment 1B.

As in Experiment 1A, percent correct scores for each subject at each reference location were normalized by subtracting the score in the colocated (separation 0°) configuration. In Fig. 3(b), the normalized data pooled across the six subjects are shown. There was no consistent change in performance across subjects with increasing separation (a repeated measures ANOVA found no significant effect $[F(6,30)=1.1,\ p>0.05]$).

D. Error analysis

To examine the different kinds of errors that listeners made in incorrect trials, errors were classified as one of six types: (i) "drop 1" errors where one keyword was incorrect; (ii) "drop 2 (same)" errors where both keywords from one of the two stimuli were incorrect; (iii) "drop 2 (different)" errors where one keyword from each source was incorrect; (iv) "switch" errors where the four keywords were reported but in incorrect pairs; (v) "drop 3" errors where three of the four keywords were not reported; and (vi) "drop 4" errors where none of the keywords were included in the response. An example of each kind of error is given in the Appendix.

Table I shows the average error rates pooled across subjects and spatial separations, for Experiments 1A (left column) and 1B (right column). In both experiments, no influence of the side of the talker (left versus right) was found for error types involving more error for one talker than the other (sign test, $p>0.05$). Thus performance was roughly equally affected by errors for the sources in the left and right hemifields. Table I shows that the majority of errors in both experiments were "drop 1" followed by smaller numbers of "drop 2 (same)," "drop 2 (different)," and "switch" errors.

To investigate which of the different error types were responsible for the changes in performance observed with increasing spatial separation in Experiment 1A, error rates were normalized by subtracting the individual error rates at 0° separation. Figure 4(a) shows these individually normalized error rates for the different error types, although "drop 3" and "drop 4" errors are not shown due to their extremely low occurrence. Error bars have been omitted for the sake of clarity. The error type that changes most dramatically with spatial separation is the "drop 2 (same)" error. Repeated measures ANOVAs on the different error rates confirmed that this was the only type to change significantly with spatial separation $[F(6,42)=5.3,\ p<0.001]$. This suggests that the main influence of spatial separation in Experiment 1A was to alter the probability that a listener would miss one of the two sources completely.

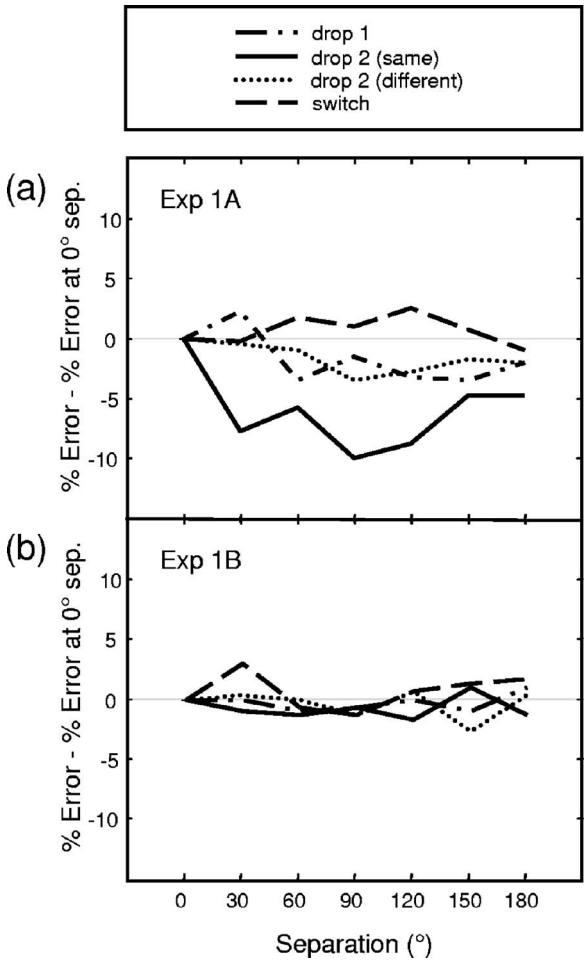


FIG. 4. Normalized error rates for incorrect trials in (a) Experiment 1A and (b) Experiment 1B. Normalization was carried out for each individual and each error type by subtracting the rate for the colocated configuration. Results are pooled across subjects. Error bars have been omitted for the sake of clarity, but statistical tests are described in the text.

Even though there was no significant change in overall performance with spatial separation in Experiment 1B, it is possible that the types of errors made depended on the separation. To investigate this possibility, error rates were normalized by subtracting the individual error rates at 0° separation as for Experiment 1A. Figure 4(b) shows these normalized error rates, although again "drop 3" and "drop 4" errors are not shown due to their extremely low occurrence. It appears that spatial separation has no consistent effect on the general pattern of errors, a conclusion confirmed using repeated measures ANOVAs on the different error types.

E. Discussion

These experiments examined the ability of listeners to track two simultaneous speech sources with minimal spectral overlap. When full HRTFs were used, performance improved with moderate separations (best performance for separations

in the range 90°–120°) but then decreased with further separation. This trend was eliminated in Experiment 1B, when spatial locations were simulated using only ITDs. Furthermore, the pattern of responses in Experiment 1A was correlated with the relative levels of the two sources at their acoustically better ears. These results strongly suggest that variations in the relative level of the two sources at the ears modulated the difficulty of the task in Experiment 1A and, ultimately, the accuracy of responses. Indeed, in the best spatial configuration (120° separation), the mean level ratio was 9 dB, meaning that each target source was 9 dB more intense in its better ear than the competing source. Interestingly, if listeners are able to use these monaural level ratios it suggests that they use the information at the two ears *independently* when tracking two sources in different hemifields.

A surprising finding in the error analysis was that there was no change in the frequency of "switch" errors as spatial separation was varied. Previous experiments using the Coordinate Response Measure corpus have found that this kind of error is quite common, especially when the talkers are of the same sex and in close spatial proximity (Brungart *et al.*, 2001). It would thus be expected that this type of error would be high in the 0° separation condition and would decrease with spatial separation. The fact that these error rates were relatively low reflects a robust ability to properly stream the two talkers in this task, i.e., to keep the two talkers as separate perceptual objects even when they are at the same spatial location. In addition, it is likely that other kinds of errors (particularly "drop 1" errors) limited performance to such an extent that the influence of "switch" errors was masked.

There was no evidence in this experiment of differential processing of sources in the left and right hemifields. This is in contrast to a large body of literature demonstrating that speech stimuli coming from the right are preferentially processed. When two speech sources are presented simultaneously to the left and right sides, a right-side advantage has been demonstrated when spatial location is determined by ear of presentation, external loudspeaker location, or even ITD only (see Darwin *et al.*, 1978 and Morais, 1978 for reviews). It is not clear why a left/right-side difference was not observed in the present study, although the differences reported in the literature are often quite small (especially when the sources are mixed at the two ears such as in our simulation rather than presented to different ears) and may be revealed only in larger data sets. Furthermore, it may be that these kinds of speech-specific effects occur less robustly for highly degraded speech signals such as those used in this experiment.

The results of Experiment 1B suggest that changes in the lateral positions of the two sound sources (produced by differences in ITD) did not influence performance on this task. One possibility is that perceived location estimates based on ITD alone are "diffuse," and that the two broad images associated with the two sources were not very clearly defined in this experiment. However, this seems unlikely based on recent data showing that when energetic masking is not the primary factor limiting performance, changes in perceived location give rise to similar improvements in selective listening tasks, regardless of which spatial cues produce the differences in perceived location (Shinn-Cunningham *et al.*, 2005). Given this, the results of Experiment 1 can be interpreted as evidence against the idea of a spotlight of auditory attention. In this task there appears to be no increased difficulty in following sources that are widely separated compared to those in close proximity. However, there is another potential explanation for this result. Despite efforts to minimize spectral overlap between the two sources, there was undoubtedly some remaining spectral overlap between the talkers, which may have been reduced by spatial separation. Thus, spatial separation may have rendered *each of the two sources* more intelligible (Gallun *et al.*, 2005; Shinn-Cunningham *et al.*, 2005). Moreover, spatial separation may improve segregation of the talkers. Both of these effects would work in opposition to any degradation in performance as sources fall outside the spatial spotlight of attention.

III. EXPERIMENT 2

A second experiment was performed to disentangle the possible opposing effects discussed above. The experiment was similar to Experiment 1, but different levels of noise were added to the speech signals in order to equate the difficulty of understanding the two talkers in the different configurations. In effect, the interference between the talkers was matched in all spatial configurations, allowing the efficiency of dividing attention to be compared directly. As there was no longer a need to reduce energetic masking, unprocessed speech was used in this experiment. Furthermore, in an effort to increase the "naturalness" of the setup, the two speech samples were chosen to have different voices and were presented from a pair of loudspeakers in a classroom. Furthermore, the locations were fixed in blocks rather than changing from trial to trial.

A. Methods

1. Subjects

Five paid subjects (ages 20–30) participated in Experiment 2. All had previous experience in auditory psychophysical studies, and one participated in Experiment 1.

2. Stimuli

Speech materials were taken from the same corpus used in Experiment 1 (described in Sec. II A 2). For each trial, two sentences spoken by different male talkers were chosen randomly from the corpus with the restriction that all keywords differed in the two sentences. No processing of the signals was done other than rms normalization, which set the relative levels of the two signals to 0 dB.

3. Procedure

Subjects were seated on a chair fitted with a headrest in a quiet, empty, carpeted classroom. The two sentences were presented from two matched Bose cube loudspeakers positioned 1 m from the listener at ear level. Stimuli were generated by PC-controlled Tucker-Davis Technologies hardware, amplified by a Crown amplifier, and sent to the loudspeakers via an eight-relay output module (KITSRUS

K108). Following each presentation, subjects indicated their responses by pressing the appropriate letter/number keys on a hand-held keypad.

Testing was done at three spatial separations about the midline: 10° (close), 90° (intermediate), and 180° (far). For each spatial configuration, subjects completed a noise calibration test as well as tests measuring single-task and dual-task performance.

The noise calibration test was designed to find the appropriate level of broadband white noise to add to the two loudspeakers such that when selectively attending to one of the two talkers, each individual listener could report keywords from the attended talker with an accuracy of 85%. Pairs of sentences were presented in noise at a variety of levels. Subjects were asked to report the color/number from the left talker only. No feedback was provided. Six noise levels were tested, with each noise level presented 25 times in a random order for a total of 150 trials in the test. The range of noise levels tested was fixed across listeners but was different for each spatial configuration. The levels, stated in decibels relative to the level of each of the speech signals, were in 6 dB steps between −20 and 10 dB (close), −14 and 16 dB, (intermediate), and −16 and 14 dB (far). For each configuration, a logistic function was fit to the raw data and the noise level corresponding to 85% performance was estimated.

In the experimental sessions that followed the noise calibration test, subjects were presented with one sentence plus noise (at the appropriate level) from each loudspeaker. Performance was measured in both single-task and dual-task conditions. Importantly, the stimulus set was identical in these two situations; only the task of the listener changed. In the single-task condition, subjects were asked to report keywords from either the left talker (as in the noise calibration test) or from the right talker (the talker to be attended was fixed within a test). Based on the noise calibration test, it was expected that performance in these tests would be approximately 85%. In the dual-task condition, listeners were asked to follow *both* talkers, and were required to respond with *two* color/number pairs (as in Experiment 1). Verbal instructions indicated that listeners should enter their response to the left talker followed by their response to the right talker (an instruction that was not given in Experiment 1), and responses were considered correct if both color/number pairs were reported correctly and in the correct order (left then right). No feedback was provided during testing.

Two tests of 100 trials each were completed for each of the three tasks conditions (left single task, right single task, dual task) in each of the three configurations. The resulting 18 tests were completed in a counterbalanced fashion (where the random ordering of the first nine tests was reversed in the second nine tests) to eliminate biases due to learning. All testing (noise calibration and single-/dual-task) was completed over six to seven sessions per subject.

4. Training

Before performing Experiment 2, subjects participated in three short training runs designed to familiarize them with the stimuli and task. In a training test, subjects were presented with pairs of sentences using one of the spatial configurations and were instructed to attend to the left talker, the right talker, or to both talkers. The combinations of configuration and task were randomly chosen for each listener, however over the course of the three training runs, each listener was exposed to each task (left, right, dual) and each spatial configuration (close, intermediate, far). Subjects indicated the color/number pair(s) they perceived and received correct-answer feedback via a written message on the hand-held response unit. A training run consisted of 100 trials.

TABLE II. Noise levels determined from the noise calibration tests for individual subjects as well as the mean across subjects. Levels are in decibels relative to the level of the speech signals, and correspond to 85% performance on the single task for each spatial configuration.

Subject	Close	Intermediate	Far
S1	−15.8	1.3	1.9
S2	−13.9	−1.7	−0.3
S3	−25.9	−3.0	−5.3
S4	−11.9	0.8	−2.1
S5	−22.3	−5.5	−4.8
mean	−16.9	−1.6	−2.1

B. Results

Table II lists the noise levels (in decibels relative to the level of the speech signals) determined from the noise calibration test for each individual in each configuration. The mean noise levels across subjects are also shown. As predicted, the task of following one of the talkers was most difficult for the close configuration so that a lower level of noise had to be added to reduce performance to 85% than for the other configurations. Note that the increase in noise level with spatial separation in these single-task trials is one estimate of the amount of spatial unmasking (in decibels) for these separations.

Mean percent correct scores across subjects are shown in Fig. 5. The upper lines represent performance in the "attend right" (dashed line) and "attend left" (dotted line) trials, respectively. These data confirm that subjects performed the single task with approximately 85% accuracy in all three configurations, although performance in the "attend right" condition was better than in the "attend left" on average (significant using a sign test, $p<0.05$). This may simply be an artifact of small differences in loudspeaker characteristics, or it may represent a "right-side advantage" as discussed in Sec. II E. Although no such asymmetry was found in Experiment 1, this may be because this subtle speech-specific bias is more prominent for clear unprocessed speech signals than for the processed stimuli used in Experiment 1. To confirm that subjects were able to distinguish the left from right loudspeaker successfully, the errors that resulted from subjects reporting the keywords from the wrong talker (i.e., reporting the right talker first or the left talker last) were counted. These errors were extremely rare (occurring in 0.9% of trials in total), confirming that listeners had no difficulty in judging the relative locations of the two talkers, even in the close configuration.

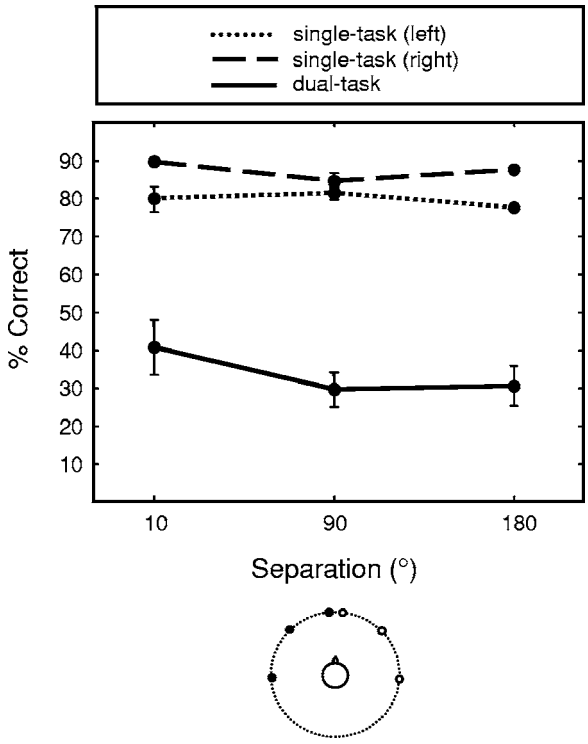


FIG. 5. Mean percent correct scores for the different spatial separations in Experiment 2 in both single-task (dashed and dotted lines) and dual-task (solid lines) trials. Single-task scores are shown separately for "attend left" and "attend right" conditions. Dual-task scores are based on trials in which both sources were correctly reported. Results are pooled across the five subjects and error bars indicate standard error of the mean.

The solid line represents performance in the dual-task trials. Performance was near 40% for close targets but near 30% for intermediate and far targets. This is slightly higher than the average performance in Experiment 1, probably due to the use of more natural and robust unprocessed speech signals.

To calculate a single measure of the "cost" associated with divided listening in the different spatial configurations, performance in the dual task was subtracted from average performance in the single task for each subject. These cost values are plotted for each subject in Fig. 6. A repeated measures ANOVA confirmed that there was a significant effect of spatial separation $[F(2,8)=9.6, p<0.05]$. Although post-hoc tests (pairwise comparisons with a Bonferroni correction) failed to reach significance, visual inspection of Fig. 6 makes it clear that the effect was due primarily to the fact that cost values were consistently smallest for the close configuration.

C. Error analysis

To further examine the different kinds of errors that listeners made in incorrect dual-task trials, errors were classified into different error types. The same error classification described in Experiment 1 was used, with the exception that drop errors associated with the left and right talkers were analyzed separately. Infrequent confusion between the left and right talkers, and combinations of a drop error and a left/right confusion, were pooled and classified as "other" errors.

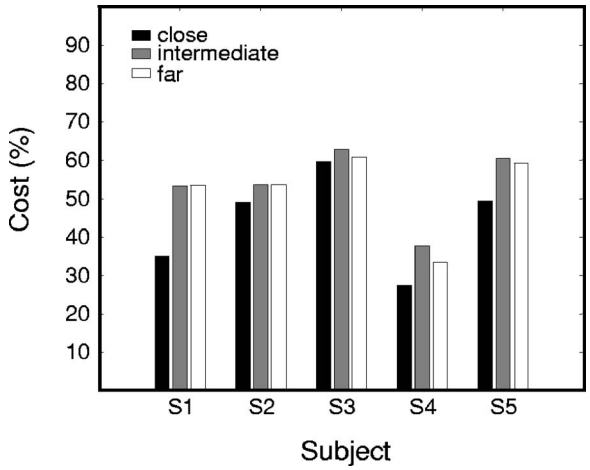


FIG. 6. Performance cost associated with responding to both targets (difference in percent correct between single-task and dual-task trials). Cost is shown for each subject in each of the three spatial configurations.

Table III shows the average error rates pooled across subjects and spatial separations. The majority of errors were of the kind "drop 1," as was the case in Experiment 1. For this experiment, however, a larger proportion of the "drop 1" errors were due to an error in reporting the right talker than the left. A smaller number of "drop 2" errors also occurred, with most of these involving errors in reporting both keywords from the right talker or one keyword from each talker. A small number of "switch" errors occurred, and these were approximately as frequent as they were in Experiment 1.

The fact that the error rates were much higher for the right source than for the left source—reversing any right-talker processing advantage observed in the single task—is most likely a result of the instructions given in this experiment. It seems that asking listeners to report the left talker first encouraged them to give *higher priority* to the left talker. In contrast, listeners in Experiment 1 were simply instructed to respond to both sources and no such asymmetry in performance was observed, presumably because attention was allocated more equally between the left and right talkers.

To investigate which of the error types drove the changes in performance observed with increasing spatial separation, error rates for the most common error types were normalized by subtracting the individual error rates at 10° separation (Fig. 7). Error bars have been omitted for the sake of clarity. This figure shows that most error types increased when spatial separation was increased, which explains the

TABLE III. Distribution of error types for incorrect trials in Experiment 2. Results are pooled across the five subjects and the three spatial separations.

Error type	% trials	
	Left	Right
Drop 1	15.6	25.8
Drop 2 (same)	2.9	6.1
Drop 2 (different)	5.3	
Switch	4.1	
Drop 3	0.8	
Drop 4	0.0	
Other	6.0	

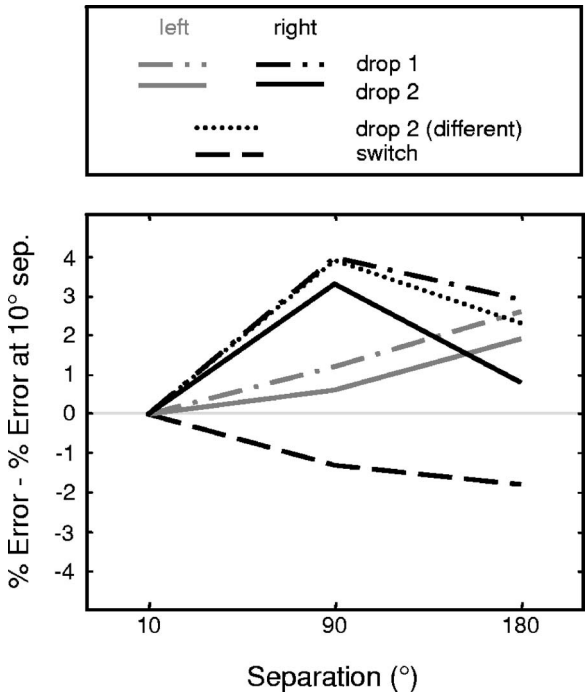


FIG. 7. Normalized error rates for incorrect dual-task trials in Experiment 2. Normalization was carried out for each individual and each error type by subtracting the rate for the "close" configuration. Results are pooled across the five subjects. Error bars have been omitted for the sake of clarity, but statistical tests are described in the text.

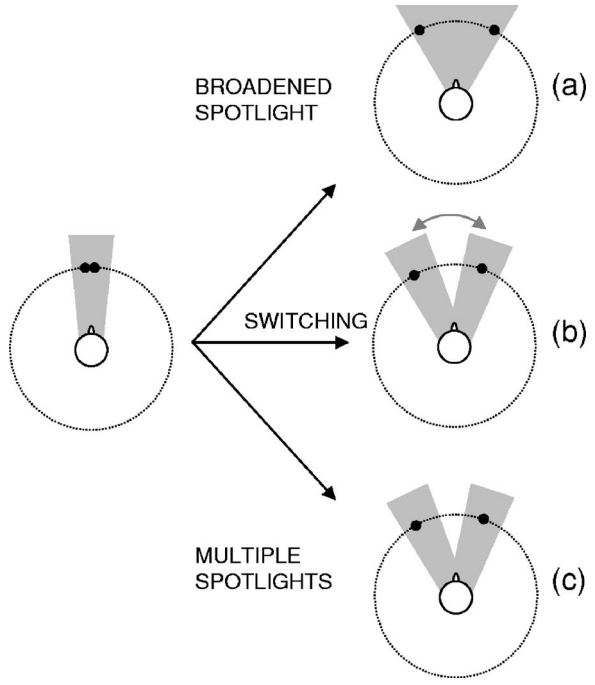


FIG. 8. If auditory spatial attention acts as a "spotlight" to enhance the perception of a relevant source and exclude others, it may be advantageous to have any sources of interest within a restricted region. Spatial separation of two targets of interest will require either (a) a broadening of the spotlight, (b) a strategy where attention is switched between the two sources, or (c) a splitting of the spotlight to form multiple spotlights.

overall decrease in performance at these larger separations. Interestingly, errors involving the right talker tended to be maximal at the 90° separation, and errors involving the left talker tended to be maximal at the 180° separation. Six repeated measures ANOVAs (one for each error type) were performed, and the effect of spatial separation reached significance only for the "drop 2 (right)" errors $[F(2,8)=6.1, p<0.05]$ (solid black line) and "drop 2 (different)" errors $[F(2,8)=9.8, p<0.05]$ (dotted black line). This suggests that the primary effect of spatial separation was to increase the probability that a listener would misreport the right talker completely or drop one keyword from each of the talkers. Note that "switch" errors *decreased* with spatial separation, consistent with spatial separation reducing confusion between elements of the two sources. Importantly, if these were not counted as errors (i.e., if the important criterion for "correctness" was simply how many of the four keywords were reported) the increase in dual-task cost as a function of spatial separation would be enhanced.

D. Discussion

These results show that there is a substantial performance cost associated with responding to two sentences compared to responding to one. This cost ranged between 30 and 60 %, depending on the subject and spatial configuration. This result is consistent with recent work (Gallun *et al.*, submitted) showing that performing two of the same kind of task (such as identifying two phrases simultaneously) results in a performance cost. Importantly, a number of factors may contribute to this cost. Not only are the attentional demands increased in the dual-task trials, but subjects must report four items instead of two and the memory load is increased. However, the main factor of interest in this study was not the magnitude or origin of the cost, but the impact of spatial configuration on the cost. It can be assumed that the memory load and general response demands were constant across spatial configurations. Furthermore, the design of the experiment meant that the accuracy with which each individual talker could be followed was constant across spatial configurations. Thus, we can be reasonably confident that changes in performance as a function of spatial separation are largely attributable to the distribution of attention between the competing talkers.

The main finding of the current experiment was that the cost associated with divided listening was smallest when the targets were in close spatial proximity. This finding is consistent with the idea that auditory spatial attention operates like a spotlight. As discussed in the Introduction, a spatially restricted attentional spotlight predicts that the simultaneous processing of two distant targets will be impaired. Several models have been put forward to describe this effect (see Fig. 8). It may be that the spotlight has to be broadened to simultaneously encompass two distant targets (the "zoom lens" model; see Eriksen and St. James, 1986) or switched rapidly between them, and in either case a decrease in processing efficiency may result. Alternatively, it may be that "multiple spotlights" of attention can be deployed (McMains and Somers, 2004; 2005), in which case there may be little or no cost in performance with spatial separation. Although the present experiments were not designed to address which of these mechanisms might apply in the auditory system, the error patterns give some preliminary indications. The fact

that the majority of errors involved one source or the other suggests that attention was focused on only one source at a time. This is further supported by the fact that when subjects gave one talker priority on the basis of response requirements (the left talker in Experiment 2), they tended to make many more errors on the other talker. These errors tended to increase with spatial separation, consistent with the spotlight being spatially restricted. On the other hand, there were also a reasonable number of errors involving both talkers, and their occurrence increased with spatial separation. This is more consistent with a "broadening" of the spotlight; listeners attended to both sources but with reduced processing efficiency as they stretched limited attentional resources to cover a greater area (Eriksen and St. James, 1986). The occurrence of some "switch" errors (although small in number) at all spatial separations lends some weight to this possibility, as including both sources within the same attentional channel could disrupt perceptual segregation of the two sources. Importantly, these patterns are also consistent with listeners employing a restricted spotlight of attention that is switched between the sources, resulting in information being missed from one source during instances when attention was on the other source. A final possibility is that listeners are able to attend to two distinct locations in space (using "multiple spotlights"), but that there is an overhead associated with splitting the attentional spotlight in this way. This kind of model has been described in vision, where subjects are able to allocate attention to two noncontiguous zones of the visual field simultaneously (Awh and Pashler, 2000; Müller *et al.*, 2003; McMains and Somers, 2004). Furthermore, a similar division of attention has been shown for frequency detection in the auditory system, where listeners primed with two different frequencies are better able to detect probe tones occurring at those frequencies and not at frequencies in between (Macmillan and Schwartz, 1975; Scharf, 1998).

Further experiments examining spatial configurations in which a distracting, nontarget stimulus is located in between two targets could tease apart whether spatial attention is split or broadened. In experiments on visual spatial attention it is assumed that intervening distracters degrade performance in cases where the spotlight is broadened because they are obligatorily attended, but not in cases where the spotlight is effectively split (McMains and Somers, 2005). Importantly, however, adding intervening distractors in an auditory task has more complex implications than it does in vision. In particular, acoustic interactions between targets and the distractor will affect the peripheral representation of the target stimuli in addition to influencing higher processes such as the spatial distribution of attention. If experiments determine that the auditory spotlight can be split, further experiments will be required to distinguish between a sustained splitting of attention and a rapid switching of attention between locations. This distinction is impossible to make with long speech stimuli such as those used here, where accuracy is based on information accumulated over tens to hundreds of milliseconds. Experiments involving much briefer stimuli may be required to address this issue (Miller and Bonnel, 1994).

One final point regarding divided listening deserves consideration. It was noted in Experiment 2 that error patterns differed for the left (presumably "higher-priority") and right (presumably "lower-priority") talkers. It is possible that these differences expose different mechanisms involved in reporting the two sources. Errors involving the left talker were less frequent than those involving the right talker, but increased steadily with spatial separation. This pattern is highly consistent with subjects focusing a spatial spotlight of attention on the left source, and attempting to share it between two locations to simultaneously process the right talker. However errors involving the right talker peaked at the intermediate configuration and dropped again for the far configuration. It may be that processing of the lower-priority source depends more on temporary storage in working memory than on the focusing of spatial attention. The increase in errors involving the right source peaks at 90° separation, which is the configuration that (on average) required the addition of the highest level of noise in this experiment (see Table II). Thus it is possible that the addition of noise to the speech signals degraded the sensory trace storing the lower-priority right talker, and/or influenced recall of the lower-priority keywords. Future experiments will be aimed at determining the influence of cognitive factors in multiple-talker listening, in particular how working memory storage and recall are affected by the spatial arrangement, signal-to-noise ratio, and task-mediated prioritization of simultaneous sources.

IV. SUMMARY AND CONCLUSIONS

These experiments examined the effect of spatial separation of two competing speech sources on the ability of listeners to report keywords from them both. In particular, the "spotlight" model of spatial attention was examined by testing the hypothesis that attending to both talkers would be more efficient if they occupied the same spatial region.

The results of Experiment 1A, which systematically examined the effect of spatial separation, show that spatial separation modulates performance predominantly due to changes in relative energy levels at the two ears. Importantly, it seems that energy ratios at the ears can be independently utilized when listening to two simultaneous talkers. When energy variations were eliminated (Experiment 1B), performance was relatively stable across different spatial configurations, suggesting that spatial attention may not act like a spotlight. Experiment 2 was performed to examine whether this lack of modulation of performance with spatial separation was the result of two opposing effects: a poorer ability to attend to both sources working in opposition to an improved separability and/or intelligibility of the competing signals. To do this, the intelligibility of the two sources was equalized across configurations by adding broadband masking noise and the "cost" of divided listening was measured. Results suggest that small separations result in a smaller performance cost when two talkers must be processed simultaneously, and are largely consistent with the spotlight model of spatial attention.

ACKNOWLEDGMENTS

This work was funded by the Office of Naval Research Grant No. N00014-04-1-0131. Frederick Gallun was supported by F32 DC006526 from NIDCD. Prity Bengani assisted with data collection. Two anonymous reviewers gave important feedback that greatly strengthened the paper.

APPENDIX: TABLE A1

TABLE A1. Examples of the different error types for one example pair of sentences. Incorrect words are indicated in bold, and substitutions between the two sentences are indicated in italics.

Error type	Example
Correct	Ready Baron go to blue one now Ready Charlie go to red two now
Drop 1	Ready Baron go to **green** one now Ready Charlie go to red two now
Drop 2 (same)	Ready Baron go to **green three** now Ready Charlie go to red two now
Drop 2 (different)	Ready Baron go to **green** one now Ready Charlie go to red **four** now
Switch	Ready Baron go to blue *two* now Ready Charlie go to red *one* now
Drop 3	Ready Baron go to **green three** now Ready Charlie go to **white** two now
Drop 4	Ready Baron go to **green three** now Ready Charlie go to **white four** now

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

Awh, E., and Pashler, H. (**2000**). "Evidence for split attentional foci," J. Exp. Psychol. Hum. Percept. Perform. **26**, 834–846.

Best, V., Ozmeral, E., Gallun, F. J., Sen, K., and Shinn-Cunningham, B. G. (**2005**). "Spatial unmasking of birdsong in human listeners: Energetic and informational factors," J. Acoust. Soc. Am. **118**, 3766–3773.

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.

Broadbent, D. E. (**1954**). "The role of auditory localization in attention and memory span," J. Exp. Psychol. **47**, 191–196.

Broadbent, D. E. (**1958**). *Perception and Communication* (Pergamon Press, London).

Bronkhorst, A. W. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acust. Acta Acust. 86, 117–128.

Brungart, D. S., and Rabinowitz, W. R. (**1999**). "Auditory localization of nearby sources. Head-related transfer functions," J. Acoust. Soc. Am. **106**, 1465–1479.

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

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects on the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Darwin, C. J., Howell, P., and Brady, S. A. (**1978**). "Laterality and localization: A "right ear advantage" for speech heard on the left," in: *Attention and Performance*, Vol. **VII**, edited by J. Renquin (Lawrence Erlbaum, Hillsdale, New Jersey), pp. 261–278.

Eriksen, C. W., and St. James, J. D. (**1986**). "Visual attention within and around the field of focal attention: A zoom lens model," Percept. Psychophys. **40**, 225–240.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Gallun, F. J., Mason, C. R., and Kidd, Jr., G. (**2005**). "Binaural release from informational masking in a speech identification task," J. Acoust. Soc. Am. **118**, 1614–1625.

Gallun, F. J., Mason, C. R., and Kidd, Jr., G. (submitted). "Task-dependent costs in processing two simultaneous auditory stimuli," Percept. Psychophys.

Kidd, G., Mason, C. R., Brughera, A., and Hartmann, W. M. (**2005**). "The role of reverberation in release from masking due to spatial separation of sources for speech identification," Acust. Acta Acust. 114, 526–536.

Macmillan, N. A., and Schwartz, M. (**1975**). "A probe-signal investigation of uncertain-frequency detection," J. Acoust. Soc. Am. **58**, 1051–1058.

Massaro, D. W. (**1976**). "Perceptual processing in dichotic listening," J. Exp. Psychol. Hum. Percept. Perform. **2**, 331–339.

McMains, S. A., and Somers, D. C. (**2004**). "Multiple spotlights of attentional selection in human visual cortex," Neuron **42**, 677–686.

McMains, S. A., and Somers, D. C. (**2005**). "Processing efficiency of divided spatial attention mechanisms in human visual cortex," J. Neurosci. **25**, 9444–9448.

Miller, J., and Bonnel, A. M. (**1994**). "Switching or sharing in dual-task line-length discrimination?," Percept. Psychophys. **56**, 431–446.

Mondor, T. A., and Zatorre, R. J. (**1995**). "Shifting and focusing auditory spatial attention," J. Exp. Psychol. Hum. Percept. Perform. **21**, 387–409.

Morais, J. (**1978**). "Spatial constraints on attention to speech," in: *Attention and Performance*, Vol. **VII**, edited by J. Renquin (Lawrence Erlbaum, Hillsdale, New Jersey), pp. 245–260.

Müller, M. M., Malinowski, P., Gruber, T., and Hillyard, S. A. (**2003**). "Sustained division of the attentional spotlight," Nature (London) **424**, 309–312.

Quinlan, P. T., and Bailey, P. J. (**1995**). "An examination of attentional control in the auditory modality: Further evidence for auditory orienting," Percept. Psychophys. **57**, 614–628.

Scharf, B. (**1998**). "Auditory Attention: The Psychoacoustical Approach," in: *Attention*, edited by H. Pashler (Psychology Press, London), pp. 75–117.

Shinn-Cunningham, B., and Ihlefeld, A. (**2004**). "Selective and divided attention: Extracting information from simultaneous sound sources," *Proc. Int. Conf. Auditory Display*, International Community for Auditory Display, Sydney, Austrailia.

Shinn-Cunningham, B. G., Ihlefeld, A., Satyavarta, and Larson, E. (**2005**). "Bottom-up and top-down influences on spatial unmasking," Acta. Acust. Acust. **91**, 967–979.

Spence, C. J., and Driver, J. (**1994**). "Covert spatial orienting in audition: Exogenous and endogenous mechanisms," J. Exp. Psychol. Hum. Percept. Perform. **20**, 555–574.

Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility," in: *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston), pp. 255–276.

Middle ear cavity and ear canal pressure-driven stapes velocity responses in human cadaveric temporal bones

Kevin N. O'Connor and Sunil Puria
Department of Otolaryngology-Head and Neck Surgery, Stanford University, 300 Pasteur Drive, Stanford, California 94305, Department of Mechanical Engineering, Mechanics and Computation Division, Stanford University, Durand Building, Stanford, California 94305, and Department of Veterans Affairs, 3801 Miranda Avenue, Palo Alto, California 94304

(Received 30 March 2006; revised 5 June 2006; accepted 9 June 2006)

Drive pressure to stapes velocity (V_{st}) transfer function measurements are collected and compared for human cadaveric temporal bones with the drive pressure alternately on the ear canal (EC) and middle ear cavity (MEC) sides of the tympanic membrane (TM), in order to predict the performance of proposed middle-ear implantable acoustic hearing aids, as well as provide additional data for examining human middle ear mechanics. The chief finding is that, in terms of the V_{st} response, MEC stimulation performs at least as well as EC stimulation below 8 kHz, provided that the EC is unplugged. Plugging the EC causes a reduced response for MEC drive below 2 kHz, due to a corresponding reduction of the pressure difference between the two sides of the TM. Between 8 and 11 kHz, the MEC drive transfer functions feature an approximately 17 dB drop in magnitude below the EC drive case, the cause of which remains unknown. The EC drive transfer functions reported here feature significantly less magnitude roll-off above 1 kHz than previous studies [with a slope of −2.3 vs −6.7 dB/octave for Aibara *et al.*, Hear. Res. **152**, 100–109 (2001)], and significantly more phase group delay (134 vs 62 μs for Aibara *et al.*). © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2221414]

PACS number(s): 43.66.Ts, 43.64.Ha, 43.64.Bt [BLM] Pages: 1517–1528

I. INTRODUCTION

The idea of implanting a hearing aid device in the middle ear cavity (MEC) has been proposed as early as 1970, on the basis that such an implant might in some cases improve acoustic performance and/or patient acceptability over conventional hearing aids (Goode, 1970). In contrast to conventional acoustic hearing aids, which inject sound into the ear canal (EC), an implanted acoustic hearing aid would inject sound directly into the MEC, where it would drive the tympanic membrane (TM) from the surface facing the middle ear rather than the surface facing the EC. The class of *acoustic* middle ear implantable hearing aids is different from the class of middle ear implantable hearing aids that vibrate the ossicular chain through direct mechanical contact (Zenner *et al.*, 1998; Hough *et al.*, 2001; Fisch *et al.*, 2001; Lenarz *et al.*, 2001; Chen *et al.*, 2004). The former may be more advantageous in that they may be easier to deploy.

While a large number of published studies contain human temporal bone measurements of EC pressure (P_{ec}) to stapes velocity (V_{st}) transfer functions (V_{st}/P_{ec}), as driven with an EC sound source (Goode *et al.*, 1994; [1] Voss *et al.*, 2000; Aibara *et al.*, 2001; Gan *et al.*, 2001; Chien *et al.*, 2006), none that the authors are aware of report transfer functions from MEC pressure (P_{mec}) to stapes velocity (V_{st}/P_{mec}), as driven with a MEC sound source. Such measurements could serve as a basis for evaluating the relative effectiveness of injecting sound into the MEC as opposed to the EC, to predict the performance of an implantable acoustic hearing aid device.

In this study, measurements of P_{ec}, P_{mec}, V_{st}, cochlear promontory velocity (V_{pr}), and noise floor were collected on four human cadaveric temporal bones. For most of these quantities, separate measurements were taken on the same preparations with sound injected into the EC and sound injected into the MEC, as well as with both an intact and a separated incudo-stapedial joint (ISJ). Additional measurements were made with MEC drive and an unplugged EC, to better replicate the most likely operational scenario for an implanted hearing aid. This collection of measurements allows several different types of transfer function to be directly compared between EC and MEC drive cases, for the same bones, thus providing ample means to examine the potential effectiveness of an implantable acoustic hearing aid.

In addition to the applicability to hearing aid research, these measurements provide opportunities to examine the mechanisms affecting sound transfer through the middle ear. In conventional circuit-type models of the middle ear, the transfer of sound through the TM and ossicular chain is implied to originate with the pressure difference between the two sides of the TM (Onchi, 1961; Zwislocki, 1962; Shaw and Stinson, 1983; Shera and Zweig, 1991; Peake *et al.*, 1992; Shera and Zweig, 1992; Rosowski and Merchant, 1995; Puria and Allen, 1998). By comparing the pressure difference to V_{st} transfer functions for EC and MEC drive cases, one can gauge the extent to which the pressure difference alone can account for the measured V_{st} response.

II. METHODS AND MATERIALS

A. Postmortem materials

A total of seven human temporal bones were used for this project. Measurements were taken on only four of the

bones, as one was used to develop the experimental procedure and two were damaged during preparation. The four bones on which measurements are reported are referred to as E1, E2, E3, and E4. E2 was extracted from a 71 year old white male, E3 from a 68 year old black male, and E4 from a 76 year old black male. Age and race information for E1 is not available. All bones were obtained from the Palo Alto VA Hospital Pathology Department, were refrigerated after extraction, and were frozen within four days.

B. Preparation of temporal bones

Each bone used in this project was initially taken from the freezer and allowed to thaw in a refrigerator for 1–3 days before preparation. Once thawed, soft tissue was removed from the bone and the EC was cleared of ear wax and other debris. Next, a surgical drill was used to open the MEC via the mastoid, exposing the stapes and round window through the facial recess. Care was taken to preserve the stapedial tendon, incus buttress, and any other structures affecting ossicular motion. The EC was not modified in this study, thus preserving the anatomical relationship between its medial portion and the tympanic membrane (see also Murugasu *et al.*, 2005). This constitutes a departure from previous methods in which the EC was removed and replaced by a tube perpendicular to the plane of the TM to allow visibility of the umbo (Goode *et al.*, 1994; Voss *et al.*, 2000; Gan *et al.*, 2001; Aibara *et al.*, 2001; Puria, 2003).

The drilled bone was then enclosed in a layer of latex, using two finger cots cut from an examination glove, in order to seal in moisture. To facilitate later identification, the latex regions covering the EC and opened MEC were marked with permanent ink. The covered bone was then placed in a mold with the markings facing up, and a plaster-like substance, KERR Model Stone (www.kerlab.com), was poured into the mold surrounding most of the bone, but leaving the marked regions partially exposed. Once the model stone finished hardening, the bone was removed from the mold and the material covering the EC and MEC was removed to allow access.

To supply sound to the MEC, a 2.5 cm polyethylene tube, obtained by removing the foam from an EARLINK eartip (www.earlink.com), was affixed to one edge of the MEC with Durelon carboxylate dental cement (www.3mespe.com), such that visibility of the footplate and round window was not impaired. During experiments, the outer end of this tube was connected to a sound source. An ER7-14C microphone probe tube (Etymotic Research; www.etymotic.com) was also cemented to the wall of the MEC, with the tip as close to the posterior stapes crus as was practical, typically within 2–5 mm.

Excess moisture was removed from the MEC using suction and small pieces of twisted tissue paper in order to make the surfaces suitable for applying retroreflective microbead targets. For all bones, targets were placed on the center of the stapes footplate, and for E2–E4 they were also applied to the bony promontory between the round and oval windows, as well as to the center of the round window. The opened MEC was then sealed by building up the sides with either DAP

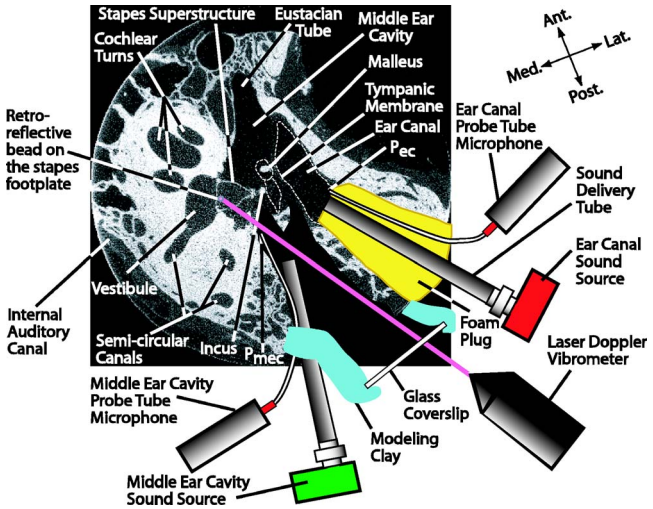


FIG. 1. (Color online) Diagram of the experimental setup superimposed on a micro-CT scan of a human middle ear cross section (30.72 mm scan diameter, with 15 μm in-plane and out-of-plane resolution). Sound is injected into either the EC or the drilled but closed MEC, and the pressure is measured on both sides of the TM using probe tube microphones. The stapes footplate velocity is measured using a laser Doppler vibrometer (LDV).

FunTak (www.dap.com) for E1, or modeling clay for E2–E4, and placing a glass coverslip on top. Figure 1 shows a drawing of the experimental setup superimposed upon a micro CT cross-sectional image of a human temporal bone (not one of the bones used in this experiment).

C. Measurement system

All measurements were made using SYSid (www.sysid-labs.com), which consists of software (version 6.5) and accompanying hardware (an Ariel DSP-16+ board). SYSid produces an output signal that can be used to drive a sound source, and synchronously measures the magnitude and phase of two input signals: one from a laser Doppler vibrometer (LDV) and another from a probe tube microphone in this case.

The output of the SYSid board was connected to a power amplifier (for E1 a Yamaha P2075, and for E2–E4 a Crown D-75) with a gain setting of 0 dB, and the output of the power amplifier was connected to a switcher box (for E2–E4) capable of directing the signal to either of two identical sound sources. The sound sources both consist of a #2955 Knowles hearing aid receiver (www.knowles.com) with a 200 Ω series resistor. An ER 3-04 plastic coupling nipple (Etymotic Research) was placed on the output port of the receiver and glued in place with epoxy. One sound source was used to drive the EC by connecting the coupling nipple to a foam EARLINK ear tip that was inserted into the EC, and the other was used to drive the MEC by connecting it to the sound delivery tube that was glued to the wall of the MEC. For E2–E4, switching between EC and MEC drive was a simple matter of toggling a switch. For E1, however, because only one driver was used, it was manually moved between the EC and MEC. The unused sound delivery tube was then plugged.

For E2–E4, pressure measurements were made using two ER-7c probe tube microphones (Etymotic Research),

one measuring pressure in the MEC and the other measuring pressure in the EC. Using a sharp instrument, the EC probe tube was fed through the foam of the eartip used for driving the EC, such that its open end could measure the pressure in the small air space between the foam plug and the TM (Fig. 1). Due to lack of visibility, the distance between the tip of the EC probe tube and the umbo could not be determined directly, but is estimated to be less than 1 cm based on how deeply the foam plug was inserted into the EC (the distance between the probe tube and the umbo was often limited by the natural angulation of the TM with respect to the EC). The two microphones were connected to another switch that was in turn connected to one of the inputs to the SYSID board (channel B). For E1, because only one ER-7c microphone was available at the time, it was manually moved between the EC and MEC by disconnecting and reconnecting to the appropriate probe tubes.

For E2–E4, velocity measurements were made using a Polytec (www.polytec.com) HLV-1000 LDV attached to a Carl Zeiss table-mounted operating microscope (www.carlzeiss.com) with the Polytec HLV MM2 accessory. The HLV MM2 contains a prism and joystick that allow the laser spot to be manipulated while being viewed through the microscope. The LDV was set to the 25 (mm/s)/V range option, the 30 kHz low-pass filter corner frequency, and the high-pass filter turned off. The output was connected to the other input of the SYSID board (channel A). For E1, a Polytec OFV302 LDV system was used instead, with the 25 (mm/s)/V range setting, 20 kHz low pass filter setting, and the fast acquisition mode setting. A separate floor-mounted Carl Zeiss microscope was used for aiming the laser beam. The LDV controller was connected to a Yamaha power amplifier (P2075) with a gain setting of 20 dB.

The sound stimulus for all measurements was a sine wave swept over 175 logarithmically spaced frequency points from 0.1 to 25 kHz, with a constant SYSID output Voltage set to either 0.1 or 0.5 V for E1 (for EC-driven and MEC-driven measurements, respectively), and 0.5 V for E2–E4. For each input frequency, a 4096 point response signal was averaged 15 times.

D. Calibration

Probe tube microphone and LDV magnitude and phase calibration procedures have been previously described (Aibara *et al.*, 2001; Puria, 2003).

E. Measurements

For each ear, four quantities were measured:

(1) V_{st}—stapes footplate velocity.
(2) P_{ec}—ear canal pressure.
(3) P_{mec}—middle ear cavity pressure.
(4) V_{pr}—velocity of the bony cochlear promontory (to estimate the artifact level).

Each of these was measured for three drive conditions:

(1) EC drive—sound delivered to the EC via a foam ear tip inserted into the EC.
(2) MEC drive—sound delivered to the MEC, with the EC ear tip still in place, to enable direct comparison with EC drive under the same experimental conditions.
(3) MEC drive (no plug)—like MEC drive, but with an open EC, to mimic the most common scenario for operation of an implanted hearing aid.

In addition, for each drive condition, the four quantities were measured with:

(1) An intact incudo-stapedial joint.
(2) A cut incudo-stapedial joint.

Noise floor measurements were taken for some of the variables by disconnecting the input signal but otherwise performing the measurements in the normal manner. Round window velocity measurements were also taken for E2–E4, but those results are not presented in this paper.

Due to various factors, not all measurements are available for all ears. In some cases measurements were added to the protocol after the first ear, and in the case of E2 the cut ISJ data were deemed invalid because the incus and stapes were later observed to still be touching. Table I lists which ears contribute to the data set for each measurement.

TABLE I. Ears contributing to the data set for each measurement type.

Meas	Intact ISJ	Cut ISJ	Noise
	1. EC drive		
V_{st}	E1, E2, E3, E4	E1, E3, E4	E2, E3, E4
P_{ec}	E1, E2, E3, E4	E1, E3, E4	E2, E3, E4
P_{mec}	E1, E2, E3, E4	E1, E3, E4	E2, E3, E4
V_{pr}	E2, E3, E4	E3, E4	⋯
	2. MEC drive, plugged ear canal		
V_{st}	E1, E2, E3, E4	E1, E3, E4	E2, E3, E4
P_{ec}	E1, E2, E3, E4	E1, E3, E4	E2, E3, E4
P_{mec}	E1, E2, E3, E4	E1, E3, E4	E2, E3, E4
V_{pr}	E2, E3, E4	E3, E4	⋯
	3. MEC drive, open ear canal		
V_{st}	E2, E3, E4	E1, E3, E4	E2, E3, E4
P_{ec}	E2, E3, E4	E3, E4	E2, E3, E4
P_{mec}	E2, E3, E4	E1, E3, E4	E2, E3, E4
V_{pr}	E3, E4	E3,E4	⋯

F. Analysis

SYSid magnitude and phase measurements, in response to tones, were loaded into MATLAB (www.mathwoks.com), where custom scripts were used to calculate transfer functions, compute statistics, and generate plots.

Plots were generated to display un-normalized pressure and velocity measurements as well as pressure-to-velocity transfer functions, among others. In the case of the un-normalized EC-driven measurements for E1, the input voltage was 0.1 V instead of 0.5 V, so for comparison purposes the affected E1 magnitudes were scaled by a factor of 5. Individual curves from E1 to E4 were plotted along with summary statistics consisting of the mean, mean+standard error of the mean ("SEM," which is defined as the standard

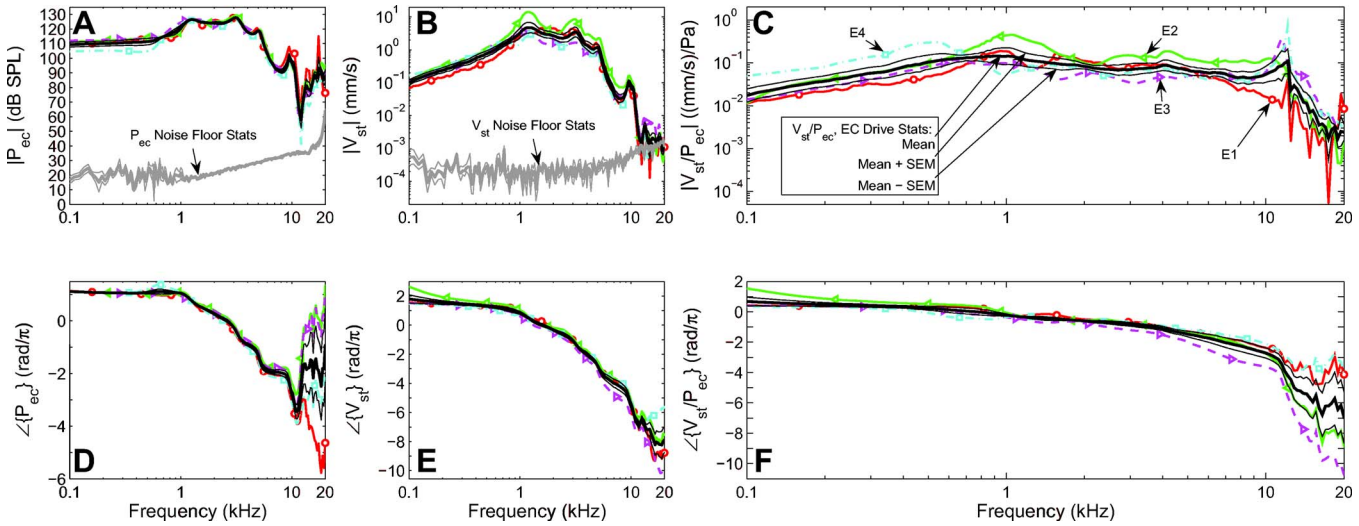


FIG. 2. (Color online) Measurements from four ears with sound injected into the EC, of EC pressure, P_{ec} [(A) and (D)], stapes velocity, V_{st} [(B) and (E)], and the pressure to velocity transfer functions, V_{st}/P_{ec} [(C) and (F)]. Panels (A)–(C) depict the magnitudes, and (D)–(F) the respective phase measurements. In addition to the four individual measurements, mean and standard error statistics are also shown in each panel, along with the noise floor in panels (A) and (B). The mean pressure magnitude lies above the noise floor by more than 6 dB for all frequencies (A), but the mean velocity falls within 6 dB of the noise above 12 kHz (B). Consequently, the "valid range" for the mean velocity and transfer function measurements [(B), (E), (C), and (F)] is considered to be 0.1–12 kHz.

deviation divided by the square root of the number of measurements), and mean-SEM. Since the individual plots are displayed on a log scale, the summary statistics were generated on a log scale and plotted so they would correctly summarize the mean and spread of the individual curves as plotted (Puria, 2003).

When plotting velocity data, two scaling adjustments were performed. First, the measured peak velocity was scaled by $1/\sqrt{2}$ to convert it to rms velocity. Since pressures were calibrated using a sound level calibrator that was already set to rms units, pressures did not need to be scaled. Second, when pointing the laser beam at the stapes footplate, the beam was not perpendicular to the plane of the footplate. Defining θ as the angle between the normal vector of the footplate and the incident laser beam, the velocity was scaled by $1/\cos(\theta)$ in order to approximate the actual velocity in the perpendicular direction.[2] The mean θ (55°) previously reported for the footplate (Aibara *et al.*, 2001) was used for all ears for all stapes velocity measurements. The promontory velocity, V_{pr}, was not adjusted for the laser angle because the laser tended to be nearly perpendicular to the promontory surface being measured.

In one plot, a comparison is made between current measurements and a past measurement (Aibara *et al.*, 2001). Because the Aibara *et al.* velocity measurements were originally reported as peak velocities, the transfer function magnitude was scaled by $1/\sqrt{2}$ as described earlier.

All group delays, τ, are computed as

$$\tau = -\frac{\Delta\phi}{\Delta\omega}, \tag{1}$$

where $\Delta\omega$ denotes a frequency interval (in units of radians/second) over which the phase mean appears as an approximately straight line when plotted on a linear frequency axis. $\Delta\phi$ denotes the phase change (in units of radians), over the $\Delta\omega$ frequency interval, of the straight line that best fits the phase mean over that interval.

III. RESULTS

A. Comparisons of measurements from individual ears

Figures 2 and 3 present the drive pressure (P_{drive}), V_{st}, and V_{st}/P_{drive} for EC and MEC drive cases, respectively. Each of the two figures is broken into six panels, labeled (A)–(F). The upper panels [labeled (A)–(C)] contain magnitudes, and the lower panels [labeled (D)–(F)] contain phases. The first column [panels (A) and (D)] contains the magnitude and phase of the drive-pressure (P_{ec} in the case of EC drive, and P_{mec} in the case of MEC drive), the second column [panels (B) and (E)] contains the magnitude and phase of the resulting stapes velocity, and the third column [panels (C) and (F)] contains the magnitude and phase of the V_{st}/P_{drive} transfer function. Curves from the four individual ears are shown with different line styles and markers to distinguish them from one another. Additionally, for all panels, the mean is plotted as a thick black line as well as the mean+SEM (a thin black line above the mean), and the mean−SEM (a thin black line below the mean). The pressure and velocity noise floor magnitude measurements are also shown, in panels (A) and (B), respectively, as mean and mean±SEM statistics. It should be noted that noise floor measurements were only taken on three of the ears (E2–E4: see Table I for a full listing of which ears were used for the various other measurements described in the following). The pressure magnitude is shown in units of dB SPL, the rms velocity magnitude in mm/s, and the transfer function magnitude in (mm/s)/Pascal. All phase measurements are in units of radians normalized by π.

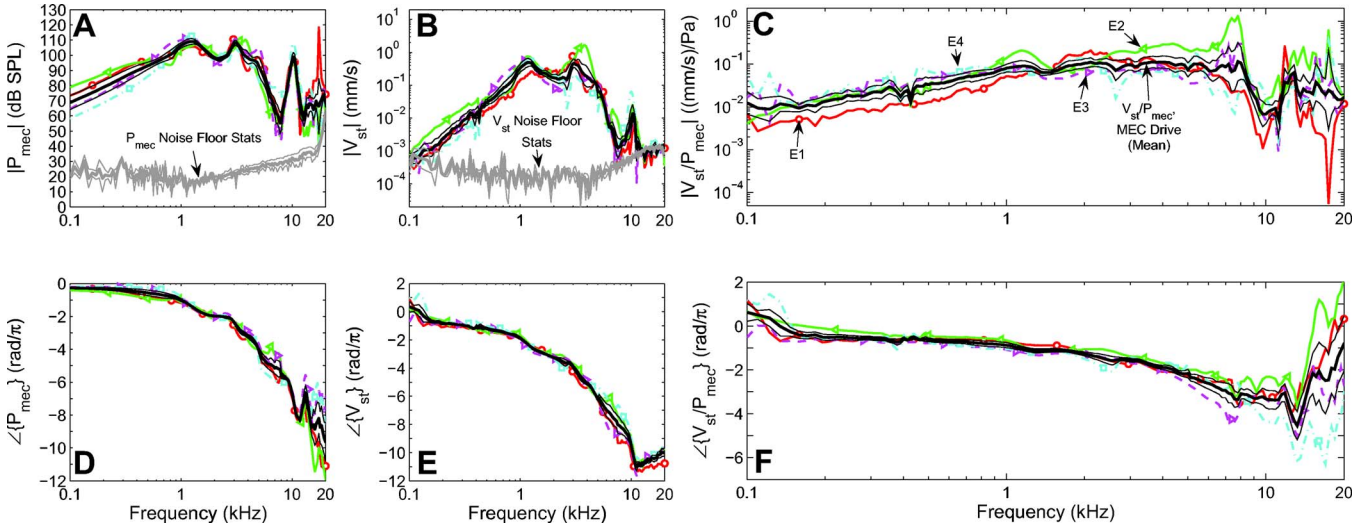


FIG. 3. (Color online) Measurements from four ears with sound injected into the MEC instead of the EC. MEC pressure, P_{mec} [(A) and (D)], stapes velocity, V_{st} [(B) and (E)], and pressure to velocity transfer functions, V_{st}/P_{mec} [(C) and (F)] are shown, in the same manner described for Fig. 2. As in Fig. 2, the mean pressure magnitude lies above the noise by more than 6 dB for the entire frequency range [(A)], but here the velocity and transfer function means are considered valid in the 0.14–11 kHz range [(B), (E), (C), and (F)].

1. EC drive (individual ears)

Referring to Fig. 2, for the EC drive case, the drive-pressure magnitude mean [panel (A)] peaks to 128 dB SPL at 3.1 kHz. The mean pressure magnitude is at least 19 dB above the noise mean (at 12 kHz), and rises as high as 107 dB above the noise (at 1.2 kHz), averaging 81 dB above the noise over the full frequency range.[3] The magnitude and phases of the four ears all look similar in shape, with the exception of the phase above 12 kHz.

The velocity magnitude curves [panel (B)] exhibit somewhat more variability in level than the pressure curves, with the maximum velocity of the mean curve reaching 4.7 mm/s (at 1.2 kHz). The mean velocity falls as low as −9.1 dB relative to the noise mean (at 20 kHz), and rises as high as 94 dB above the noise mean (at 1.4 kHz). Using as the criterion for validity that the velocity mean be at least 6 dB above the noise mean, the velocity mean is valid over the 0.1–12 kHz range, over which the mean averages 59 dB above the noise mean.[3]

The transfer function magnitude [panel (C)] exhibits a fair amount of variability across ears. Generally speaking, the transfer functions all rise for low frequencies, peak, proceed as a relatively flat line, then fall off above 10 kHz, after peaking in some cases. The low frequency rising slope is 6 dB/octave for the mean (based on the 0.1–0.6 kHz range), with the individual ears each having roughly the same slope, in spite of their differences in level. From 1 to 8 kHz, the mean shows a slope of −2.3 dB/octave. E3 and E4 feature a prominent peak at 11.5 and 12.15 kHz, respectively. The SEM over the 0.1–12 kHz valid range varies from 1.2 (at 2.2 kHz) to 7 dB (at 11.5 kHz), with a mean of 2.6.

The low-frequency phase is similar for E1, E3, and E4 [panel (F)], and E1, E2, and E4 for high frequencies up to 12 kHz. In the 1.1–11.1 kHz region the group delays are 122 (E1), 133 (E2), 185 (E3), and 97 μs (E4) with the mean phase having a group delay of 134 μs over the same range. The phase SEM, over the valid range, varies from 0.05 (at 1.9 kHz) to 0.45 rad/π (at 11.8 kHz), with a mean of 0.16.

2. MEC drive (individual ears)

MEC drive pressure magnitudes [Fig. 3, panel (A)] exhibit more variability than the EC drive case at low frequencies. Instead of a flat low frequency slope, as seen in the EC drive case, the mean rises with a slope of 10.1 dB/octave, from 0.1 to 1 kHz. The maximum of the mean curve is 109 dB SPL (at 1.24 kHz), versus the maximum of 128 dB SPL for the EC drive mean curve. The pressure magnitude falls off rapidly between 5 and 7.8 kHz, and peaks suddenly around 10.3 kHz before falling again to another local minimum around 13.2 kHz. The P_{mec} mean varies between 17.1 (at 19.5 kHz) and 97.8 dB (at 1.12 kHz) above the noise mean, and averages 65.4 dB above the noise over the full frequency range.

The velocity magnitude [panel (B)] begins and ends close to the noise floor, with the mean varying from −4.4 (at 13.2 kHz) to 76.6 dB (at 1.12 kHz) relative to the noise mean. Over the valid 0.14–11 kHz range, the velocity mean averages 46.8 dB above the noise mean. The low frequency slope of the mean is 17 dB/octave, over the 0.2–1 kHz range. The magnitude falls off sharply around 5.2 kHz and stays down except for the local peak around 10.3 kHz (corresponding to the peak observed in the pressure measurements).

The MEC drive transfer function magnitude [panel (C)] exhibits a rising slope of 6.2 dB/octave over the 0.2–1 kHz range, followed by a slightly positive slope of 0.16 dB/octave over the 1–8 kHz range, and a drop of around 24 dB from 8 kHz to the upper limit of the valid frequency range (11 kHz). Overall the SEM over the valid 0.14–11 kHz range varies from 1.06 (at 1.5 kHz) to 9.5 dB (at 7.6 kHz), with a mean of 2.9.

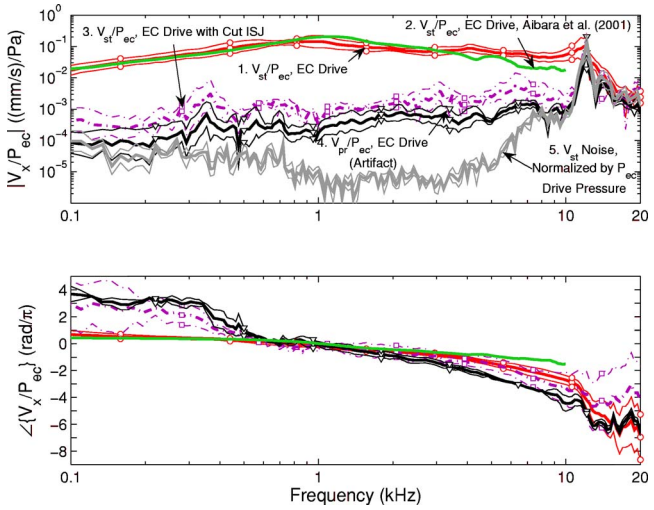


FIG. 4. (Color online) Ear canal-driven pressure to velocity transfer functions (V_x/P_{ec}, where x is either "st" for stapes or "pr" for promontory) for the current study with intact ossicles (1), for the 2001 Aibara *et al.* study (2), and for the current study with a cut ISJ (3), along with the artifact measurement of the cochlear promontory velocity normalized by the drive pressure (4), and the velocity noise floor magnitude normalized by the drive pressure (5). The top panel contains the magnitude and the bottom panel the phase measurements. As mentioned in Fig. 2, (1) is considered valid below 12 kHz. (3) and (4) are considered valid in the 0.1–9.8 and 0.27–7.6 kHz ranges, respectively.

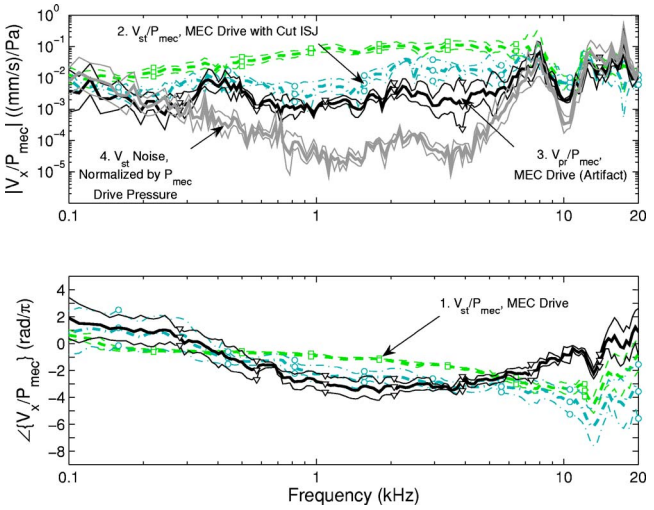


FIG. 5. (Color online) MEC-driven pressure to velocity transfer functions (V_x/P_{mec}, where x is either "st" for stapes or "pr" for promontory), all for the current study, with intact ossicles (1) and a cut ISJ (2), as well as an artifact measurement (3), and velocity noise normalized by the drive pressure (4). As mentioned in Fig. 3, (1) is considered valid in the 0.14–11 kHz range. (2) and (3) are considered valid in the 0.25–7.6 and 0.31–7 kHz ranges, respectively, with the frequencies immediately surrounding 10 kHz also considered valid.

The MEC drive transfer function phase [panel (F)] exhibits reasonable agreement across ears within the valid frequency range. The group delay values in the relatively linear 1.1–7 kHz region are: 168 (E1), 125 (E2), 222 (E3), and 134 μs (E4), with the mean curve having a group delay of 162 μs. The SEM ranges from 0.044 (at 0.46 kHz) to 0.45 rad/π (at 7.2 kHz), with mean of 0.14 over the valid 0.14–11 kHz frequency range.

B. Comparisons between intact ISJ, cut ISJ, artifact, and scaled noise measurements

Figures 4 and 5 again feature V_{st}/P_{drive} transfer function statistics for EC and MEC drive cases, respectively, but this time they are plotted alongside a number of additional curves. Measurements of velocity artifact normalized by drive pressure are shown, for which the artifact velocity was measured by aiming the laser beam at retroreflective targets placed on the cochlear promontory, a bony surface rigidly connected to the rest of the temporal bone. This measurement illustrates how a presumably rigid portion of the bone vibrates in the presence of the stimulus, thus serving as a useful baseline for evaluating the amount of movement observed for the stapes footplate. Measurements of stapes velocity in the absence of ossicular coupling are also shown, measured after cutting the ISJ, remeasuring stapes velocity, and normalizing by the drive-pressure. Finally, the V_{st} noise floor data have also been normalized by the ordinary (intact ISJ) drive pressure measurements (magnitude only) to give an appropriately scaled noise reference. In Fig. 4, the V_{st}/P_{ec} mean from the Aibara *et al.* (2001) study, based on measurements from 11 ears, is shown for purposes of comparison. The SEM curves for the Aibara *et al.* data were not available for the magnitude measurements and are not shown. One methodological difference between these two studies is that the Aibara *et al.* study used an artificial EC, whereas the current study used the natural EC.[4] With the exception of the Aibara *et al.* curve, all measurements shown in Figs. 4 and 5 are statistical summaries (mean and mean±SEM) of the available measurements from the individual ears. The normal V_{st}/P_{drive} case is based on four measurements, and the cut ISJ, promontory (artifact), and noise statistics are based on three measurements (see Table I for details).

1. EC drive (intact and cut ISJ, artifact, and scaled noise)

The mean magnitude of V_{st}/P_{ec} with an intact ISJ for the current study closely matches the Aibara *et al.* mean up to 0.7 kHz (Fig. 4). The Aibara *et al.* mean then proceeds to rise above that of the current study and peak at a slightly higher frequency: 0.21 (mm/s)/Pa at 1.15 kHz for Aibara *et al.* versus 0.14 (mm/s)/Pa at 0.92 kHz for the current mean. From 1.5 to 10 kHz the Aibara *et al.* mean curve has a slope of −6.7 vs −2.3 dB/octave for the current mean over the 1–8 kHz range. The Aibara *et al.* phase closely resembles the current mean phase up to around 2 kHz, above which it becomes apparent that the Aibara *et al.* mean curve exhibits a significantly smaller group delay: on the order of 62 (based on the 2–10 kHz region) vs 134 μs (based on the 1.1–11.1 kHz region) for the current case.

The cut ISJ mean lies above the scaled noise mean by at least 6 dB over the 0.1–9.8 kHz range, and as much as 59.5 dB (at 3 kHz), and averages 36.1 dB above the mean over this range. The artifact mean lies above the scaled noise mean by at least 6 dB from 0.27 to 7.6 kHz, and as much as 49.4 dB (at 1.4 kHz), averaging 29.1 dB above the mean over this range. Over the shared 0.1–9.8 kHz valid range, the current V_{st}/P_{ec} (intact ISJ) mean ranges from 20.8 (at

6.2 kHz) to 53.9 dB (at 1 kHz) above the cut ISJ mean, and averages 35.5 dB above it. Over the shared 0.27–7.6 kHz valid range, the V_{st}/P_{ec} (intact ISJ) mean ranges from 30 (at 7.4 kHz) to 63.5 dB (at 0.48 kHz) above the artifact mean, and averages 46.6 dB above it. Finally, over the shared 0.27–7.6 kHz valid range, the cut ISJ mean lies above the artifact mean, by at least 3.9 dB (at 2.5 kHz) and as much as 17.9 dB (at 0.37 kHz), and averages 10.9 dB above the artifact mean.

The artifact phase mean exhibits a rather high apparent group delay of 232 μs (based on the 2–7.6 kHz range), considerably higher than that of the current V_{st}/P_{ec} (intact ISJ) mean of 134 μs (based on the 1.1–11.1 kHz range). The cut ISJ phase mean exhibits a group delay of 172 μs (based on the approximately linear 1.1–9.8 kHz range), though the individual curves (not shown) exhibit highly variable slopes, which account for the large SEM for low and high frequencies.

2. MEC drive (intact and cut ISJ, artifact, and scaled noise)

Unlike the EC drive case in which cut ISJ and artifact measurements are normalized by the drive pressure in the EC, in the MEC drive case (Fig. 5) they are normalized by the drive pressure in the MEC. The cut ISJ and artifact means lie above the scaled noise mean by more than 6 dB between 0.25 and 7.6 kHz, and 0.31 and 7 kHz, respectively. The cut ISJ mean lies above the scaled noise mean by as much as 65.2 dB (at 3.5 kHz), and averages 36.4 dB above it. The artifact mean lies above the scaled noise mean by as much as 43.8 dB (at 3.5 kHz), and averages 26.1 dB above it. Both cut ISJ and artifact means become valid again in the vicinity of 10 kHz, as the signal to noise ratio (SNR) improves with a peak in the drive pressure (see Fig. 3). The cut ISJ mean lies above the artifact mean by between −4.2 (at 0.42 kHz) and 24.5 dB (at 3.7 kHz), averaging to 12 dB above the artifact mean over the shared valid 0.31–7 kHz range. The V_{st}/P_{mec} (intact ISJ) mean lies above the cut ISJ mean by between 4.4 (at 7.6 kHz) and 28.8 dB (at 0.92 kHz), averaging 15.8 dB above the cut case mean (in the shared 0.25–7.6 kHz valid range), and lies above the artifact by between 4.2 (at 0.43 kHz), and 41.1 dB (at 0.98 kHz), averaging 28.4 dB above the artifact mean (in the shared 0.31–7 kHz valid range).

C. V_{st}/P_{drive} (EC drive, MEC drive, and MEC drive with an open EC)

Figure 6 shows the V_{st}/P_{drive} transfer functions for the three drive cases (EC drive, MEC drive, and MEC drive with an open EC). Since the EC drive and MEC drive statistics have already been described and shown separately (in Figs. 2–5), the focus here is on how they differ from one another and how they relate to the MEC drive, open EC case.

The EC and MEC drive magnitudes exhibit similar mean low frequency slopes (6 and 6.2 dB/octave, respectively), but the MEC drive magnitude is on average 9.4 dB lower from 0.1 to 0.7 kHz. The two curves begin to overlap around 1.7 kHz, continuing to around 6 kHz. Above 6 kHz,

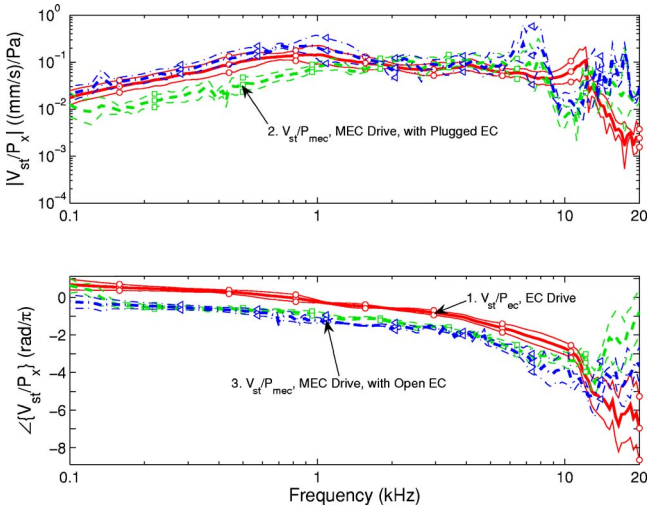


FIG. 6. (Color online) Comparison of drive pressure to stapes velocity transfer functions (V_{st}/P_x, where x is either "ec" for EC, or "mec" for middle ear cavity) for EC drive (1), MEC drive with the ear plug still in the EC (2), and MEC drive with the ear plug removed from the EC (3). As mentioned in Figs. 2 and 3, (1) and (2) are considered valid in the 0.1–12 and 0.14–11 kHz ranges, respectively. (3) is considered valid in the 0.1–11.5 kHz range.

the EC drive case rises to a peak while the MEC drive case dips 17.4 dB below the EC drive curve around 10.3 kHz. Within the valid frequency range for MEC drive (0.14–11 kHz, compared to 0.1–12 kHz for EC drive) the chief differences in magnitude between EC and MEC drive appear below around 1.5 kHz, and above 8 kHz. The MEC drive, open EC case, valid between 0.1 and 11.5 kHz (limited by the velocity, based on measurements not shown), more closely resembles the EC drive case than the MEC drive case does, up to around 6 kHz, above which it exhibits some of the same features as the MEC drive case. The MEC drive, open EC case mean lies slightly above the EC drive case (by 2.6 dB on average), with overlapping SEM, up to around 1.8 kHz. The overall SEM for the MEC drive, open EC case ranges from 0.2 (at 0.65 kHz) to 12.4 dB (at 7.4 kHz), with a mean of 3.2, based on the valid 0.1–11.5 kHz frequency range.

Both MEC drive phases look similar below around 7 kHz, above which the SEM for the open EC case spreads and the mean drops below the closed EC case. The EC drive case is generally around π radians higher than the MEC drive cases below 8 kHz, implying that the TM is moving the ossicles in the opposite direction, as expected. The mean phase group delay of the MEC drive, open EC case is 157 μs (based on the approximately linear 1.1–11.1 kHz region), compared with 134 and 162 μs for the EC and MEC drive cases, respectively. The MEC drive, open EC phase SEM ranges from 0.01 (at 1.4 kHz) to 0.88 rad/π (at 10.3 kHz), with a mean of 0.21.

D. V_{st} normalized by the pressure difference between the two sides of the TM

In Fig. 7, V_{st} for EC and MEC drive cases is normalized by the pressure difference across the TM from the drive side to the opposite side in each case. For the EC drive case this

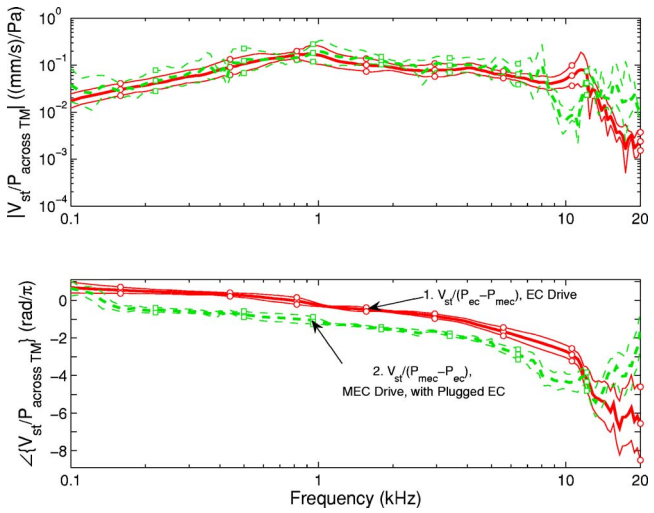


FIG. 7. (Color online) Comparison of pressure difference across the TM to stapes velocity transfer functions for EC drive (1) and MEC drive with the ear plug still in the EC (2). In each case, the pressure difference is calculated by subtracting the pressure on the opposite side of the TM from the pressure on the side being driven. (1) and (2) are considered valid in the 0.1–12 and 0.14–11 kHz ranges, respectively.

corresponds to $V_{st}/(P_{ec}\text{-}P_{mec})$ with P_{ec} being the drive-pressure, and for the MEC drive case it corresponds to $V_{st}/(P_{mec}\text{-}P_{ec})$ with P_{mec} being the drive-pressure. Again only statistics are shown, based on four ears in both cases.

The mean magnitude for the EC drive case looks very similar to that of V_{st}/P_{ec}, whereas the MEC drive case changes significantly from V_{st}/P_{mec} when normalized by the pressure difference, to closely resemble the EC drive case below 8 kHz. The valid frequency ranges, limited again by the velocity measurements, are 0.1–12 kHz for EC drive and 0.14–11 kHz for MEC drive. The magnitude means show relatively close agreement from 0.14 kHz up to around 8.5 kHz, with the MEC drive mean on average 1.4 dB higher than the EC drive mean and the SEM of the two cases overlapping for the entire 0.14–8.5 kHz range. Above 8.5 kHz the behaviors differ, with the MEC drive mean dipping downward, and the EC drive mean rising to a peak. The phases are roughly parallel and around π radians apart below 8 kHz.

IV. DISCUSSION

A. Comparisons to previous studies

Ear canal-driven stapes velocity transfer functions (or the equivalent information presented in the form of stapes displacement measurements) have been reported by several groups, including Goode *et al.* (1994), Voss *et al.* (2000), Gan *et al.* (2001), Aibara *et al.* (2001), and Chien *et al.* (2006). The authors are not aware of any earlier reports of MEC-driven transfer functions.

The most notable differences observed between the current study and previous studies concern the slope of the high-frequency magnitude and the group delay from the phase of the V_{st}/P_{ec} transfer functions. All other studies show the high-frequency transfer function with a significantly steeper high-frequency magnitude roll-off than the current study, and phase group delay of the current study tends to be somewhat larger than that of the other studies.

Generally speaking, the mean SEM values for the current study are around twice as high as those of the other studies. Much of the difference observed in SEM values can be attributed to the lower sample size for the current experiment (N=4) compared to the other studies (where N ranges from 6 to 18). If the current standard deviation stayed the same, but the sample size matched that of the other studies, the SEM values would become quite comparable to (and often better than) those of the other studies. While the smaller sample size of the current study warrants some degree of caution in interpreting the results, the authors feel that the inter-ear variation of the current study is sufficiently similar to that of the other studies to justify consideration of these results.

B. Block diagram of the experimental setup

Figure 8 contains a block diagram representing the experimental setup with a few simplifying assumptions, much of which is based on the model presented by Peake *et al.* (1992) and by Shera and Zweig (1992). Switch SW1 controls the state of the EC: in position (A), the EC sound source, represented by pressure source P_{S1} with series impedance Z_{S1}, has been inserted in the EC. This is true for the EC drive and MEC drive with plugged EC cases. In position (B), the EC is open (the plug is removed), so the radiation impedance of the EC is attached instead, represented by Z_{Rad}. This applies to the MEC drive, open EC case. The pressure in the EC, P_{ec}, is depicted as being the pressure across either of these impedance blocks, and the positive terminal of P_{ec} is directly connected to the positive terminal of P_{tm}, and represents the pressure at the TM on the EC side. These connections represent one of the simplifying assumptions of the model, namely: the pressure inside the EC is assumed to be uniform in space. In reality, the pressure at the output of the sound source [terminal (A) of SW1] is at a slightly different point in space from the tip of the probe tube where P_{ec} is measured, and the pressure on the EC side of the TM is not a single point at all, but could refer to the pressure at some point anywhere over the entire surface of the TM. Considering that the typical dimensions of the EC airspace between the sound source and the TM are fairly small, and that the actual spacing between these points is estimated to be less than 1 cm, spatial variation of pressure in the EC is believed to be minimal below 8 kHz or so.

The EC pressure is shown as being the sum of the pressure difference between the two sides of the TM, P_{tm}, and the pressure inside the MEC, P_{mec}. P_{mec} is the pressure measured by the MEC probe tube, and in the diagram it is equated to the pressure at the output of the MEC sound source, represented as P_{S2} with series impedance Z_{S2}, and the pressure on the MEC side of the TM, represented as the negative terminal of P_{tm}. As in the EC case, equating these three points is a simplifying assumption since in reality the points are separated in space. The MEC airspace as it is prepared in this experiment is considerably larger overall than the enclosed

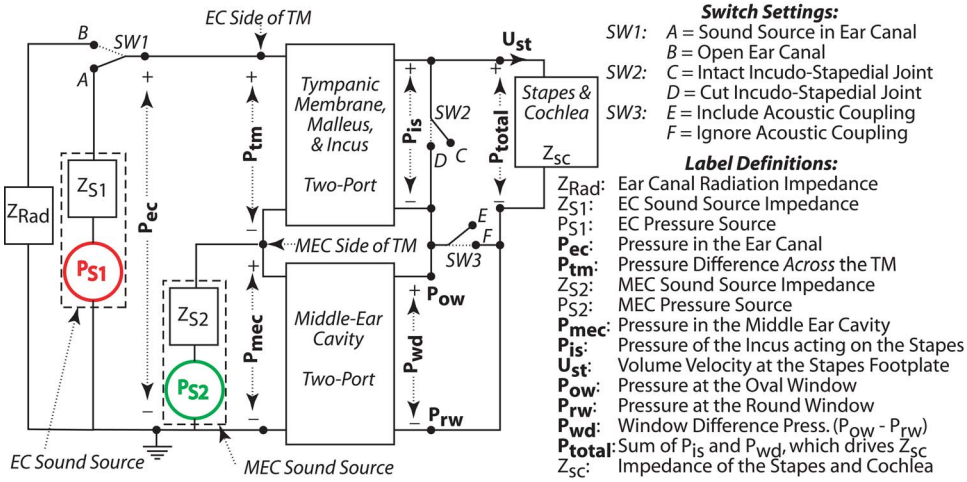


FIG. 8. (Color online) Block diagram approximately representing the experimental setup in terms of acoustic variables, impedance and two-port blocks, pressure sources, switches, and various interconnections, with mechanisms included for both ossicular and acoustic coupling of sound to the cochlea. Switch SW1 controls whether the EC is plugged or open, P_{ec} and P_{mec} are pressures in the EC and MEC, respectively, and P_{tm} is the pressure difference across the TM from the EC to the MEC side. P_{tm} drives the two-port block representing the TM, malleus, and incus, which produces output P_{is}, the equivalent pressure due to the incus pushing on the stapes. P_{mec}, in turn, drives a two-port representing the MEC, which produces output P_{wd}, the pressure difference between the oval and round windows. The total pressure across the block representing the stapes and cochlea, Z_{sc}, is composed of a part due to the ossicular chain (P_{is}), and a part due to acoustic coupling (P_{wd}). Dividing the sum of the ossicular and acoustic coupling pressure components by Z_{sc} produces the total volume velocity of the stapes, U_{st}, which is approximately equal to V_{st} times the area of the stapes footplate. Switch SW2 controls whether the ISJ is cut or intact, and SW3 allows the acoustic coupling contribution to either be included or ignored, thus enabling a further simplification of the model.

EC air space, but the typical separations between the TM, probe tube, and output of the sound source tend to be on the order of 1 cm. The pressure across the TM, P_{tm}, drives a two-port block representing the tympanic membrane, malleus, and incus. The output of the block, P_{is}, represents the effective pressure that the incus exerts on the stapes due to their mechanical connection. In the case where the incudo-stapedial joint is separated, SW2 is in the (D) position, to indicate that this effective termination is zero. The positive terminal of P_{is} is connected to an impedance block representing the stapes and cochlea, Z_{sc}, and the volume velocity of the stapes footplate, U_{st}, is shown entering this block, where U_{st} is approximately equal to V_{st} times the area of the stapes footplate.

The pressure measured in the MEC, P_{mec}, is connected to a two-port block representing the MEC airspace. At the output of this block, one terminal represents the pressure at the oval window, P_{ow}, and the other terminal represents the pressure at the round window, P_{rw}. This is the one place in the model where spatial variation of pressure is not ignored. The pressure difference between the oval and round windows is labeled as P_{wd}. A switch, SW3, is included to allow the effects of acoustic coupling to be included (E position) or ignored (F position). In the E position, the total pressure across the Z_{sc} block, P_{total}, is the sum of P_{is} and P_{wd}, such that $U_{st}=(P_{is}+P_{wd})/Z_{sc}$. In the F position, however, the pressure at the oval and round windows is forced to be equal, and P_{total} is made to equal only the pressure due to the ossicular chain, P_{is}. For cases where acoustic coupling is expected to be small, such as at low frequencies where spatial pressure variation is expected to be small, ignoring acoustic coupling in the model (SW3=F) greatly simplifies the analysis, because the MEC and TMOC blocks can then be viewed by the left-hand side of the circuit in terms of their input impedances, Z_{tmoc} and Z_{mec}.

The acoustic coupling mode of the model (SW3=E) is useful to account for the presence of volume velocity with the ISJ cut (SW2=D). Without acoustic coupling, the model would predict zero volume velocity in this case, which is not true, as Figs. 4 and 5 illustrate.

No mechanisms are included to account for bone conduction effects, and for higher frequencies spatial pressure variation effects in the EC and MEC may become significant.

1. EC drive in terms of the block diagram

In the case of EC drive, SW1 is in the A position and the pressure source P_{S1} is active while the other source P_{S2} is turned off and acts like a short circuit. P_{S1} sets up a pressure signal that transfers over Z_{S1} to set up the pressure measured in the EC, P_{ec}. P_{ec} is then divided into P_{tm}, the pressure across the TM, and P_{mec}, the pressure measured in the MEC. Ignoring acoustic coupling (SW3=F), P_{tm} causes the TM, malleus, and incus to move, which causes the incus to exert a force on the head of the stapes, viewed in terms of the equivalent pressure as P_{is}. The only role of the MEC airspace in this case (where SW3=F) is as part of a voltage divider in establishing P_{tm}. Given equivalent input impedances Z_{tmoc} for the TM, ossicles, and cochlea (TMOC) without acoustic coupling, and $Z_{mec}\|Z_{S2}$ for the MEC airspace including the MEC sound source in parallel, the relationship between the pressure driving the TMOC block, P_{tm}, and P_{ec} is then

$$\frac{P_{tm}}{P_{ec}} = \frac{(P_{ec} - P_{mec})}{P_{ec}} = \frac{Z_{tmoc}}{(Z_{tmoc} + Z_{mec} \| Z_{S2})}. \quad (2)$$

Thus, if $|Z_{tmoc}| \gg |Z_{mec} \| Z_{S2}|$, Eq. (2) becomes approximately equal to 1, with P_{mec} becoming negligibly small and P_{tm} becoming approximately equal to P_{ec}, such that essentially all of the pressure in the EC is used to drive the TMOC part of the circuit. When P_{mec} is relatively small, the actual P_{wd} can also be expected to be relatively small, so the actual acoustic coupling component is not expected to be significant when $|Z_{tmoc}| \gg |Z_{mec} \| Z_{S2}|$.

2. MEC drive with a plugged EC, in terms of the block diagram

In this case SW1 is in the A position since the EC plug and sound source is still inserted in the EC. P_{S2} is active, but P_{S1} is not active and should be treated as a short circuit. Again ignoring acoustic coupling, P_{S2} sets up the input pressure which transfers over Z_{S2} to establish the pressure measured in the MEC, P_{mec}. This pressure is then divided between $-P_{tm}$ (negative because P_{mec} is connected to its negative terminal), which drives the TMOC block, and P_{ec}, which is the pressure across the impedance of the EC sound source, Z_{S1}. Here, P_{tm} relates to P_{mec} as follows:

$$\frac{P_{tm}}{P_{mec}} = -\frac{(P_{mec} - P_{ec})}{P_{mec}} = -\frac{Z_{tmoc}}{(Z_{tmoc} + Z_{S1})}. \quad (3)$$

Note that, in Eq. (3), if Z_{S1} is equal to or larger than Z_{tmoc} in magnitude and their angles are the same, then the magnitude of P_{tm}/P_{mec} will be less than 0.5, indicating that the pressure transferred to the ossicular chain is at least 6 dB lower than the measured drive pressure.

3. MEC drive with an open EC, in terms of the block diagram

In this case, the only difference from the MEC drive, closed EC case is that SW1 is in the B position, indicating that the EC is open and is represented by its radiation impedance, Z_{Rad}. Ignoring acoustic coupling again, one finds that

$$\frac{P_{tm}}{P_{mec}} = -\frac{(P_{mec} - P_{ec})}{P_{mec}} = -\frac{Z_{tmoc}}{(Z_{tmoc} + Z_{Rad})}. \quad (4)$$

C. Analysis of EC and MEC drive differences below 8 kHz

Figure 7 shows that the differences between EC drive and MEC drive (plugged EC) V_{st}/P_{drive} transfer functions essentially disappear below 8 kHz when V_{st} is normalized by the pressure difference across the TM instead of by the drive pressure. Figure 9 shows how these pressure differences (P_{ec}-P_{mec} for EC drive, or P_{mec}-P_{ec} for MEC drive) relate to the drive pressure for each of the three drive types. In the magnitude panel, one can see that EC drive stays within 2 dB of 0 dB all the way to 11 kHz, suggesting that for EC drive the pressure in the EC is a good indicator of the pressure driving the TM and ossicles. In the case of MEC drive with the open EC (valid in the 0.14–11.6 kHz range), the

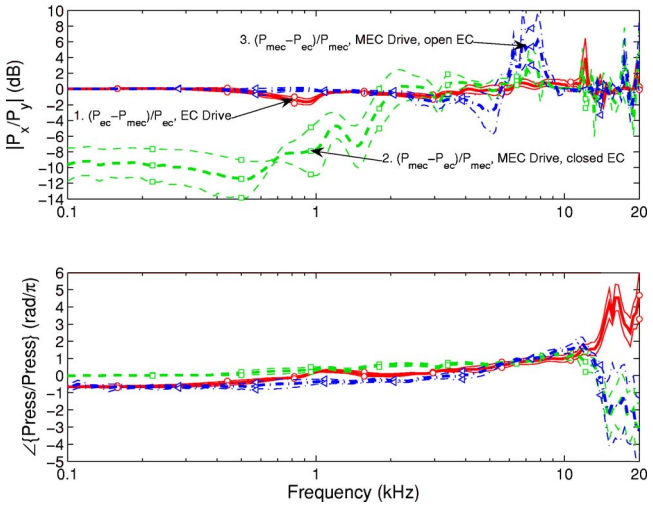


FIG. 9. (Color online) Comparison of transfer functions relating the drive pressure to the pressure difference across the TM (P_x/P_y, where P_x is either "P_{ec}-P_{mec}" or "P_{mec}-P_{ec}," and P_y is either "P_{ec}" or "P_{mec}," for EC and MEC drive, respectively) for EC drive (1), MEC drive with the ear plug still in the EC (2), and MEC drive with the ear plug removed from the EC (3). As in Fig. 7, the pressure differences are found by subtracting the pressure on the opposite side of the TM from the pressure on the side being driven. (1) and (2) are considered valid over the full frequency range, but (3) is only considered valid in the 0.14–11.6 kHz range.

mean magnitude also stays quite close to 0 dB up to 4 kHz. MEC drive with the plugged EC, however, shows a mean drop of 10 dB or more up to 0.5 kHz, with a gradual rise bringing it close to the other curves by 2 kHz. This indicates that by plugging the EC, the pressure difference across the TM is reduced, which in turn can explain the reduced response below 2 kHz for the V_{st}/P_{mec} (plugged EC) transfer function. The earlier observation that normalizing V_{st} by the pressure difference across the TM can account for the observed difference between the transfer functions (Fig. 7) is consistent with this. The difference between the plugged MEC drive case and the other cases below 2 kHz can be interpreted in terms of the impedances represented in the block diagram, as described earlier. Because the magnitude of P_{tm}/P_{mec} for the MEC drive, plugged EC case is smaller than the other two cases below 2 kHz, the above-presented equations indicate that, below 2 kHz, the magnitude of Z_{S1} should be larger than the magnitude of the parallel combination of Z_{mec} and Z_{S2} [Eqs. (2) and (3)], and should also be larger than the magnitude of Z_{Rad} [Eq. (4)]. Clearly, having high impedance on the opposite side of the TM can reduce its ability to move. Above 2 kHz, the three curves generally stay fairly close together (with the exception of 6–8 kHz and above 11 kHz). The V_{st}/P_{drive} transfer functions in turn show similarities up to 6 kHz, above which the MEC drive curves show different behavior from the EC drive case.

D. EC and MEC drive differences above 8 kHz

The differences between MEC drive (closed canal) and EC drive transfer functions below 8 kHz appear to be well-explained by the difference in pressure drop across the tympanic membrane due to the plugged EC, indicating that, *below* 8 kHz, *the sound transfer mechanism from the TM to the cochlea behaves the same whether the sound source is on the*

EC side or the MEC side. For higher frequencies, however, a dip centered around 10 kHz appears in the MEC drive measurements whether or not they are normalized by the pressure difference (Figs. 6 and 7), suggesting that something else is responsible for the observed dip.

The 7–13 kHz region for MEC drive features drops and peaks in pressure that result in only some of the velocity readings having an adequate SNR for some frequencies (Fig. 3). The drop, peak, and following drop of the MEC drive pressure magnitude at 7.8, 10.3, and 12.8 kHz appear to be the result of half, three-fourths, and full wave resonances inside the 2.5-mm-long plastic sound delivery tube used to drive both the EC and MEC, based on the similarity between the tube length and the computed partial wavelengths at those frequencies. As a result of this up and down behavior in pressure, the velocity follows a similar pattern, such that at the low points in pressure, the velocity SNR is low (but still greater than 6 dB in the mean below 11 kHz), but at the height of the peak (10.3 kHz) the velocity SNR is sufficient for all ears.

Referring to the 10 kHz vicinity in Figs. 6 and 7, where the velocity SNR is not in doubt, the transfer function indicates that the MEC drive measurements consistently see around 17 dB less velocity than the EC drive measurements on average. The magnitude drop is consistent across all ears for MEC drive, but additional study is required to determine the nature of this effect.

E. Group delay

The group delay, as computed from the phase of a transfer function [Eq. (1)], is typically interpreted as indicating the amount of time needed for the output signal of an input-output system (the numerator of the transfer function) to appear once the input signal (the denominator of the transfer function) has entered the system. In the case of the V_{st}/P_{drive} transfer functions (Fig. 6), the group delay could theoretically be interpreted as the amount of time needed for the drive pressure to travel from the microphone tip to the TM, to be converted by the TM into mechanical motion of the umbo, and to be transferred by the ossicular chain into motion of the stapes footplate. In the current experiment, the mean V_{st}/P_{drive} group delay numbers are 134, 162, and 157 μs for EC, MEC, and MEC (open EC) drive conditions, respectively. Viewed in terms of the distance that sound could propagate in air over that amount of time (assuming the speed of sound in air is 34 500 cm/s), the equivalent distances are 4.6, 5.6, and 5.4 cm. Given that the relevant dimensions of the temporal bone preparations used in this study (i.e., the combined distances from the drive-side microphone tip to the TM surface, from the TM surface to the umbo, and from the umbo to the stapes footplate) are considerably smaller (see Fig. 1) than these computed distances, it appears that the effective propagation speed from the EC to the stapes footplate is less than the speed of sound in air. The group delay from the Aibara *et al.* study, 62 μs, translates into a propagation distance in air of 2.1 cm, which is considerably less than the current study, and a group delay of 112 μs from Voss *et al.* translates into 3.9 cm, which is only a little less than the current study. Ear canal pressure to vestibular pressure transfer functions reported in Puria *et al.* (1997) and Puria (2003), in turn, have group delays of 98 and 117 μs, respectively, which translate to 3.4 and 4.0 cm propagation distances in air. With the possible exception of the Aibara *et al.* data, these group delay numbers appear to suggest that some phenomena within the tympanic membrane and/or ossicular chain cause an effective slowing-down of the transferred waves.

The EC drive artifact and cut ISJ measurements (Fig. 4) exhibit rather high group delays, 232 and 172 μs, respectively, which translate into equivalent air propagation distances of 8 and 6 cm. It is unclear why the phases of these measurements exhibit such high apparent group delays, but they are most likely the result of a measurement artifact, perhaps exacerbated by low signal levels, rather than an actual indication of wave propagation delay.

V. SUMMARY

Injecting sound into the MEC produces a stapes velocity response similar to that from injecting sound into the EC, below 8 kHz, provided that the EC is unobstructed. Blockage of the EC reduces the response to MEC stimulation below 2 kHz, due to a resulting reduction of the pressure difference between the two sides of the tympanic membrane. In the vicinity of 10 kHz, the response to MEC stimulation appears to drop by as much as 17 dB below the EC-driven response.

The current set of stapes velocity transfer function measurements consistently exhibit less high frequency magnitude roll-off than earlier studies, and significantly more phase group delay than earlier studies.

ACKNOWLEDGMENTS

Work supported in part by the Stanford Department of Otolaryngology, by Grant No. DC005960 from the NIDCD of NIH, and by a gift from Implanted Acoustics, Inc. The authors thank Dr. Toshiki Maetani for help with temporal bone dissections.

[1]Goode *et al.* (1994) reported stapes displacement instead of stapes velocity.
[2]Recent measurements by Chien *et al.* (2006) suggest that the angle correction may not be appropriate at frequencies above 2 kHz.
[3]As mentioned in Sec. II, the input voltage for E1 was smaller than that of the other ears for the EC drive case (0.1 V instead of 0.5 V). In Fig. 2, panels (A) and (B), the individual pressure and velocity magnitude measurements for E1 were scaled by a factor of 5 so as to enable the shapes of the four curves to be directly compared. The original magnitude data for E1 (not shown) are therefore closer to the noise than they appear in the figure. The signal to noise values, and the valid frequency ranges mentioned in the text were computed using the original, i.e., unscaled, E1 magnitudes.
[4]Another methodological difference is that the reported magnitude mean from the Aibara *et al.* (2001) study was first computed on a linear scale and then transformed to a logarithmic scale, whereas the current means were computed from the beginning on log-transformed data. A consequence of this is that the mean for Aibara *et al.* appears on a log scale to be skewed more toward the individual curves with higher signal levels than toward those with lower levels.

Aibara, R., Welsh, J. T., Puria, S., and Goode, R. L. (**2001**). "Human middle-ear sound transfer function and cochlear input impedance," Hear. Res. **152**, 100–109.

Chen, D. A., Backous, D. D., Arriaga, M. A., Garvin, R., Kobylek, D., Littman, T., Walgren, S., and Lura, D. (**2004**). "Phase 1 clinical trial results of the Envoy system: A totally implantable middle ear device for sensorineural hearing loss," Otolaryngol.-Head Neck Surg. **131**, 904–916.

Chien, W., Ravicz, M. E., Merchant, S. N., and Rosowski, J. J. (**2006**). "The effect of methodological differences in the measurement of stapes motion in live and cadaver ears," Audiol. Neuro-Otol. **11**, 183–197.

Fisch, U., Cremers, C. W., Lenarz, T., Weber, B., Babighian, G., Uziel, A. S., Proops, D. W., O'Connor, A. F., Charachon, R., Helms, J., and Fraysse, B. (**2001**). "Clinical experience with the Vibrant Soundbridge implant device," Otol. Neurotol. **22**, 962–972.

Gan, R. Z., Dyer, R. K., Wood, M. W., and Dormer, K. J. (**2001**). "Mass loading on the ossicles and middle ear function," Ann. Otol. Rhinol. Laryngol. **110**, 478–485.

Goode, R. L. (**1970**). "An implantable hearing aid—State of the art," Trans. Am. Acad. Ophthalmol. Otolaryngol. **74**, 128–139.

Goode, R. L., Killion, M., Nakamura, K., and Nishihara, S. (**1994**). "New knowledge about the function of the human middle-ear; development of an improved analog model," Am. J. Otol. **15**, 145–154.

Hough, J. V., Dyer, R. K., Jr., Matthews, P., and Wood, M. W. (**2001**). "Early clinical results: SOUNDTEC implantable hearing device phase II study." Laryngoscope **111**, 1–8.

Lenarz, T., Weber, B. P., Issing, P. R., Gnadeberg, D., Ambjornsen, K., Mack, K. F., and Winter, M. (**2001**). "Vibrant Sound Bridge System. A new kind of hearing prosthesis for patients with sensorineural hearing loss. 2. Audiological results," Laryngorhinootologie **80**, 370–387.

Murugasu, E., Puria, S., and Roberson, J. B., Jr. (**2005**). "Malleus-to-footplate versus malleus-to-stapes-head ossicular reconstruction prostheses: temporal bone pressure gain measurements and clinical audiological data," Otol. Neurotol. **26**, 572–582.

Onchi, Y. (**1961**). "Mechanism of middle ear," J. Acoust. Soc. Am. **33**, 794.

Peake, W. T., Rosowski, J. J., and Lynch, T. J., III (**1992**). "Middle-ear transmission: Acoustic versus ossicular coupling in cat and human," Hear. Res. **57**, 245–268.

Puria, S. (**2003**). "Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions," J. Acoust. Soc. Am. **113**, 2773–2789.

Puria, S., and Allen, J. B. (**1998**). "Measurements and model of the cat middle ear: Evidence of tympanic membrane acoustic delay," J. Acoust. Soc. Am. **104**, 3463–3481.

Puria, S., Peake, W. T., and Rosowski, J. J. (**1997**). "Sound-pressure measurements in the cochlear vestibule of human-cadaver ears," J. Acoust. Soc. Am. **101**, 2754–2770.

Rosowski, J. J., and Merchant, S. N. (**1995**). "Mechanical and acoustic analysis of middle ear reconstruction," Am. J. Otol. **16**, 486–497.

Shaw, E. A. G., and Stinson, M. R. (**1983**). *The Human External and Middle Ear: Models and Concepts* (Delft, The Netherlands).

Shera, C. A., and Zweig, G. (**1991**). "Phenomenological characterization of eardrum transduction." J. Acoust. Soc. Am. **90**, 253–262.

Shera, C. A., and Zweig, G. (**1992**). "Middle-ear phenomenology: The view from the three windows," J. Acoust. Soc. Am. **92**, 1356–1370.

Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2000**). "Acoustic responses of the human middle ear," Hear. Res. **150**, 43–69.

Zenner, H. P., Maassen, M. M., Plinkert, P. K., Zimmermann, R., Baumann, J. W., Reischl, G., and Leysieffer, H. (**1998**). "First implantation of a totally implantable electronic hearing aid in patients with inner ear hearing loss," (in German), HNO **46**, 844–852.

Zwislocki, J. J. (**1962**). "Analysis of middle ear function. I. Input Impedance." J. Acoust. Soc. Am. **34**, 1514–1523.

Auditory filters and the benefit measured from spectral enhancement

Jeffrey J. DiGiovanni[a] and Padmaja Nair
Auditory Psychophysics and Signal Processing Laboratory, Ohio University, Athens, Ohio 45701

(Received 20 December 2005; revised 15 June 2006; accepted 16 June 2006)

Algorithms designed to improve speech intelligibility for those with sensorineural hearing loss (SNHL) by enhancing peaks in a spectrum have had limited success. Since testing of such algorithms cannot separate the theory of the design from the implementation itself, the contribution of each of these potentially limiting factors is not clear. Therefore, psychophysical paradigms were used to test subjects with either normal hearing or SNHL in detection tasks using well controlled stimuli to predict and assess the limits in performance gain from a spectrally enhancing algorithm. A group of normal-hearing (NH) and hearing-impaired (HI) subjects listened in two experiments: auditory filter measurements and detection of incremented harmonics in a harmonic spectrum. The results show that NH and HI subjects have an improved ability to detect incremented harmonics when there are spectral decrements surrounding the increment. Various decrement widths and depths were compared against subjects' equivalent rectangular bandwidths (ERBs). NH subjects effectively used the available energy cue in their auditory filters. Some HI subjects, while showing significant improvements, underutilized the energy reduction in their auditory filters.
 [DOI: 10.1121/1.2221554]

PACS number(s): 43.66.Ts, 43.66.Fe [DOS] Pages: 1529–1538

I. INTRODUCTION

Speech intelligibility has been identified as one of the foremost problems for hearing-aid users. Although hearing aids provide amplification to compensate for the degree of hearing loss by normalizing loudness and improving thresholds, they do not sufficiently ameliorate the distortion introduced by the damaged cochlea that leads to poor spectral resolution abilities in individuals with sensorineural hearing loss (SNHL) (Plomp, 1978). One of the methods proposed to resolve this issue is spectral enhancement. It refers to the process of selectively amplifying spectral peaks in a speech signal while the spectral valleys or troughs are attenuated or remain unaffected. The goal of this strategy is to effectively improve the signal-noise ratio, leading to better intelligibility.

Numerous researchers have been studying the efficacy of spectral enhancement strategies producing contradictory results (Simpson *et al.*, 1990; Baer *et al.*, 1993; Giguère and Smoorenburg, 1998; Franck *et al.*, 1999; Miller *et al.*, 1999; Lyzenga *et al.*, 2002; DiGiovanni *et al.*, 2005). Miller *et al.* (1999) conducted a physiologic study using a spectral enhancement algorithm called contrast-enhanced frequency shaping (CEFS). They presented both unmodified vowels and vowels processed using CEFS to acoustically traumatized cats and studied their auditory neural responses. Results indicated that phase-locked neural representation of vowel formants was substantially improved with the CEFS algorithm. Giguère and Smoorenburg (1998) used cochlear models to assess spectral enhancement. By simulating the excitation pattern for the vowel /æ/for a normal ear and an ear with a 50% outer hair cell loss, they found no difference in the excitation patterns for the damaged cochlea with and without spectral enhancement. This led them to conclude that spectral enhancement is not a viable solution to the poor spectral resolution caused by broadened auditory filters of the damaged cochlea.

Again, speech perception tests using spectral enhancement algorithms have yielded mixed results. Simpson *et al.* (1990) tested subjects with sensorineural hearing loss on speech-in-noise tests and found that they showed improved results in word and sentence identification tasks. They found an increase of 6.4% in identification of consonant-vowel-consonant words, and an 11.4% increase for Bench-Kowal-Bamford (BKB) sentences using spectral enhancement. Baer *et al.* (1993) changed certain parameters of the same algorithm and carried out further testing; their results however did not show the significant amount of improvement shown by Simpson *et al.* (1990). The only condition in which they were able to demonstrate a significant benefit occurred when spectral enhancement was combined with wide dynamic range compression. In this condition, an 0.8-dB improvement in SNR was noted. A similar study by Franck *et al.* (1999) found that combining spectral enhancement with dynamic range compression reduced the benefit of the spectral enhancement scheme. Although they found an improvement in vowel perception scores in the "spectral enhancement alone" and "spectral enhancement along with single channel compression" conditions, the final consonant perception in noise scores worsened considerably. Additionally, the best speech intelligibility scores were found in the unprocessed condition (Franck *et al.*, 1999). More recently, Lyzenga *et al.* (2002) combined a spectral expansion algorithm with a different algorithm that worked to decrease the effects of upward spread of excitation and observed a 1-dB improvement

[a]Electronic mail: digiovan@ohio.edu

in speech reception threshold (SRT). The range of results for these different studies includes a decrease in intelligibility to a modest improvement.

The incongruity of the results of these studies indicates that the theory of spectral enhancement is inherently flawed or has been inadequately implemented. In a previous study, DiGiovanni *et al.* (2005) attempted to assess the theoretical viability of spectral enhancement. Both normal-hearing and hearing-impaired subjects were required to detect and discriminate a narrow-band peak (the approximate width of a vowel formant) in broadband noise. Frequencies flanking the target narrow-band signal were decremented to improve the spectral contrast and consequently the signal-to-noise ratio. The spectral decrements surrounding the target stimuli were either 100 or 200 Hz wide, and their depth was either 3 or 6 dB (i.e., their intensity was reduced by 3 or 6 dB compared to all other bands including the target band). The test conditions ranged from no decrements to a maximum decrement 6 dB deep and 200 Hz wide. Results showed that spectral decrements significantly improved detection and discrimination of spectral peaks in noise for both normal-hearing and hearing-impaired individuals, leading the authors to the conclusion that spectral enhancement is a viable premise. However, as the stimulus used was broadband noise, conclusions regarding the effect of spectral enhancement on speech were limited. The authors further noted that the amount of benefit shown by hearing-impaired subjects, although significant, was less than the benefit demonstrated by normal-hearing subjects. They posited that the broadened auditory filters of the damaged cochlea in individuals with sensorineural hearing loss limited the amount of measured improvement, though auditory filters were not measured in that study.

The goal of the present study was to examine the effects of increasing spectral contrast on detectability of raised harmonics in a broadband harmonic spectrum. A second goal was to explore the relationship of auditory filter bandwidths and the improvements in increment detection when spectral enhancements are added. In the first experiment, auditory filters were measured for all subjects. In the second experiment, both normal-hearing and hearing-impaired subjects were required to detect a narrow-band peak (approximately the width of a vowel formant) in a harmonic spectrum with a fundamental frequency of 50 Hz and harmonics up to and including 5000 Hz. Spectral decrements of varying width and intensity relative to the overall spectrum level were inserted into the bands flanking the peak.

II. EXPERIMENT 1: AUDITORY FILTERS WIDTHS

Individuals with sensorineural hearing loss experience greater difficulty in speech discrimination tasks and this has, in part, been attributed to their wider auditory filters. DiGiovanni *et al.* (2005) found that, although hearing-impaired subjects showed a significant improvement with spectral enhancement, they demonstrated less benefit as compared to normal-hearing subjects. In the present study, the auditory filter bandwidth of hearing-impaired subjects was measured at 2 kHz, since this frequency lies well inside the region where second formants mostly occur. Next, listeners' benefit of spectral enhancement was related to their auditory filter widths to assess the hypothesis that wider auditory filters lead to poorer benefit from spectral enhancement. Quantitative relationships between a subject's improvement and their auditory filter bandwidth will be made.

TABLE I. Audiometric thresholds and ages of subjects in the HI group.

		Audiometric thresholds (kHz)							
Subject	Age (years)	0.25	0.5	1.0	1.5	2.0	3.0	4.0	8.0
I1	66	20	25	40	45	50	45	50	65
I2	75	15	20	35	⋯	40	⋯	50	60
I3	78	55	45	45	⋯	40	⋯	55	80
I4	67	55	60	55	45	40	35	35	65

A. Methods

1. Subjects

Two groups of subjects participated in this study: four individuals with normal hearing (NH group) and four individuals with sensorineural hearing loss (HI group). Informed consent was obtained from all subjects and they were paid for participation.

The NH group had audiometric thresholds of 15 dB HL or better from 0.25 to 8 kHz and a negative report of auditory pathology. The HI group had mild to severe sensorineural hearing loss, with thresholds ranging from 35 to 50 dB HL from 1.5 to 3.0 kHz. Audiograms for this second group were measured at octave intervals, and interoctave intervals when appropriate, from 0.25 to 8.0 kHz. Audiometric data for the HI group are shown in Table I. Also, to be included in the experiment, participants were required to have a relatively constant threshold microstructure around the test frequency, 2.0 kHz. For frequencies between 1.5 and 2.5 kHz, thresholds were measured at 25-Hz intervals to ensure that hearing sensitivity did not vary more than 10 dB in the region around 2.0 kHz. Thresholds measured between 1.5 and 2.5 kHz were obtained using a computer-implemented version of Békésy discrete-frequency audiometry. Participants pressed a button on a computer keyboard to signal "louder" or "softer," which raised or lowered the signal level in 1.5-dB steps. Participants were instructed to press the louder button if the signal was inaudible and the softer button if it was audible. The stimulus for the Békésy audiometry task was a 250-ms tone (with 20-ms raised cosine ramps). Thresholds for each frequency were based on the mean of the last 10 of 12 reversals in stimulus level direction.

2. Stimuli

The stimuli were generated digitally at a sampling rate of 24.414 kHz using a computer (Dell Dimension, DIM 4550). The computer was equipped with a signal processor (Tucker-Davis Technologies, RP2.1) 24-bit digital-to-analog converter and, after filtering and attenuation (Tucker-Davis Technologies, PA-5), sounds were presented to one ear of a listener through an earphone (Telephonics, TDH-39P). Bands of noise and 2-kHz pure tones were generated using the

aforementioned hardware. The stimuli were 500 ms in duration with a 500-ms interstimulus interval. The pure tone had a raised cosine ramp of 10 ms. The noise was low-pass filtered at 8.0 kHz and was generated using a Gaussian distribution. The noise was fixed in level at 46 dB/Hz, which is equivalent to an overall level of 85 dB SPL. The pure tone varied in level depending on the subject's response.

3. Procedure

A three-interval forced-choice procedure was used to measure the subjects' auditory filter bandwidths. A two-up one-down rule was used to track 70.7% on the psychometric function (Levitt, 1971). Each block consisted of ten reversals and the threshold was calculated from the final eight reversals. The step size was 4 dB for the first two reversals and decreased to 2 dB for the remainder of the block. In any given trial, two of the intervals contained notched noise (standard intervals) while one of the intervals contained both the noise and a 2-kHz tone centered in the spectral gap of the noise (target interval). f_{lo} and f_{hi} defined the cutoff frequencies for the high-pass and low-pass conditions, respectively. The subjects' task was to determine the interval containing the tone. Correct answer feedback was given following each response for a given trial. The subject was seated in a sound attenuating booth. Each subject underwent a minimum of two hours of practice before data collection began. Three blocks were collected for a given point, or notched-noise condition, for a total of 15 blocks per subject. The three cutoff frequencies that were collected were derived from g values of 0.0, 0.2, and 0.4, resulting in five points per filter (0,0; 0.2, 0.2; 0.4, 0.4; 0.2, 0.4; 0.4, 0.2) (Glasberg and Moore, 1990) where

$$g = \frac{|f - f_0|}{f_0} = \frac{\Delta f}{f}. \tag{1}$$

For instance, for $g=0.20$ and $f_0=2.0$ kHz, $f_{lo}=1.6$ kHz and $f_{hi}=2.4$ kHz.

B. Results

The auditory filters were modeled from the collected data using Patterson *et al.*'s (1982) formula for curve fitting [Roex(p,r)]:

$$W(g) = (1 - r)(1 - pg)e^{-pg} + r, \tag{2}$$

where p determines the passband width and the rate of fall of the skirts of the filter and r is the dynamic range limiter.

Auditory filter bandwidths were derived by fitting the five points of data into the two-parameter Roex(p,r) model. Table II shows the equivalent rectangular bandwidths (ERBs) of the modeled data along with the subject age and group. The average ERBs were 276 and 473 Hz for the NH and HI subjects, respectively.

C. Discussion

Auditory filter bandwidths measured in this experiment largely agree with those established in literature. Dubno and Dirks (1989) measured filters in nine normal-hearing subjects at a spectrum level of 40 dB/Hz. At 2.0 kHz, their subjects had ERBs ranging from 220 to 320 Hz. The range for our NH subjects was 222–300 Hz. They also measured auditory filters in hearing-impaired subjects. Their hearing-impaired subjects had ERBs ranging from 400 to 1640 Hz, whereas the ERBs for the HI subjects in this study ranged from 356 to 648 Hz. Glasberg and Moore (1986) also measured auditory filter bandwidths in normal and hearing-impaired ears. Their ranges were in agreement with the current data as well as with that of Dubno and Dirks' (1989).

TABLE II. ERBs and ages for HI and NH subjects.

Subject	Age (years)	ERB (Hz)
I1	66	647.6
I2	75	449.1
I3	78	355.7
I4	67	440.6
HI Average		473.3
N1	21	222.4
N2	21	300.2
N3	24	296.7
N4	24	283.0
NH average		275.5

The difference between the performance of the NH and HI groups can be attributed to a loss in frequency resolution. Broadened auditory filters in hearing-impaired subjects is a common finding (Glasberg and Moore, 1986; Dubno and Dirks, 1989). The current data also reveal that the HI subjects performed worse than the NH subjects. Overall, our HI subjects' had an average ERB 1.7 times that of the NH subjects' ERBs.

III. EXPERIMENT 2: INCREMENT DETECTION THRESHOLDS

DiGiovanni *et al.* (2005) demonstrated that spectral enhancement showed promise as a viable method to improve speech intelligibility in hearing-impaired individuals. The stimulus used in their study was broadband noise with a 100-Hz-wide narrow-band peak at 2 kHz. Spectral decrements were inserted adjacent to this peak in order to provide spectral contrast and subsequently improve the signal-to-noise ratio. Although the study demonstrated significant improvement for both normal-hearing and hearing-impaired subjects, it did not indicate whether the results could be generalized to speech stimuli as well. In this study, subjects listened in detection tasks using a broadband harmonic spectrum. Test conditions included various depths and widths of spectral decrements to assess their relationship with improved performance on spectral enhancement tasks, if any. Furthermore, a second set of conditions was tested with stimuli that differed in one respect; a three-harmonic, fixed 10-dB increment was added to the standard and target intervals. These conditions were included to test the possibility that by introducing a first formantlike fixed increment, upward spread of masking may reduce the benefit of enhancing the spectrum.

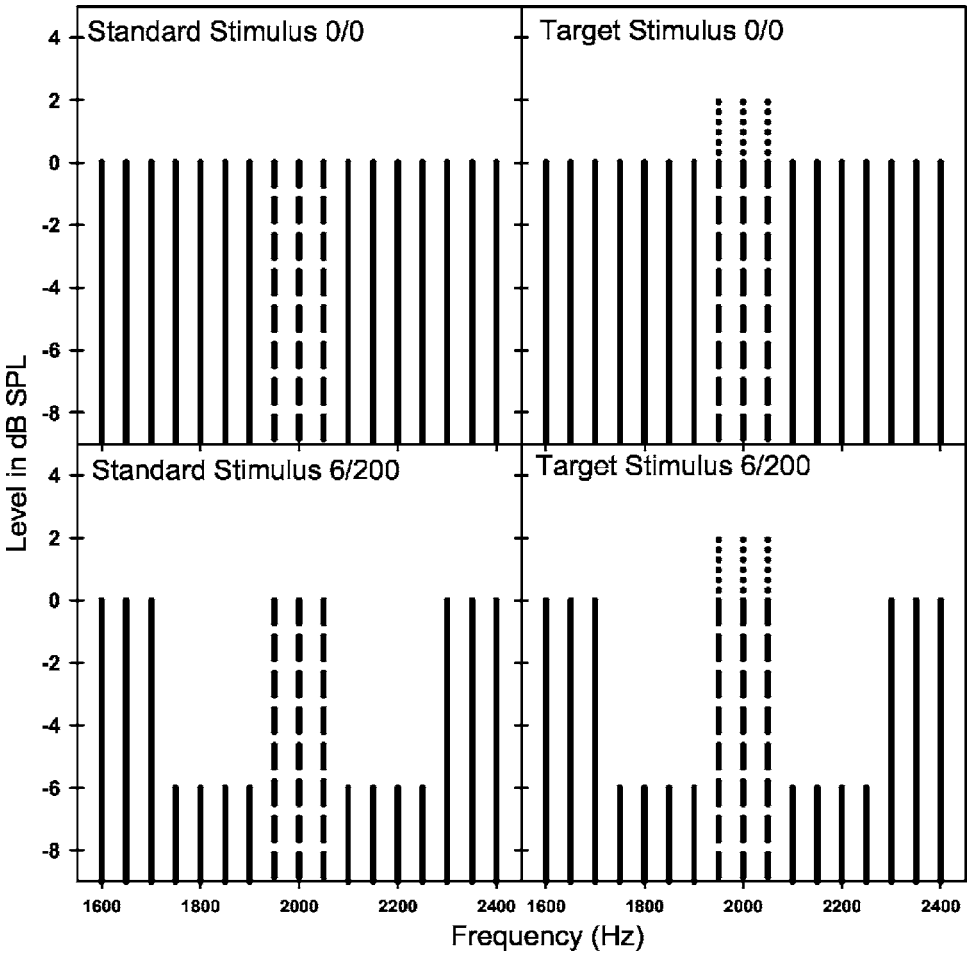


FIG. 1. Schematic representations of the stimuli used in experiment 2. Shown in the two left panels are the standard stimuli for the 0/0 (upper panel) and the 6/200 (lower panel) conditions. Shown in the two right panels are the target stimuli for the 0/0 (upper panel) and 6/200 (lower panel) conditions. In this example the target stimuli are shown with an increment of 2 dB relative to the spectrum level of the harmonic spectrum.

A. Methods

1. Subjects

Subjects used were the same as in experiment 1.

2. Stimuli

Two sets of conditions were run, one with the standard stimulus and another with a stimulus that had a first formant-like increment in its frequency spectrum, called the F1 stimulus. While the standard stimulus for this experiment was a harmonic spectrum with F_0=50 Hz (band limited at 5000 Hz), the F1 stimulus had an additional 100-Hz-wide, 10-dB increment inserted at the first formant region (i.e., 950, 1000, and 1050 Hz). The stimuli were 500 ms in duration with a 500-ms interstimulus interval. The target stimulus for both sets of conditions had a 100-Hz-wide (three harmonics), narrow-band increment centered at 2 kHz. Spectral decrements of varying widths were inserted adjacent to the 2 -kHz increment. The convention used to refer to these conditions will be the depth (dB)/width (Hz). For instance, a 6 -dB-deep, 200-Hz-wide decrement will be referred to as 6/200. The control condition with no decrements will be referred to as 0/0. F1 stimuli are denoted similarly, but with the addition of F1 in the designation (e.g., 6/200F1). Figure 1 shows a schematic of part of the spectrum for the 0/0 and 6/200 conditions for both the standard and the target stimuli.

3. Procedure

A three-alternative forced-choice task was used to measure the subjects' ability to detect the narrow-band increment in a harmonic series. A two-down one-up paradigm tracking 70.7% on the psychometric function was used (Levitt, 1971). The step size was 3 dB for the first two reversals and 1 dB for the remaining reversals. A block of trials ended after the completion of ten response reversals. The first two reversals were discarded and threshold was the mean of the final eight reversals. The standard harmonic spectrum was presented in any two of three intervals in a given trial, and in the other interval, the harmonic spectrum with the 2-kHz increment was presented. The interval order was determined from a uniform, random distribution. The subjects were instructed to pick the interval that was different from the other two and respond using the computer keyboard. Feedback was given on a trial-by-trial basis.

Several conditions were collected: 0/0, 6/200, 6/400, 9/200, 9/400, 6/400F1, and 9/400F1 by varying the width and depth of the spectral decrements surrounding the 2 -kHz increment. The order of conditions was randomized within subjects. A minimum of three blocks per condition were measured to calculate thresholds. More blocks were run if the standard deviation exceeded 3 dB until the standard deviation was below 3 dB. The number of blocks never exceeded five.

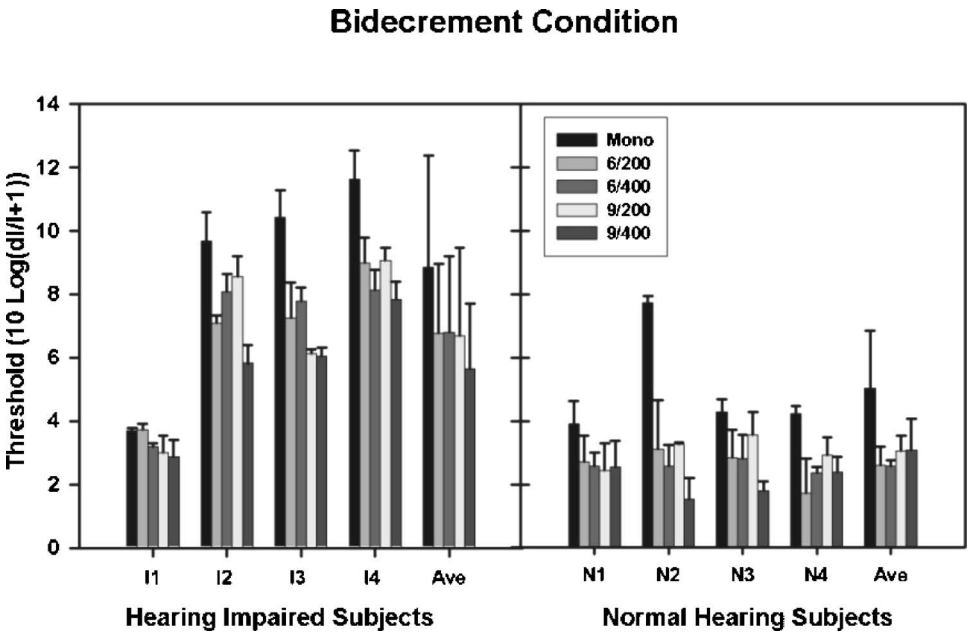


FIG. 2. The increment detection results of the NH and HI groups for the bidecrement conditions in experiment 2 are shown. The data are presented after they were converted into ΔL (i.e., the difference between the centerband spectrum level of the signal and standard intervals). Data for the HI subjects are shown in the left panel and data for the NH subjects are shown in the right panel. The increment thresholds in dB for each subject as well as the group average are shown with corresponding standard deviation bars. The 0/0 condition is referred to as Mono.

B. Results

Increment detection thresholds for both subject groups are shown in Figs. 2 and 3. Narrow-band target thresholds were converted from $10\log(\Delta I/I)$ to $10\log(\Delta I/I+1)$ in order to demonstrate the relative difference at threshold between the signal increment and the harmonic spectrum. The spectral enhancement conditions are thus expressed in terms of decibel increments above the spectrum level of the harmonic series. The spectral enhancement conditions ranged from no enhancement (0/0) to a maximum enhancement in the 9/400 condition. Additionally, F1 conditions (9/400F1 and 6/400F1), which included a spectral increment of 10 dB at 950, 1000, and 1050 Hz, were tested. Benefit as it is understood in this study is the difference between threshold in a given enhancement condition relative to the 0/0 condition.

A repeated measures analysis of variance was performed on the entire listener pool to measure the effects of spectral enhancement in the standard and the F1 conditions. A significant group interaction was found on the standard condition [$F(1,6)=6.42$, $p<0.05$], indicating that, overall, normal-hearing subjects had better thresholds than the hearing-impaired group. Additionally, significant effects of decrement depth [$F(2,5)=10.66$, $p<0.05$] and width [$F(2,5)=10.51$, $p<0.05$] were found, indicating that greater the depth and width of the decrements, the better the detection thresholds for both groups. No significant depth by group [$F(2,5)=1.45$, $p=0.318$] or width by group [$F(2,5)=0.43$, $p=0.67$] interactions were found, suggesting that there was no differential effect of decrement depth or width on either of the two groups. The three-way interaction (depth × width × group) was not significant [$F(4,3)=0.48$, $p=0.76$]. Overall, the analysis along with the data shown in Fig. 4 suggests that both groups showed greatest benefit in the condition with the greatest spectral decrement (9/400).

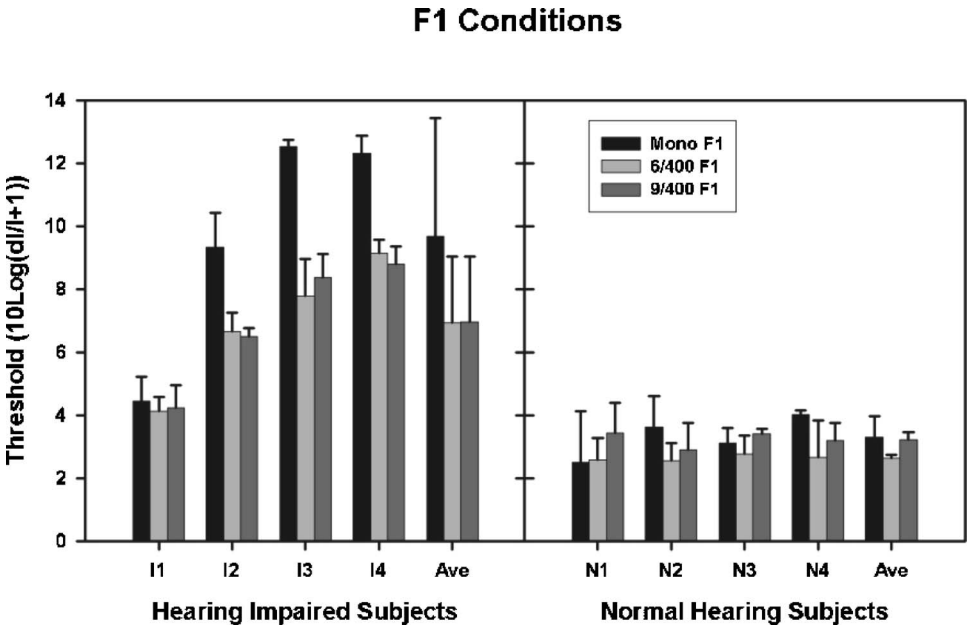


FIG. 3. As in Fig. 2, but now showing results for the F1 stimuli.

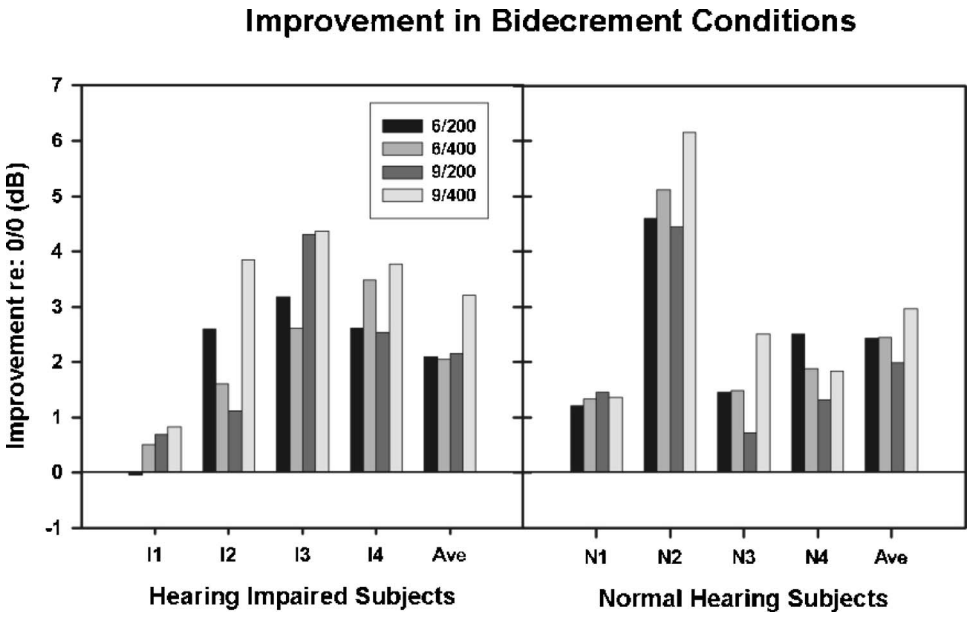


FIG. 4. The benefit measured for each decrement condition in experiment 2 is shown relative to the 0/0 condition. The comparisons were made from the ΔL converted data. The left panel shows individual benefit along with group average for the HI group and the right panel shows the same data for the NH group.

For the F1 conditions, a significant group interaction was found $[F(1,6)=14.25,\ p<0.05]$, indicating that the normal-hearing subjects had better thresholds than the hearing-impaired subjects. Also, significant effects of decrement depth $[F(2,5)=13.46,\ p<0.05]$ and width $[F(1,6)=9.99,\ p<0.05]$ were found, indicating that wider and deeper decrements led to better detection thresholds in the F1 conditions as well. The significant depth by group interaction $[F(2,5)=6.1,\ p<0.05]$ suggests that there was a differential effect of decrement depth on each group. Furthermore, a significant interaction between width and depth was found $[F(2,5)=13.46,\ p<0.05]$, suggesting that the relative importance of the decrement width varied depending on the depth of the decrement. The three-way interaction (depth × width × group) was also significant $[F(2,5)=6.1,\ p<0.05]$.

These analyses show that spectral enhancement demonstrated significant benefit for both normal-hearing and hearing-impaired listeners. The F1 condition, which approximates a steady-state vowel more closely than the standard stimulus, suggests that spectral enhancement could be viable for real-life situations. The greatest amount of improvement was seen for the most extreme 9/400 spectral enhancement condition, and the least improvement was seen for the 6/200 condition.

C. Discussion

The findings from experiment 2 agree with the findings from DiGiovanni *et al.* (2005). In their study, one of the experiments was similar to the standard stimulus in this study, except that they used noise rather than a harmonic complex. Both studies showed that the HI subjects performed worse than the NH. While they suggested that there may be a link between the measured improvement and auditory filter bandwidths, auditory filters were not measured and therefore this relationship could not be directly evaluated. From Figs. 4 and 5 it can be seen that the current HI subjects showed similar improvement to the NH subjects for the standard conditions and greater improvement for the F1 conditions, whereas the NH group in DiGiovanni *et al.* (2005) showed more improvement than the HI subjects. Overall, the findings of their study and the current study are in agreement. It is clear that both the NH and HI subjects in this study demonstrated benefit in detection of a set of harmonics when the spectrum was enhanced.

While both the NH and HI subjects generally demonstrated benefit in the F1 conditions, the NH subjects did not show nearly as much benefit as the HI subjects, as shown in Fig. 5. On average, the HI subjects realized a 2.72-dB benefit in the F1 conditions while the NH subjects only showed a 0.38-dB improvement. This finding is contrary to our prediction. We predicted that subject performance, and consequently their improvement, would be related to the analysis band that was used, namely their auditory filter bandwidth in equivalent rectangular bandwidth (ERB) units. Our NH subjects had consistently narrower auditory filters than our HI subjects, which led to our prediction that NH subjects would perform better on this task as well as show greater improvement.

Despite an average improvement of almost 3 dB for both the F1 conditions, the amount of improvement varied among the HI subjects. It is difficult to predict if this variability will carry over to a spectral-enhancement algorithm used in a real-world environment. Another analysis was performed to better understand the effect of F1 on performance. The original analysis compared the enhanced conditions to the 0/0 condition separately for experiments 1 and 2. However, a comparison to the 0/0 data is not required. To understand the impact of F1, the three conditions (0/0, 6/400, and 9/400) can be directly compared between the standard and F1 data from experiment 2. To do this, the 0/0 threshold is subtracted from the 0/0F1 threshold. This is repeated for the 6/400(F1) and the 9/400(F1) conditions. When this analysis is done, most of the findings are as expected. That is, the 6/400 and 9/400 conditions have the same or worse thresholds for both groups. This is also true for the 0/0 condition for the HI subjects. However, the 0/0 threshold for the NH

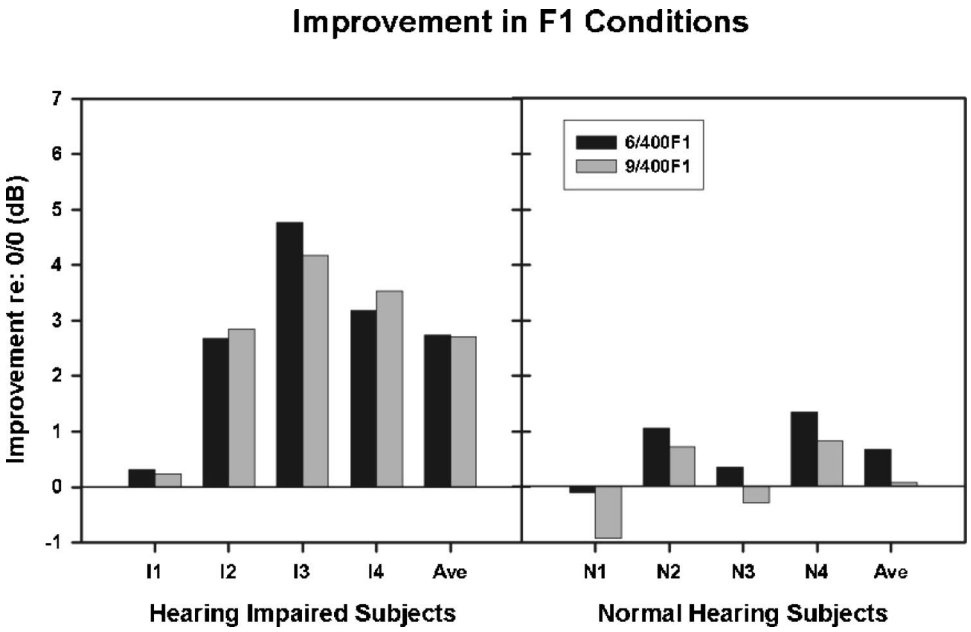


FIG. 5. As in Fig. 4, but now showing results for the F1 stimuli.

group was better for all subjects in the presence of F1. The source of this finding is uncertain for two reasons: (1) all subjects were trained for at least 4 h prior to data collection and (2) the presentation order was randomized. All other findings, however, were as expected. Therefore, the finding that HI subjects show more improvement in the F1 conditions should be made with caution.

IV. GENERAL DISCUSSION

The effective power within an auditory filter may be the significant factor in determining performance in experiment 2 (Bilger, 1978; Summers and Leek, 1994). Similar to Summers and Leek's (1994) ripple detection results, our HI subjects performed worse than the NH subjects. Despite this, the HI subjects showed improvement in the spectrally enhanced conditions. Summers and Leek (1994) posited that broadened auditory filters resulted in a flattened internal spectrum leading to the reduced performance. In the current study, auditory filter data show that the HI subjects have broader auditory filters than the NH subjects. This might explain the poorer overall performance of the HI group. If the overall power within the auditory filter is the determining factor for performance, then the relative signal-to-noise ratio or, rather, the increment-to-spectrum ratio would decrease with increasing bandwidth. Furthermore, this would explain the improvement measured for both groups when spectral decrements are inserted. Removing energy from the background spectrum results in an increase in the relative power of the increment to the spectrum. From this, it follows that little or no benefit would be realized for extremely wide filter bandwidths. As the filter widens, the effect of removing energy from part of the spectrum becomes less significant and, ultimately, will have a negligible impact on performance. However, since the subject groups were not age matched, differences in performance between groups may be partially a result of age differences rather than auditory-filter bandwidths.

To better understand the amount of improvement related to auditory filter bandwidths, two quantitative models were developed. Multiple cues have been well described to account for detection of increments in broadband signals (Formby *et al.*, 1994). Two of these cues are relevant to this analysis: overall-energy and spectral-profile cues. The overall-energy cue simply assumes that performance improvements can be predicted by the energy difference between the standard stimulus and target stimulus within a band. The spectral-profile cue assumes that the listener utilizes changes in the spectral profile of between the standard and target stimulus without the contribution of the increment. (Formby *et al.*, 1994; Heinz and Formby, 1999). While neither of these cues incorporate profile analysis *per se*, a profilelike analysis could have been used by the listener to detect the increment (Green, 1988). Using these cues, two models were developed: (1) the overall-energy model and (2) the spectral-profile model. These models were developed to predict the maximum possible benefit subjects could realize, depending on the possible listening cues.

The overall-energy model assumed that the main cue for detection was the comparison of overall energy in a given bandwidth (i.e., the auditory filter bandwidth) between the standard and comparison stimuli. To calculate benefit, the standard stimulus used the 0/0 condition. The overall level was calculated for 0/0 and each enhancement condition within each subject's auditory filter bandwidth. To do this, the number of harmonics within the auditory filter was calculated. Next, the overall energy for these harmonics was calculated considering the individual levels of the increment harmonics (without the pedestal), the decremented harmonics in the enhancement condition, and the fixed-level harmonics in the event the auditory filter bandwidth was wider than the decrement width plus the increment width. In this manner, as the auditory filter bandwidth exceeded the width of the increment plus decrement, less benefit would be predicted. The benefit for a particular condition was the difference in energy between these two stimuli.

The spectral-profile model assumed listeners are using the spectral differences within a band as the cue. The model

TABLE III. Comparison of benefit predicted by overall-energy model, spectral profile model, and the measured benefit for each subject.

	I1 (648 Hz)			I2 (449 Hz)			I3 (356 Hz)			I4 (441 Hz)		
	Energy (dB)	Profile (dB)	Measured (dB)	Energy (dB)	Profile (dB)	Measured (dB)	Energy (dB)	Profile (dB)	Measured (dB)	Energy (dB)	Profile (dB)	Measured (dB)
6/200	3.5	6.0	–0.04	3.2	4.4	2.59	2.9	6.0	3.17	3.5	6.0	2.62
6/400	3.5	6.0	0.51	4.2	6.0	1.60	2.9	6.0	2.62	3.5	6.0	3.48
9/200	4.9	9.0	0.69	4.4	6.3	1.11	4.0	9.0	4.30	4.9	9.0	2.54
9/400	4.9	9.0	0.82	6.0	9.0	3.85	4.0	9.0	4.37	4.9	9.0	3.77
	N1 (222 Hz)			N2 (300 Hz)			N3 (297 Hz)			N4 (283 Hz)		
6/200	1.9	6.0	1.21	2.9	6.0	4.60	1.9	6.0	1.46	1.9	6.0	2.50
6/400	1.9	6.0	1.33	2.9	6.0	5.12	1.9	6.0	1.48	1.9	6.0	1.88
9/200	2.6	9.0	1.46	4.0	9.0	4.44	2.6	9.0	0.72	2.6	9.0	1.32
9/400	2.6	9.0	1.37	4.0	9.0	6.16	2.6	9.0	2.50	2.6	9.0	1.83

assumes that listeners are not incorporating the level of the increment with the pedestal as part of the cue. Therefore, the energy within a given bandwidth was calculated by comparing the spectral differences between the standard and target stimuli of the magnitude spectrum within the auditory filter but outside the increment. Therefore, the number of harmonics within each subject's auditory filter was calculated. The three increment harmonics were subtracted. The remaining harmonics were then considered. For the standard stimulus, the overall level within the auditory filter minus the level of the increment was calculated by adding the levels of the remaining harmonics. For the target stimulus, the overall level within the auditory filter minus the level of the increment level was calculated by adding the level of the fixed-level and decremented harmonics, the number of which depended on the enhancement condition and the auditory filter bandwidth. The predicted improvement was calculated as the difference between the level of this band for the 0/0 condition and an enhancement condition.

The maximum achievable benefit was calculated using the subjects' measured auditory filter bandwidths for both models. Using the subjects' own bandwidths allows for the most direct test of the model. The model predictions along with the measured benefit are shown in Table III. Several qualitative observations can be made. First, for both models, greater depth always predicted more benefit. Increasing width, however, required an analysis band that exceeded the subjects' ERB in order to predict greater benefit. This impacted the HI subjects more due to their wider ERBs. Subject data shown in Fig. 4 indicate that, on average, greater depths and widths contributed to improved performance. Second, both models predict less benefit for broader ERBs. However, subject data did not consistently agree with this trend. HI subjects provided a greater range of ERBs, but their performance did not monotonically vary with their ERB. It is likely that a greater pool of subjects, especially HI subjects having broader ERBs, would smooth out this trend. Third, both models overpredicted the benefit. It may be that subjects do not make ideal use of the cues. The spectral-profile model consistently overpredicted the data while the overall-energy model predicted numbers within the proximity of the data.

To quantify the predictive value of each of the models, an error analysis was performed. The error analysis involved a calculation of the root-mean-square (rms) error, in dB, between the predicted and measured benefit. The results of this analysis are shown in Table IV. The overall-energy model has a modest 1.84-dB rms error for all subjects and all conditions combined. The spectral-profile model, however, has almost three times the rms error than the overall-energy model. Furthermore, the spectral-profile model overpredicts the measured benefit for every condition and every subject. The error analysis for the overall-energy model was calculated for the NH and HI group means separately to assess if the model was more accurate for one of the groups. The rms error was 1.19 and 2.32 dB for the NH and HI groups, respectively. This reveals that the overall-energy model is a better predictor of benefit for the NH listeners than the HI listeners. Therefore, subjects likely are using the overall energy in a particular analysis band as a partial or total cue, that determines their performance in these listening tasks. Moreover, the model predicts the NH performance within 1.19 dB rms error while the HI error is approximately double. It may be that the HI listeners in this study could not utilize the maximal benefit. However, it appears that the HI subjects had a greater variability than the NH subjects for this model. From Table III, subject I3 shows excellent agreement between predicted and measured improvement. I4 shows a lesser degree of agreement, and I1 and I2 showed the least agreement. Given this variability for HI subjects, it is possible that the auditory filter data may not be sufficient to describe the performance benefit for all HI subjects.

This study measured auditory filters and increment detection thresholds for NH and HI subjects to elaborate on a suggested relationship between detection thresholds and auditory filters as discussed in DiGiovanni *et al.* (2005). DiGiovanni *et al.* (2005) measured increment detection thresholds

TABLE IV. Error analysis for predicted and measured benefit.

	rms error (dB)	
	Energy	Profile
All subjects	1.8	5.3
NH group	1.2	5.5
HI group	2.3	5.2

as well as frequency discrimination for a narrow-band increment in a broadband noise. They found that subjects improved similarly in both tasks, thus both tasks were providing the same type of information. The current study limited the measurement to increment detection thresholds largely due to this fact as well as that the usage of a harmonic spectrum did not lend itself to a frequency discrimination task.

Despite the fact that at high levels an increment may have a fairly broad excitation, literature has consistently shown that the important aspect of neural coding an increment (or formant) in a broadband spectrum (e.g., a vowel spectrum) is through synchrony and not neural excitation, and even at high presentation levels that cause a broadband excitation, the frequency specificity of synchrony was maintained (Sachs and Young, 1980; Horst *et al.*, 1985). In this manner, if a spectral peak can be "reintroduced" into the signal, it follows that regardless of the bandwidth of excitation, synchrony will increase local to the place of the spectral peak in a similar way that formants are coded at higher levels. Therefore, an algorithm that can partially restore the spectral representation of speech will likely improve speech intelligibility.

V. CONCLUSIONS

Based on psychophysical, physiological, and clinical studies, it appears that modest improvements can be achieved by enhancing peaks in a spectrum by either selectively amplifying the peak or attenuating the energy adjacent to a peak. Based on the current data and those from literature (as cited), several inferences can be made: (1) deeper and wider decrements are better, at least to 400 Hz wide and 9 dB deep; (2) auditory filter bandwidths are related to improvement in enhancement conditions; (3) F1 should not be amplified significantly (Miller *et al.*, 1999); (4) amplification should be focused on F2 and F3 frequency regions (Miller *et al.*, 1999); (5) the effects of combined processing (e.g., spectral enhancement and compression) are not well understood (Franck *et al.*, 1999); and (5) since it is well known that consonant perception is a critical factor for speech intelligibility, any spectral enhancement algorithm needs to account for consonant perception.

These experiments suggest that further research and the development of an algorithm for spectral enhancement remain worthy ventures. Historical arguments suggest that broadened auditory filters eliminate the possible benefit from spectral enhancement. However, animal studies have shown that spectral enhancement is effective in restoring the formant representation at the neural level (Miller *et al.*, 1999). DiGiovanni *et al.* (2005) showed promising results by testing normal and hearing-impaired listeners in two psychophysical studies. To further these findings using stimuli that are more speechlike, two psychophysical experiments were performed to test the NH and HI benefit from spectral enhancement. These experiments included (1) auditory filter measurements and (2) narrow-band signal detection with and without adjacent spectral decrements as well as with and without an added 10-dB fixed increment to simulate the effect of a lower formant (i.e., the F1 condition). Even though the HI subjects performed worse overall, they showed more improvement than the NH subjects for both experiments. Therefore, we conclude that the NH and HI listeners generally benefit from spectral enhancements when discriminating a narrow-band signal in a broadband harmonic spectrum. Furthermore, HI subjects received more benefit from the enhancements, contrary to the finding of DiGiovanni *et al.* (2005). Finally, the analysis of auditory filter bandwidths showed that the NH listeners effectively use all the energy in their ERB, while some of the HI subjects underperform the predictions made by the overall-energy model, suggesting that, for these subjects, the reduced performance is not fully accounted for by their wider ERBs. It is possible that this is a result of the physiological limitation imposed by SNHL. In sum, spectral enhancement remains a viable goal to make modest improvements in the local signal-to-noise reduction for the detection of spectral peaks.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their insightful comments on an earlier version of the manuscript. This work was supported by an Ohio University Scholarly Activity Award.

Baer, T., Moore, B. C. J., and Gatehouse, S. (**1993**). "Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality, and response times," J. Rehabil. Res. Dev. **30**, 49–72.

Bilger, R. C. (**1978**). "A revised critical band hypothesis," in *Hearing and Davis: Essays Honoring Hallowell Davis*, edited by S. K. Hirsh, D. H. Eldridge, I. J. Hirsh, and S. R. Silverman (Washington U.P., St. Louis, MO), pp. 191–198.

DiGiovanni, J. J., Nelson, P. B., and Schlauch, R. S. (**2005**). "A psychoacoustic evaluation of spectral enhancement," J. Speech Lang. Hear. Res. **48**, 1–15.

Dubno, J. R., and Dirks, D. D. (**1989**). "Auditory filter characteristics and consonant recognition for hearing-impaired listeners," J. Acoust. Soc. Am. **38**, 1666–1680.

Formby, G., Heinz, M. G., Luna, C. E., and Shaheen, M. K. (**1994**). "Masked detection thresholds and temporal integration for noise band signals," J. Acoust. Soc. Am. **96**, 102–114.

Franck, B. A., Sidonne, C., van Kreveld-Bos, G. M., Dreschler, W. A., and Verschuure H. (**1999**). "Evaluation of spectral enhancement in hearing aids, combined with phonemic compression," J. Acoust. Soc. Am. **106**, 1452–1464.

Giguère, C., and Smoorenburg, G. F. (**1998**). "Computational modeling of outer hair cell damage: Implications for hearing aid signal processing," in *Psychophysics, Physiology and Models of Hearing*, edited by T. Dau, B. Kollmeier, and V. Hohmann (World Scientific, Singapore), pp. 155–165.

Glasberg, B. R., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear implants," J. Acoust. Soc. Am. **79**, 1020–1033.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Deviation of auditory filter shapes from notch-noise data," Hear. Res. **47**, 103–138.

Green, D. M., *Profile Analysis* (Oxford U. P., New York, **1988**), p. 93.

Heinz, M. G., and Formby, C. (**1999**). "Detection of time and bandlimited increments and decrements in a random-level noise," J. Acoust. Soc. Am. **106**, 313–326.

Horst, J. W., Javel, E., and Farley, G. R. (**1985**). "Extraction and enhancement of spectral structure by the cochlea," J. Acoust. Soc. Am. **78**, 1898–1901.

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

Lyzenga, J., Festen, J. M., and Houtgast, T. (**2002**). "A speech enhancement scheme incorporating spectral expansion evaluated with simulated loss of frequency selectivity," J. Acoust. Soc. Am. **112**, 1145–1157.

Miller, R. L., Calhoun, B. M., and Young, E. D. (**1999**). "Contrast enhancement improves the representation of /eh/-like vowels in the hearing-impaired auditory nerve," J. Acoust. Soc. Am. **106**, 2693–2708.
Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788–1803.
Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.
Sachs, M. B., and Young, E. D. (**1980**). "Effects of nonlinearities on speech encoding in the auditory nerve," J. Acoust. Soc. Am. **68**, 858–875.
Simpson, M., Moore, B. C. J., and Glasberg, B. R. (**1990**). "Spectral enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners," Acta Oto-Laryngol. **469**, 101–107.
Summers, V, and Leek, M. R. (**1994**). "The internal representation of spectral contrast in hearing-impaired listeners," J. Acoust. Soc. Am. **95**, 3518–3528.

The spatial unmasking of speech: Evidence for better-ear listening

Barrie A. Edmonds and John F. Culling
School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff, CF10 3AT, United Kingdom

(Received 12 January 2006; revised 31 May 2006; accepted 23 June 2006)

Speech reception thresholds (SRTs) were measured for target speech presented concurrently with interfering speech (spoken by a different speaker). In experiment 1, the target and interferer were divided spectrally into high- and low-frequency bands and presented over headphones in three conditions: monaural, dichotic (target and interferer to different ears), and swapped (the low-frequency target band and the high-frequency interferer band were presented to one ear, while the high-frequency target band and the low-frequency interferer band were presented to the other ear). SRTs were highest in the monaural condition and lowest in the dichotic condition; SRTs in the swapped condition were intermediate. In experiment 2, two new conditions were devised such that one target band was presented in isolation to one ear while the other band was presented at the other ear with the interferer. The pattern of SRTs observed in experiment 2 suggests that performance in the swapped condition reflects the intelligibility of the target frequency bands at just one ear; the auditory system appears unable to exploit advantageous target-to-interferer ratios at different ears when segregating target speech from a competing speech interferer. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2228573]

PACS number(s): 43.66.Rq, 43.66.Ba, 43.66.Pn [GDK] Pages: 1539–1545

I. INTRODUCTION

Listeners often experience the compelling impression that they can pick up any of a range of voices around them by focusing their attention on the appropriate direction (Cherry, 1953). This introspection is supported by the fact that listeners achieve better recognition of concurrent voices when these voices come from different directions. The phenomenon, known as spatial unmasking, can be measured as the improvement in the masked threshold for the detection or identification of a target sound as a function of the angle of separation, in the horizontal plane, between the target and an interfering sound.

The perceived direction of a sound source in the horizontal plane is largely determined by two binaural cues (Rayleigh, 1876; 1907). Interaural level difference (ILD) refers to the difference in sound level at the two ears caused by different distances to the source (i.e., the inverse-square law) and, at high frequencies, by the occluding effect of the head (i.e., head shadow). Interaural time difference (ITD) refers to the brief time lag in the signal at the more distant ear caused by the finite speed of sound. ITD is the dominant cue to sound direction (Wightman and Kistler, 1992). These cues also contribute to the spatial unmasking of speech (Bronkhorst and Plomp, 1988). However, the fact that sound localization and spatial unmasking exploit the same cues need not imply a common mechanism. Rather, it is thought that the auditory system exploits ITD and ILD independently to bring about monaural and binaural improvements in the target-to-interferer ratio (Bronkhorst and Plomp, 1988; Culling *et al.*, 2006; 2004; Edmonds and Culling, 2005a; Hawley *et al.*, 2004; Zurek, 1993). ILD is mainly associated with monaural gains in audibility at the ear with the better target-to-interferer ratio, while ITD gives rise to improvements in the audibility of a target sound through binaural interaction (Colburn, 1977; Durlach, 1972; Jeffress, 1948; Levitt and Rabiner, 1967; Rabiner *et al.*, 1966).

Zurek's (1993) model of spatial unmasking predicts the effect of spatial separation on speech intelligibility by calculating the target-to-interferer ratio due to head shadow and binaural unmasking across a series of 1/3-octave frequency bands and comparing these estimates against the articulation index (French and Steinberg, 1947). Here, we focus on how Zurek's (1993) model estimates the monaural contribution to speech intelligibility. The total monaural estimate of the model can be defined as either the contribution of the monaural frequency bands providing the better target-to-interferer ratios regardless of ear (which Zurek termed the "better-bands" rule) or the contribution of the monaural frequency bands at the ear providing the better overall target-to-interferer ratio (which Zurek termed the "better-ear" rule). The predicted difference in speech intelligibility for the better-bands and better-ear rules is negligible in normal listening situations, as head shadow will tend to cause one ear to have the better target-to-interferer ratio in all frequency bands. The model is agnostic as to whether the auditory system employs a better-bands rule or a better-ear rule. Nonetheless, the better-bands rule assumes that the human auditory system is free to select monaural information independently from frequency channel to frequency channel across the two ears. However, the authors are unaware of any attempts to empirically test between the predicted effects of the better-ear and better-bands rules on speech intelligibility.

The experiments reported here compare the predictions of the better-bands and better-ear rules by exploring the effect of splitting stimuli spectrally across the two ears using a

paradigm similar to the one described by Edmonds and Culling (2005b). High- and low-frequency bands of target and interfering speech materials were presented using infinite ILDs (i.e., each frequency band of each signal was presented to one ear only) in order to maximize the target-to-interferer ratio across the ears. We reasoned that if listeners really can exploit the better-bands rule then it should be possible to present the high- and low-frequency bands of a target utterance to different ears in the presence of a concurrent interferer without observing any impact on speech intelligibility. However, if listeners are only able to attend to the ear providing the best overall target-to-interferer ratio (i.e., they employ the better-ear rule), then performance could be poorer in this condition than when target and interferer are simply presented to different ears.

II. EXPERIMENT 1

A. Participants

Nine Cardiff University undergraduate students were recruited and awarded course credit in return for their participation. All participants reported normal hearing and spoke English as their first language.

B. Stimuli

Sentences from the MIT recordings of the male speaker CW reading the Harvard Sentence Lists (IEEE, 1969) were used as target items and sentences from the speaker DA (another male speaker from MIT recordings of the Harvard sentence lists) were used as interferer items. These sentence materials were filtered into high- and low-frequency bands using a 512-point, linear-phase, finite impulse response (FIR) filter with very steep (>1000-dB/octave) cutoffs. High- and low-frequency cutoffs spanned a splitting frequency, leaving a spectral notch of 1 equivalent rectangular bandwidth (ERB) (Moore and Glasberg, 1983). These spectral notches were the same as those employed by Edmonds and Culling (2005b): 700–802 Hz (i.e., a splitting frequency of *750 Hz*), 1411–1594 Hz (i.e., a splitting frequency of *1500 Hz*), and 2822–3287 Hz (i.e., a splitting frequency of *3000 Hz*)[1]. These splitting-frequency manipulations allowed the high- and low-frequency regions of the speech materials to be presented in different interaural configurations. Three interaural configurations of target and interferer were used: *monaural*, *dichotic*, and *swapped*. In the monaural configuration the target speech and interferer were both presented to one ear. In the dichotic configuration the high- and low-frequency bands of a target sentence were presented to the left ear and the high- and low-frequency bands of an interfering sentence were presented to the right ear. Last, in the swapped configuration a high-frequency target band and a low-frequency interferer band were presented to the left ear while a low-frequency target band and a high-frequency interferer band were presented to the right ear. Thus, a total of nine conditions (3 interaural configurations × 3 splitting frequencies) was used in this experiment.

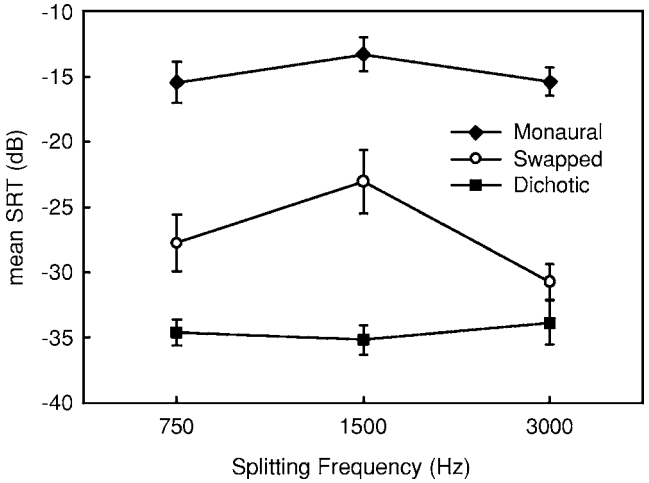


FIG. 1. Mean SRTs of the monaural (diamonds), dichotic (squares), and swapped (circles) conditions as a function of splitting frequency (750, 1500, and 3000 Hz). Error bars show standard error.

C. Procedure

Stimuli were presented to the listener using a TDT AP2 array processor via a TDT psychoacoustics rig (DD1, FT6, PA4, HB6) over Sennheiser HD 590 headphones in a single-walled IAC sound-attenuating booth. Speech reception thresholds (SRTs) were measured using a 1-up/1-down adaptive threshold method (Levitt, 1971) in which a set of ten target sentences was presented against a fixed-level (~70-dB) interferer sentence (a novel interfering sentence was used for each SRT measurement). The initial target-to-interferer ratio was −28 dB. The first stimulus was repeated, each time with a 4-dB-more-intense target sentence, until the listener judged that they could understand half of the sentence and attempted to transcribe it on a computer terminal. When the listener pressed "return" at the end of their transcription, the actual target sentence appeared below the listener's transcript with five keywords in capitals. The listener self-marked his or her identification of the five keywords. If more than two keywords were correctly identified, the level of the subsequent target sentence was reduced by 2 dB. Otherwise it was increased by 2 dB. The mean of the last eight target-to-interferer ratios derived in this way was taken as the SRT.

D. Results and discussion

Figure 1 shows the mean SRTs for the listeners in experiment 1. SRTs were highest (i.e., worst performance) in the monaural condition, lowest (i.e., best performance) in the dichotic condition, and of intermediate level in the swapped condition. SRTs in the swapped condition appeared to vary with splitting frequency, producing a chevron-like pattern of SRTs (i.e., the SRTs were lower for this condition at the 750 and 3000-Hz splitting frequencies than they were when this interaural configuration was presented with a 1500-Hz splitting frequency).

A two-way repeated-measures analysis of variance was performed on the SRTs for all nine conditions (3 interaural configurations × 3 splitting frequencies). There was a significant main effect of interaural configuration [$F(2,16)$

$=114.71$, $p<0.001$] and splitting frequency [$F(2,16)=3.82$, $p<0.05$]. There were no other significant F ratios. There was no significant interaction between interaural configuration and splitting frequency [$F(4,32)=21.20$, $p=0.088$]. Tukey *posthoc* comparisons of interaural configuration confirmed that all three conditions differed significantly from each other [$q>7$, $p<0.001$]. *Posthoc* analysis of splitting frequency revealed that only SRTs in the 1500 and 3000-Hz conditions were significantly different [$q=3.75$, $p<0.05$].

If the monaural advantage described by Zurek (1993) is due to the selection of information using the better-bands rule (i.e., the frequency channel at either ear providing the most favorable target-to-interferer ratio) rather than the better-ear rule (i.e., the ear providing the best target-to-interferer ratio over all frequency channels), then one would expect there to be no difference in the SRTs measured for the dichotic and swapped conditions. However, there was a significant difference between the SRTs of the dichotic and swapped conditions; this result suggested that listeners exploit the better-ear rule rather than the better-bands rule.

Although there was no significant interaction between splitting frequency and interaural configuration in this study, we noticed a chevron-shaped pattern of SRTs in the swapped condition, which suggests that listeners were particularly disadvantaged by the 1500-Hz splitting frequency in this condition. The fact that performance improves when a larger proportion of information is presented to one ear or the other suggests that SRTs in the swapped condition were determined solely by the target information present in the frequency channels at the ear providing the most information about the target. As the 1500-Hz splitting frequency removes target speech that is roughly in the middle of the speech range in informational terms (Fletcher and Galt, 1950), neither ear provides the listener with sufficient information to achieve optimal SRTs. We decided to further investigate the apparent effect of splitting frequency on the intelligibility of speech in the swapped condition in a second experiment.

III. EXPERIMENT 2

This experiment explored the contribution of high- and low-frequency speech bands to speech intelligibility. It was conducted to confirm whether or not listeners combine any information about the target that is presented to different ears or whether they rely solely on the information available at one ear. Two new conditions were designed to yield SRTs that varied systematically with splitting frequency. One ear was designated a "target-only" ear and received either a high-frequency target band (high-contribution condition) or a low-frequency target band (low-contribution condition); the proportion of target speech at the target-only ear in the high-contribution and low-contribution conditions depended on the splitting frequency. The remaining proportion of target speech (e.g., the low-frequency band in the high-contribution condition) was presented to the contralateral ear with the high- and low-frequency interferer bands. We reasoned that the intelligibility of the target speech in these conditions should increase (i.e., yield lower SRTs) as the range of frequencies available at the target-only ear increased. Consequently, if the better-ear rule is employed then one would expect the pattern of SRTs in the swapped condition to closely match the pattern of SRTs found in one or the other of these new conditions at different splitting frequencies. For example, the target-to-interferer ratio at the better ear in the swapped condition might mirror those of the low-contribution condition when the splitting frequency is high (i.e., 3000 Hz) and those of the high-contribution condition when the splitting frequency is low (i.e., 750 Hz). However, if listeners have any ability to exploit the better-bands rule rather than the better-ear rule, then performance in the swapped condition should be better than that achieved for these high-contribution and low-contribution conditions, as more frequency bands will be available to the listener.

A. Participants

A new group of nine undergraduate students from Cardiff University was recruited and rewarded with course credit for their participation. Again, all participants reported normal hearing and spoke English as their first language.

B. Stimuli

The target and interfering speech materials were again presented as high- and low-frequency bands divided at splitting frequencies of *750*, *1500*, and *3000 Hz* in three interaural configurations: *swapped* (the left ear received a high-frequency target band and a low-frequency interferer band; the right ear received a low-frequency target band and a high-frequency interferer band), *high-contribution* (a high-frequency target band was presented to the left ear; the high- and low-frequency interferer bands and a low-frequency target band were presented to the right ear), and *low-contribution* (a low-frequency target band was presented to the right ear; the high- and low-frequency interferer bands and a high-frequency target band were presented to the left ear).

C. Results and discussion

Figure 2 shows how the swapped-condition thresholds relate to the high-contribution and low-contribution condition SRTs. These thresholds vary systematically with increasing splitting frequency in all three conditions; again, a chevron-like pattern of SRTs was observed in the swapped condition. When a splitting frequency of 750 Hz was employed the low-contribution condition SRTs were highest (i.e., intelligibility was poorest), while SRTs in the high-contribution and the swapped conditions were lowest. When a splitting frequency of 1500 Hz was used, SRTs in all three conditions were intermediate and indistinguishable. When the 3000-Hz splitting frequency was employed, SRTs in the low-contribution and swapped conditions were lowest, while those of the high-contribution condition were highest.

A two-way repeated measures analysis of variance was performed on the SRTs, with two within-subjects factors (3 interaural configurations $\times 3$ splitting frequencies). There was a significant main effect of interaural configuration

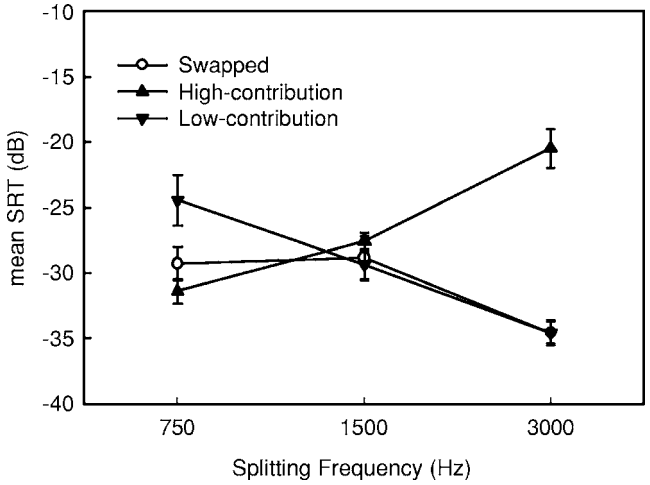


FIG. 2. Mean SRTs of the swapped (circles), low-contribution (upright triangles), and high-contribution (inverted triangles) conditions as a function of splitting frequency (750, 1500, and 3000 Hz). Error bars show standard error.

$[F(2,16)=9.88, p<0.001]$ and a significant interaction between interaural configuration and splitting frequency $[F(4,32)=22.24, p<0.001]$.

Tukey *posthoc* tests for the interaction between interaural configuration and splitting frequency revealed the following differences. For comparisons of interaural configuration within the 750-Hz level of splitting frequency, the low-contribution condition SRTs were significantly different from both the high-contribution condition $[q=5.7, p<0.001]$ and swapped-condition $[q=3.97, p<0.05]$ SRTs; there was no significant difference between the SRTs of the swapped and high-contribution conditions $[p>0.05]$ at this splitting frequency. Comparisons of interaural configuration within the 1500-Hz level of splitting frequency revealed no significant differences between the SRTs of the three interaural configurations. For comparisons of interaural configuration within the 3000-Hz level of splitting frequency, the mean SRT of the high-contribution condition was significantly different from both the swapped and the low-contribution condition SRTs $[q>11.58, p<0.001]$; the SRTs of the swapped and low-contribution conditions were not significantly different $[p>0.05]$ at this splitting frequency.

Comparisons of splitting frequency within interaural configuration revealed the following differences. SRTs for the swapped condition at a splitting frequency of 3000 Hz were significantly different from those at 1500 and 750 Hz $[q>4.3, p<0.05]$. SRTs for the high-contribution condition at a splitting frequency of 3000 Hz were significantly different from those at 1500 and 750 Hz $[q>5.8, p<0.001]$. SRTs for the low-contribution condition at all three splitting frequencies were significantly different from one another $[q>4.0, p<0.05]$.

The overall similarity of mean SRTs in the swapped condition to those observed in the high-contribution and low-contribution conditions at different splitting frequencies suggests that the intelligibility of speech in the swapped condition was dependent upon the frequency bands at the ear providing the better target-to-interferer ratio. We conclude that this effect is consistent with the better-ear rule of listening and not the better-bands rule.

It is tempting to suppose that the high thresholds observed in the low-contribution condition with 750-Hz splitting frequency and the high-contribution condition with 3000-Hz cutoffs may reflect performance based solely on those narrow frequency regions. However, given the restricted range of frequencies available, it is quite possible that the better ear was, in each case, the one with the interfering voice and that these thresholds represent masked thresholds based on a relatively wide frequency band.

IV. GENERAL DISCUSSION

Zurek's (1993) model of spatial unmasking describes how speech intelligibility is dependent on monaural and binaural contributions to the target-to-interferer ratio across a series of frequency bands. We tested, in two experiments, whether the monaural advantage to spatial unmasking reflects an all-or-nothing strategy for exploiting information at one ear or the other, or whether the auditory system is able to take advantage of the ear with the better target-to-interferer ratio within each frequency channel. These experiments were designed as a simple test of the two strategies (i.e., better ear or better bands) that Zurek (1993) described in his model. In experiment 1, speech intelligibility was better in the dichotic condition than in the swapped condition. This result is inconsistent with the better-bands rule, which would predict that thresholds in these two conditions should be identical. Experiment 2 showed that performance in the swapped condition was virtually indistinguishable from the best thresholds observed in the high-contribution and low-contribution conditions at a given splitting frequency; this suggests that speech intelligibility in the swapped condition was dependent on the target frequency channels at just one ear. Consequently, we suggest, these data indicate clear support for a better-ear interpretation of Zurek's model. While this conclusion is not particularly novel (i.e., it only confirms a prediction of Zurek's model), it is most definitely surprising, as a number of related studies suggest that the better-bands rule should have prevailed.

Brungart *et al.* (2005), for instance, used the Zurek's better-bands rule to explain some unexpected effects in their study of precedence-based segregation (see Sec. IV C for further details). They manipulated the extent to which the "echo" of a masking stimulus was delayed and found that this manipulation brought about an improvement in the intelligibility of target speech heard against broadband noise at some delays and not others. After an inspection of their stimuli, they concluded that the addition of a delayed masker at certain delay intervals introduced spectral notches in the spectrum of the masker, which allowed listeners to exploit advantages in the target-to-interferer ratio at the frequencies of these notches. As these notches appeared in different frequency regions in both ears, Brungart *et al.* reasoned that, in order for listeners to take full advantage of these notches, the auditory system might exploit information at the ear providing the most favorable target-to-interferer ratio within each

frequency band (i.e., the better-bands rule). A simulation of Zurek's (1993) model of spatial unmasking confirmed that these gains in intelligibility, afforded by the changes in target-to-interferer ratio at these notches, was consistent with the increases in performance observed for these conditions. It would be interesting to know if the better-ear rule would also have been able to account for the changes in performance that were observed by Brungart *et al.* (2005), as the results of the current investigation (against speech interference) suggest that listeners are only able to exploit notches in the masker spectrum at only one of the ears. Unfortunately, Brungart *et al.* (2005) did not report whether this was the case or not. Moreover, our own data were collected using speech as an interferer, rather than the speech-shaped noise used by Brungart *et al.* Below, we consider how the results of the current dichotic listening experiments relate to other bodies of research in the literature.

A. Band-independent processing of ITD

The idea that the auditory system is able to select and integrate information from different frequency channels has been successfully demonstrated binaurally (Akeroyd, 2004; Culling and Summerfield, 1995; Edmonds and Culling, 2005b). It is somewhat intriguing, therefore, that the results of the current investigation suggest that the auditory system does not employ a similar strategy for monaurally exploiting target-to-interferer ratios. Previously, in an analogous set of experiments to the ones described here, Edmonds and Culling (2005b) looked for evidence of across-channel processing in the spatial unmasking of speech using ITD. Three experiments were designed to test whether the segregation of spatially separated sounds is dependent on the consistency of ITD across different frequency bands, in particular, whether or not the binaural gain in speech intelligibility was constrained to the exploitation of a single ITD across frequency. It was found that, as long as the target and interferer had a different ITD in each frequency channel, speech intelligibility was unaffected. As integration of ITD across frequency is important for localizing sounds, this result suggested that improvements in speech intelligibility for these stimuli reflected the benefit of binaural unmasking and that perceived location was a relatively unimportant cue. The fact that the monaural and binaural components of spatial unmasking display such strikingly different (i.e., better-ear vs band-independent processing) characteristics suggests that the mechanisms underlying the exploitation of ITD and ILD might be separate and distinct.

B. The spectral fusion of sounds across the ears

Early work in the field of dichotic listening (e.g., Broadbent, 1955; Broadbent and Ladefoged, 1957; Cutting, 1976) suggests that listeners can fuse together different spectral regions of a sound source when they are presented separately but simultaneously to opposite ears to form a single auditory image. Broadbent and Ladefoged (1957), for example, presented listeners with a dichotic speech stimulus in which the first and second formants of a synthesized sentence were simultaneously presented to different ears. They found that as long as the two formants shared the same fundamental frequency ($F0$), listeners would describe hearing a single voice in a single location just as they might report for an unmodified stimulus.

While the nature of these spectral fusion experiments is quite different from those of the current investigation, a number of parallels can be drawn. Both sets of experiments require the listener to piece together different parts of a speech stimulus that has been spectrally divided across the ears. In the spectral fusion experiments described above, the speech stimulus is split at a point dividing the first and second formants. Here, a specific splitting frequency divides the speech into two halves; although this manipulation does not take into account the formant structure of the stimuli, $F0$ information is preserved. Thus, although the target speech is spectrally split across the two ears both frequency regions of the target share the same $F0$. Consequently, one might expect spectral fusion to occur for the swapped-target stimuli in the current paper; if this was the case, SRTs should reflect contribution of all target frequency bands and not just those at one ear. Furthermore, predictions about the effect of spatial cues could be based on differences between the perceived locations of the two fused auditory images.

We can make a number of different predictions about how listeners might have performed, in the swapped condition of the current paper, if spectral fusion of swapped speech bands occurred. First, spectral fusion across the two ears might cause listeners to perceive the target and interferer as having the same location (e.g., in the center of the head). If this was the case, thresholds might match those observed in the baseline condition. Second, the fused bands of target speech might be perceived to have a slightly different location from the interferer (e.g., both heard slightly off-center, but in opposite hemispheres). Third, spectral fusion across the ears might result in the target and interferer being perceived to have distinctly different locations (e.g., target at one ear and interferer at the opposite ear). If this was the case, the competing voices should be easy to segregate and thresholds should be similar to those in the dichotic condition. The data are obviously inconsistent with the first and third scenarios, as thresholds in the swapped condition fell somewhere in between those achieved for the baseline and dichotic conditions. The second scenario sounds appealing, at least until one attempts to explain the data of experiment 2 using the same scenario. Using this scenario, the target and interferer of the high- and low-contribution conditions in experiment 2 should be perceived by listeners to be well separated (i.e., the target should be perceived to be off-center and the interferer at the opposite ear). This spatial separation, however, does not provide any additional benefit to the listener; the thresholds of the high- and low-contribution conditions accurately predict performance in the swapped condition. This suggests that the benefit of spectral fusion observed in "quiet" does not extend to more complex listening situations (e.g., involving the segregation of target speech from a concurrent speech interferer), and perhaps indicates a problem with grouping[2] by fundamental frequency ($F0$) across the ears in the swapped condition.

The possibility that failure to group by $F0$ may underlie

the pattern of results would be consistent with the findings of Culling and Darwin (1993). They found that correct identification rates for pairs of simultaneous vowels with different *F*0's was largely unaffected by a manipulation in which the *F*0's of the two vowels were swapped between the first and second formants. They thus found little evidence for across-formant grouping by *F*0. Bird and Darwin (1998) later showed that a similar manipulation of competing sentences was ineffective at disrupting their intelligibility. These experiments both attempted to disrupt performance by confusing the grouping process. The present experiment may offer a complementary form of evidence because, in this case, a lack of across formant grouping by *F*0 may have impaired listeners' ability to perform a task in which such grouping was needed.

C. Spatial release from informational masking

A number of studies in the spatial unmasking literature suggest that attending to perceived location can be particularly important for providing a release from informational masking (Brungart, 2001; Brungart *et al.*, 2001; 2005; Freyman *et al.*, 2001; 2004; 1999; Kidd *et al.*, 2005; 1994). Informational masking is thought to reflect a failure to segregate or a propensity to confuse sounds (especially speech sounds); it is thought to produce an excess of masking (i.e., a rise in the threshold for identifying a target over and above that attributable to energetic masking). The impact of informational masking on the segregation of sound sources can be avoided, however, if the listener is provided with some means to focus their attention on the target item. Freyman *et al.* (1999), for example, demonstrated how perceived location can help listeners focus their attention on the target utterance and obtain a release from the informational masking. In Freyman's study, an illusory spatial separation between target and masker (presented over spatially separated loudspeakers) was achieved with the addition a delayed copy of the masker (i.e., to simulate a room reflection) at the target loudspeaker. The "echo" invoked the precedence effect (Litovsky *et al.*, 1999; Zurek, 1987), which creates the illusion of spatial separation while substantially disrupting the cues for binaural unmasking. Under these circumstances, an improvement in speech intelligibility against speech maskers was observed but not against noise maskers; this effect was characterized as a release from informational masking.

Previously, Edmonds and Culling (2005a; 2005b) found little support for the idea that perceived location is important for spatial unmasking. However, the present paper appears to suggest that listeners could only exploit information at one location (i.e., ear), as listeners were foiled by the swapped condition of experiment 1 (i.e., producing much higher threshold than one might expect if perceived location was not important to spatial unmasking). Given the design of the experiments described here, one could choose to interpret the results of the current paper in terms of release from informational masking. The results of the swapped condition are a prime candidate for such interpretation, as there is no energetic masking in this condition (i.e., the target and interferer are never physically mixed in a given frequency channel). The elevated thresholds in the swapped condition (i.e., compared to those observed in the dichotic condition), for example, could be characterized as perceptual interference in the segregation of the competing speech streams. Alternatively, if one keeps to the low-level (i.e., nonattentional) framework of Zurek's (1993) model, the SRTs in the swapped condition can be explained as the contribution of target frequency bands at each ear to the intelligibility index. That is, the listener simply listens to the ear providing the better level of intelligibility.

V. CONCLUSIONS

When segregating competing voices, listeners do not benefit from favorable target-to-interferer ratios in different frequency bands if these bands are presented to different ears; the SRTs observed in the current investigation appear to have been determined solely by the quality of target information presented to a single ear. In terms of Zurek's (1993) model, this result is more consistent with the better-ear rule than the better-bands rule for monaural listening. Further investigation will be required to determine if this better-ear mechanism is differentially sensitive to different forms of masking, as it is unclear how grouping by fundamental frequency and perceived location, for example, affect this process.

ACKNOWLEDGMENTS

Barrie Edmonds was supported by the UK EPSRC (Grants GR/M96155 and GR/S11794).

[1]The 1-ERB-wide notch was originally introduced by Edmonds and Culling (2005b) in order to reduce the possibility of confounding binaural interactions in frequency channels close to the splitting frequency. However, as the current investigation employs infinite ILDs and not ITDs the spectral notch is not necessary, but was kept for the sake of consistency across these similarly themed research papers.

[2]Grouping is distinct from fusion. Fusion is the binding together of different elements into a single auditory image. Grouping, as the term implies, involves identifying different sound elements as belonging to distinct auditory objects.

Akeroyd, M. A. (**2004**). "The across frequency independence of equalization of interaural time delay in the equalization-cancellation model of binaural unmasking," J. Acoust. Soc. Am. **116**, 1135–1148.

Bird, J., and Darwin, C. J. (**1998**). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, Q. Summerfield, and R. Meddis (Whurr, London), pp. 263–269.

Broadbent, D. E. (**1955**). "A note on binaural fusion," Q. J. Exp. Psychol. **7**, 46–47.

Broadbent, D. E., and Ladefoged, P. (**1957**). "On the fusion of sounds reaching different sense organs," J. Acoust. Soc. Am. **29**, 708–710.

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., Simpson, B. D., and Freyman, R. L. (**2005**). "Precedence-based speech segregation in a virtual auditory environment," J. Acoust. Soc. Am. **118**, 3241–3251.

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, C. (**1953**). "Some experiments on the recognition of speech, with

one and with two ears," J. Acoust. Soc. Am. **25**, 975–979.
Colburn, H. S. (**1977**). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," J. Acoust. Soc. Am. **61**, 525–533.
Culling, J. F., and Darwin, C. J. (**1993**). "Perceptual separation of simultaneous vowels: Within and across-formant grouping by *F*0," J. Acoust. Soc. Am. **93**, 3454–3467.
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.
Culling, J. F., Edmonds, B. A., and Hodder, K. I. (**2006**). "Speech perception from monaural and binaural cues," J. Acoust. Soc. Am. **119**, 559–565.
Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2004**). "The role of head-induced interarual time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **116**, 1057–1065.
Cutting, J. E. (**1976**). "Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening," Psychol. Rev. **83**, 114–140.
Durlach, N. I. (**1972**). "Binaural signal detection: Equalization and cancellation theory," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), pp. 369–462.
Edmonds, B. A., and Culling, J. F. (**2005a**). "The role of head-related time and level cues in the unmasking of speech in noise and competing speech," Acta. Acust. Acust. **91**, 546–553.
Edmonds, B. A., and Culling, J. F. (**2005b**). "The spatial unmasking of speech: Evidence for within-channel processing of interaural time delay," J. Acoust. Soc. Am. **117**, 3069–3078.
Fletcher, H., and Galt, R. H. (**1950**). "The perception of speech and its relation to telephony," J. Acoust. Soc. Am. **22**, 89–151.
French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.
Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.
Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.
Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.
Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.
IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.
Jeffress, L. A. (**1948**). J. Comp. Physiol. Psychol. **41**, 35.
Kidd, G., Mason, C. R., Brughera, A., and Hartmann, W. M. (**2005**). "The role of reverberation in release from masking due to spatial separation of sources for speech identification," Acta. Acust. Acust. **91**, 526–536.
Kidd, G. J., Mason, C. R., Deliwala, P. S., and Woods, W. S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Levitt, H., and Rabiner, L. R. (**1967**). "Predicting binaural gain in intelligibility and release from masking for speech," J. Acoust. Soc. Am. **42**, 820–829.
Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.
Moore, B. C., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitatory patterns," J. Acoust. Soc. Am. **74**, 750–753.
Rabiner, L. R., Laurence, C. L., and Durlach, N. I. (**1966**). "Further results on binaural unmasking and the EC model," J. Acoust. Soc. Am. **40**, 62–70.
Rayleigh, L. (**1876**). "On perception of the direction of a source of sound," Nature (London) **14**, 32–33.
Rayleigh, L. (**1907**). "On our perception of sound direction," Philos. Mag. **8**, 214–232.
Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.
Zurek, P. M. (**1987**). "The precedence effect," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer, New York), pp. 85–105.
Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston), pp. 255–276.

The ability of the parasitoid fly *Ormia ochracea* to distinguish sounds in the vertical plane[a)]

Ben J. Arthur[b)] and Ronald R. Hoy
Department of Neurobiology and Behavior Mudd Hall, Cornell University, Ithaca, New York, 14853

(Received 31 January 2006; revised 22 May 2006; accepted 21 June 2006)

The parasitic fly *Ormia ochracea* localizes its host, field crickets, by homing in on their calling song. Previous phonotactic studies indicate that their sound localization ability in azimuth is extraordinarily acute, but the fly's ability to localize the elevation of sound sources has not been tested to date. Here we show that in a freely-walking closed-loop Y-maze task elevational performance is well above chance, but slightly below the fly's performance in azimuth. Immobilizing the head or the halteres (sensory organs of balance) slightly lowered elevational discrimination, but performance was still well above chance. Because ormiine ears are thought to be symmetric and the pure-tone models of a cricket's call used in these experiments contained little to no spectral bandwidth, additional studies will be needed to elucidate the underlying cues. Nonetheless, it is clear that while walking, *Ormia* flies are capable of distinguishing sound sources that differ only in elevation. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2225936]

PACS number(s): 43.66.Qp, 43.66.Gf, 43.64.Bt, 43.64.Ha [JAS] Pages: 1546–1549

I. INTRODUCTION

Ormia ochracea is an acoustic parasitoid of field crickets. Gravid female flies use calling male crickets as hosts for larvae, finding them in the first place by listening for their mating songs (Cade, 1975). The azimuthal sound localization acuity of tethered female *Ormia* has been demonstrated to be two degrees when walking on a trackball (Mason *et al.*, 2001), which matches both humans (Hartmann and Rakerd, 1989) and barn owls (Knudsen *et al.*, 1979). Such an acuity is surprising given the 0.5 mm distance between the ears, but can be explained through a tympanal hearing organ (Robert *et al.*, 1992; Lakes-Harlan and Heller, 1992) in which the two tympana are mechanically coupled (Miles *et al.*, 1995; Robert *et al.*, 1996). To date, however, little data exist concerning elevational ability. One would think that such an ability might be well developed since flying *Ormia* can localize singing crickets, both in nature (Walker, 1993) and in experimental flight rooms (Ramasauer and Robert, 2000; Müller and Robert, 2001). Moreover, in the frequent case that a gravid *Ormia* alights near the cricket instead of landing directly on it, she is confronted with the need to navigate through a complex microterrain, a task for which elevational discrimination might be essential. This study addresses the elevational ability of *Ormia* through the use of a vertical Y-maze in which flies freely walk towards sound sources varying in elevation.

II. METHODS

Females were tested on both horizontal (azimuthal) and vertical (elevational) Y-mazes for which the angle between the branches was about 85° (Fig. 1). The mazes were constructed of metal window screen mesh and suspended in the air with dental floss and steel weights in a 0.5 m^3 box lined with sound-absorbing foam (Sonex, Illbruck Acoustic, Minneapolis, MN). All three branches of both maze types were 1 cm wide, just wide enough to accommodate the leg span of *Ormia,* and 9 cm long, making the total path length from the base through the bifurcation to the speaker 18 cm. Artificial cricket songs (4800 Hz carrier, 13 ms pulses at 45 Hz, 2 ms rise and 5 ms fall times) were synthesized (Matlab 7.0.1, The MathWorks, Natick, MA) by a computer (Dimension 8200, Dell, Round Rock, TX; 2.4 GHz Pentium 4, Intel, Santa Clara, CA; Windows XP, Microsoft, Redmond, WA) and played (RP2.1, PA5, System 3, Tucker Davis Technologies, Alachua, FL) from one of two miniature speakers (MDR-ED228, Sony; PM655, Harmon Kardon) placed at the end of each branch. The sound level was adjusted to be 80 dB SPL (re. 20 μPa) at the base of the Y and checked to make sure it was free of any harmonic distortion (4134 1/2 in. condenser microphone, 2608 measuring amplifier, 4220 pistonphone, Brüel and Kjær, Nærum, Denmark). Gravid flies were taken from our laboratory colony that originated from wild flies caught in Bradenton, Florida (Wineriter and Walker, 1990) and allowed to acclimate to the box for 5 min before testing began. Prior to this acclimatization period two of the five experimental groups had either the heads or halteres immobilized with wax under cold anaesthesia (5 min on ice; 5 min recovery). Several test trials were run in white light to show flies that there was a path on which they could walk towards the cricket song. For these trials the placement of the flies on the maze was initially on the arms of the Y near the speakers and then over the course of several trials the placement was gradually moved back through the bifurcation towards the base of the Y. Twenty real trials were then performed in red light (590–838 nm, Sylvania, Danvers, MA) to quantify the phonotactic ability of the fly. Two scoring methods were used: a strict criterion counted as correct only those trials for

[a)]This work was presented in "Sound localization ability of the parasitoid fly *Ormia ochracea* in the elevational plane," International Congress of Neuroethology, Nyborg, Denmark, August 2004.

[b)]Electronic mail: bja28@cornell.edu

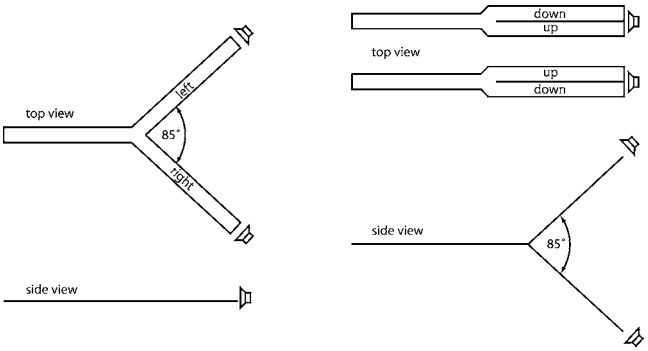


FIG. 1. The three mazes used in these experiments. (Left) The azimuthal maze. (Right) The two elevational mazes. Note that elevational mazes were *not* simply azimuthal mazes rotated 90°. Rather the walking surface of the initial branch of both maze types was normal to gravity. Whereas the two bifurcating branches for azimuthal mazes were splayed 85° in the horizontal plane, in the elevational mazes these two branches were side-by-side when viewed from above and spread 85° in the vertical plane. Two chiral elevational mazes were used to control for the confound between azimuth and elevation due to the asymmetry at the bifurcation. Speakers for the elevational mazes were placed on the midline so as to minimize any azimuthal cues.

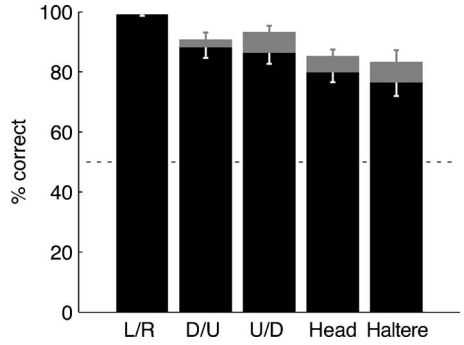


FIG. 2. Localization performance in five experimental conditions using two scoring methods. Results for the azimuthal maze are shown in the left bar, and for the elevational mazes in the four bars on the right. Black indicates the mean percentage of correct trials as judged by the strict criterion, and grey indicates the additional trials scored as correct by the more relaxed criterion (see Methods). Horizontal dashed line indicates the chance level at 50%. All conditions were significantly above chance; the D/U and U/D elevational conditions were not significantly different than each other; azimuthal performance was significantly above performance in the combined D/U and U/D elevational conditions; and the Head and Haltere conditions were on the borderline of being significantly different from the combined D/U and U/D conditions. Ten separate flies were used in each condition with 15–20 trials for each fly. Percent correct was computed for each fly, and then averaged across flies to get the mean and standard error bars.

which the flies went straight to the correct speaker, stepping for no more than one body length on the wrong branch; a more relaxed criterion counted as correct the additional trials for which the flies initially chose the wrong branch at the bifurcation in the Y and then, before reaching the terminus, turned around and went toward the correct speaker. Flies that were unresponsive to cricket song, repeatedly flew off the maze, or showed a consistent bias towards one side were not included in the analysis. Performance differences between groups were quantified by a binomial ANOVA done in R (R Foundation for Statistical Computing, Vienna, Austria). The null hypothesis that all groups had equal means was compared to a model in which the two chiral elevational mazes were combined but the others were left to vary, as well as one in which all groups could vary. For both the strict and relaxed criteria, the null hypothesis was rejected ($p<1e\text{-}8$) and treating the two elevational groups separately did not account for significantly more of the variance ($p>0.4$). Comparisons with chance were then calculated using the combined model and omitting the intercept term, and comparisons between groups by including it, with the resulting p values being reported in the next section. A small subset of flies, not included in the analysis of performance, was videotaped using either conventional (DCR-TRV38, Sony) or high-speed (250 fps, MotionScope HR1000, Redlake Camera) cameras. Taping was not done routinely because of concerns about the sound field, the limited angular resolution of the information gained, and increased experimental difficulty due to the flies' tendency to hide in the many facets of the camera if they flew off the maze.

III. RESULTS

The data presented here consist of the five following experimental conditions: (1) an azimuthal maze (L/R); (2+3) two chirally nonequivalent elevational mazes, a left-handed (D/U) and right-handed (U/D) form; (4) an elevational maze with the fly's head waxed to its thorax (Head); and (5) an elevational maze with the fly's halteres waxed to the calypters (Haltere). The azimuthal task was performed to compare our current experimental paradigm with previous work (Mason *et al.*, 2001), and the latter two conditions served to search for potential mechanisms by which *Ormia* might detect elevational cues. The heads were immobilized specifically because ormiine ears are located on the thorax in the narrow cleft separating it from the head, and so even small movements of the head could effect the sound field surrounding the ears. The halteres were immobilized because they were observed to oscillate in response to auditory stimulation in tethered walking flies (personal observation; data not shown), and it was desired to see if this was necessary for successful discrimination.

Ten separate flies were used in each of the five conditions (50 flies total) and 15–20 trials were run on each fly (992 trials total). As judged by the strict criterion, performance was significantly above chance in all conditions (Fig. 2; $p<0.00001$). While performance for the azimuthal maze was near perfect at 99±0.6%, the average for the two elevational mazes, which were not significantly different than each other ($p>0.5$), dropped significantly compared to azimuthal performance to 87±1.9% ($p<0.005$). Immobilizing the head lowered elevational performance to 80±3.2%, but not significantly ($p>0.05$). Immobilizing the halteres significantly lowered elevational performance to 76±3.5% ($p<0.01$).

In the azimuthal trials, flies typically oriented (yawed) their body towards the sound as they ran along the base of the Y towards the bifurcation, indicating that the location of the sound was perceived well before a decision had to be made. A similar change in the angular pitch of the thorax was sometimes observed in elevational trials, although this observation was far less robust. High-speed video was used to ascertain whether flies might rotate their bodies to convert

world-centered elevation into body-centered azimuth, or rotate their heads relative to their thorax to induce a change in the sound field near the tympana on the thorax. While small rolls of the thorax were observed to naturally occur during the normal walking gate, no pronounced or obviously intentional rolls of a great magnitude were observed. Nor was the head ever observed to make large rolls independently of the thorax except while grooming between trials. This last observation is consistent with the results described above of only a minimal decrement in performance when the heads were waxed to the thorax.

When tested in each of the four elevational conditions, a portion of the initial wrong decisions made by the fly at the bifurcation were corrected before the terminus of the incorrect branch was reached. Even when these trials are scored as correct using the more relaxed criterion, the results presented above are similar: (a) performance is significantly above chance for all conditions ($p<0.00001$), (b) azimuthal performance ($99\pm0.6\%$) is slightly greater than elevational performance ($92\pm1.5\%$; $p=0.021$), and (c) immobilizing the head or the halteres slightly lowered elevational performance ($85\pm2.8\%$, $p=0.042$ and $83\pm3.0\%$, $p=0.013$, respectively). Note that points (b) and (c) are on the borderline of being significant, having p values between 0.01 and 0.05.

IV. DISCUSSION

The data presented here show that phonotaxis towards synthetic cricket song by freely walking flies is at performance levels well above chance in both the elevational and azimuthal planes. Azimuthal performance was significantly, albeit slightly, better than elevational performance, possibly due to a left/right asymmetry in the design of the vertical Y mazes: whereas the flies' tendency to rotate their bodies towards the speaker naturally guided them through the bifurcation in the azimuthal task, the vertical mazes did not permit this because they had the up and down branches to the left and right of each other. The ability of flies to localize in elevation at all is intriguing because ormiine ears are thought to be symmetric and the artificial cricket song used here was tonal. Hence neither the interaural intensity difference cues used by barn owls due to their asymmetric ears (Kaup, 1862; Payne, 1971; Olsen *et al.*, 1989) nor the monaural spectral cues used by humans for broadband noise stimuli (Angell and Fite, 1901; Roffler and Butler, 1967) are available to flies as elevational cues.

To investigate one possible strategy that might enable *Ormia* to localize elevationally distinct sources, an experimental manipulation of waxing the head to the thorax was performed. Given that ormiine ears are in the cleft separating the head from the thorax, head movements relative to the thorax could affect the sound field. It has long been known that flies move their heads during tethered visual fixation tasks (Land, 1973; Geiger and Poggio, 1977), and recently head movements have also been shown to occur in free-flight during body saccades (Schilstra and van Hateren, 1998). The purpose is thought to be the stabilization of gaze to minimize visual blur and enhance object recognition. Head movements can be stimulated by other modalities as well, including tactile stimulation of the halteres (Sandeman and Markl, 1980) and auditory stimulation (personal observation). The latter was observed during presentation of synthetic cricket calls while tethered *Ormia* were walking on a trackball. Though modifying the sound field through head movements could be one strategy *Ormia* use to overcome the limitations of symmetric ears and a tonal stimulus, the data here show that they are not necessary, as elevational performance is only slightly degraded when such movements are prevented.

A second manipulation performed here was preventing the movement of the halteres by waxing them to the calypters. The halteres are a pair of oscillating pendulums protruding from the thorax just posterior to the wings which are thought to act as sensory gyroscopes to stabilize flight (Derham, 1713; Fraenkel and Pringle, 1938). Halteres also flap when not in flight however (Fraenkel, 1939; Schneider, 1953; and personal observation in *Ormia*), and it has been hypothesized that such oscillations sense angular accelerations of the surface that flies are walking or resting on (Miller, 1977). The data of Sandeman and Markl (1980) address this point by showing that ablating the halteres has a significant effect on a walking fly's ability to compensate for rotations of teeter-totters and turntables. The effect was so "slight," however, that the authors conclude that sensory input from haltere oscillations are predominantly used during flight. Because it is still essentially an open question then as to why halteres oscillate while flies are not flying, we immobilized them during a portion of our elevational trials. The results reported here confirm that halteres are not necessary for normal walking behavior in that the effect on discrimination performance was small. It seems likely then that the slight decrement in performance both in our data, as well as in Sandeman and Markl (1980), is perhaps merely due to the short-term effects associated with haltere immobilization or removal (e.g., novelty of immobility, surgical trauma, cold anaesthesia). Were this true, one would expect performance difficulties to be abated if more time were allotted to the recovery period, as Fraenkel (1939) found for haltereless flight endurance, spontaneous take off, and sensitivity to startle.

In summary, the Y-maze data presented here show successful phonotaxis towards sounds in the elevational plane by freely-walking *Ormia ochracea* even when their head and halteres are immobilized. Further experiments will be necessary, however, to determine the underlying cues used. Open-loop behavioral studies, head-related transfer function (HRTF) measurements, and dichotic stimuli would in particular facilitate a direct comparison with strategies used by other species.

ACKNOWLEDGMENTS

Thanks to James Booth for help with the statistics, to Cole Gilbert for help with the haltere literature, to Bob Wyttenbach and one anonymous reviewer for commenting on the manuscript, and to the National Institute on Deafness and Other Communication Disorders (NIDCD) for funding.

Angell, J. R., and Fite, W. (**1901**). "The monaural localization of sound," Psychol. Rev. **8**, 225–245.
Cade, W. (**1975**). "Acoustically orienting parasitoids: Fly phonotaxis to cricket song," Science **190**, 1312–1313.
Derham, W. (**1713**). *Physico-Theology*, 1st ed. (W. Innys, London), p. 406.
Fraenkel, G. (**1939**). "The function of the halteres of flies (diptera)," Proc. Zool. Soc. A **109**, 69–78.
Fraenkel, G., and Pringle, J. W. S. (**1938**). "Halteres of flies as gyroscopic organs of equilibrium," Nature (London) **141**, 919–920.
Geiger, G., and Poggio, T. (**1977**). "On head and body movements of flying flies," Biol. Cybern. **25**, 177–180.
Hartmann, W. M., and Rakerd, B. (**1989**). "On the minimum audible angle: A decision theory approach," J. Acoust. Soc. Am. **85**, 2031–2041.
Kaup, J. J. (**1862**). "Monograph of the strigidae," Trans. Zool. Soc. Lond. **4**, 201–260.
Knudsen, E. I., Blasdel, G. G., and Konishi, M. (**1979**). "Sound localization by the barn owl (*Tyto alba*) measured with the search coil technique," J. Comp. Physiol. [A] **133**, 1–11.
Lakes-Harlan, R., and Heller, K. G. (**1992**). "Ultrasound-sensitive ears in a parasitoid fly," Naturwiss. **79**, 224–226.
Land, M. F. (**1973**). "Head movement of flies during visually guided flight," Nature (London) **243**, 299–300.
Mason, A. C., Oshinsky, M. L., and Hoy, R. R. (**2001**). "Hyperacute directional hearing in a microscale auditory system," Nature (London) **410**, 686–690.
Miles, R. N., Robert, D., and Hoy, R. R. (**1995**). "Mechanically coupled ears for directional hearing in the parasitoid fly *Ormia ochracea*," J. Acoust. Soc. Am. **98**, 3059–3069.
Miller, P. L. (**1977**). "Haltere activity in a flightless hippoboscid fly, *Crataerina pallida*," J. Insect Physiol. **23**, 855–860.
Müller, P., and Robert, D. (**2001**). "A shot in the dark: The silent quest of a free-flying phonotactic fly," J. Exp. Biol. **204**, 1039–1052.
Olsen, J. F., Knudsen, E. I., and Esterly, S. D. (**1989**). "Neural maps of interaural time and intensity differences in the optic tectum of the barn owl," J. Neurosci. **9**, 2591–2605.
Payne, R. S. (**1971**). "Acoustic location of prey by barn owls (*Tyto alba*)," J. Exp. Biol. **54**, 535–573.
Ramsauer, N., and Robert, D. (**2000**). "Free-flight phonotaxis in a parasitoid fly: Behavioural thresholds, relative attraction and susceptibility to noise," Naturwiss. **87**, 315–319.
Robert, D., Amoroso, J., and Hoy, R. R. (**1992**). "The evolutionary convergence of hearing in a parasitoid fly and its cricket host," Science **258**, 1135–1137.
Robert, D., Miles, R. N., and Hoy, R. R. (**1996**). "Directional hearing by mechanical coupling in the parasitoid fly *Ormia ochracea*," J. Comp. Physiol. [A] **179**, 29–44.
Roffler, S. K., and Butler, R. A. (**1967**). "Factors that influence the localization of sound in the vertical plane," J. Acoust. Soc. Am. **43**, 1255–1259.
Sandeman, D. C., and Markl, H. (**1980**). "Head movements in flies (*Calliphora*) produced by deflexion of the halteres," J. Exp. Biol. **85**, 43–60.
Schilstra, C., and van Hateren, J. H. (**1998**). "Stabilizing gaze in flying blowflies," Nature (London) **395**, 654.
Schneider, G. (**1953**). "Die halteren der schmeissfliege (*Calliphora*) als sinnesorgane und als mechanische flugstabilisatoren," Zeitschrift fur Vergleichende Physiologie **35**, 416–458.
Walker, T. J. (**1993**). "Phonotaxis in female *Ormia ochracea* (diptera: Tachinidae), a parasitoid of field crickets," J. Insect Behav. **6**, 389–410.
Wineriter, S. A., and Walker, T. J. (**1990**). "Rearing phonotactic parasitoid flies [*Diptera*: *Tachinidae*, *Ormiini*, *Ormia spp*.]," Entomophaga **35**, 621–632.

Underwater auditory localization by a swimming harbor seal (*Phoca vitulina*)

Anais Bodson, Lars Miersch, Bjoern Mauck, and Guido Dehnhardt[a)]
General Zoology & Neurobiology, University of Bochum, ND 6/33, D-44780 Bochum, Germany

(Received 7 April 2006; revised 23 May 2006; accepted 12 June 2006)

The underwater sound localization acuity of a swimming harbor seal (*Phoca vitulina*) was measured in the horizontal plane at 13 different positions. The stimulus was either a double sound (two 6-kHz pure tones lasting 0.5 s separated by an interval of 0.2 s) or a single continuous sound of 1.2 s. Testing was conducted in a 10-m-diam underwater half circle arena with hidden loudspeakers installed at the exterior perimeter. The animal was trained to swim along the diameter of the half circle and to change its course towards the sound source as soon as the signal was given. The seal indicated the sound source by touching its assumed position at the board of the half circle. The deviation of the seals choice from the actual sound source was measured by means of video analysis. In trials with the double sound the seal localized the sound sources with a mean deviation of 2.8° and in trials with the single sound with a mean deviation of 4.5°. In a second experiment minimum audible angles of the stationary animal were found to be 9.8° in front and 9.7° in the back of the seal's head. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2221532]

PACS number(s): 43.66.Qp, 43.80.Lb [WWA] Pages: 1550–1557

I. INTRODUCTION

Although pinnipeds do not possess an active sonar system like toothed whales, underwater acoustic information might play an important role in the life of these aquatic mammals, especially during passive listening. For example, vocalizations are largely used for communication, particularly in the context of reproduction. Male Weddell seals (*Leptonychotes weddellii*) appear to use visual and acoustic displays to claim three-dimensional underwater territories around breathing holes (Bartsh *et al.*, 1992). For harp seals (*Phoca groenlandica*) it is suggested that their calls serve to promote herd formation and courtship (Serrano and Terhune, 2002; Watkins and Schevill, 1979). In leopard seals (*Hydrurga leptonyx*), broadcast calls are used underwater for long-distance communication by mature females, e.g., to advertise their sexual receptivity, and possibly by mature males while searching for mates (Rogers *et al.*, 1996). As in territorial male bearded seals (*Erignathus barbatus*) the characteristic trills are significantly longer than in roamers, it is suggested that trill duration may be a useful indicator of male quality. Territorial male bearded seals may be successful individuals while roaming males may be younger animals or males in poorer condition that are unable to maintain an aquatic territory (Van Parijs *et al.*, 2003; Cleator and Stirling, 1990).

Male harbor seals (*Phoca vitulina*) produce simple stereotyped roar vocalizations. The function of these vocalizations appears to be primarily advertising the presence of a male in breeding condition, and has therefore been suggested to be used in male-male competition and/or as a reproductive advertisement display to attract females (Bjørgesaeter *et al.*, 2004; Hanggi and Schusterman, 1994; Hayes *et al.*, 2004; Van Parijs *et al.*, 1999, 2000a, b). In addition to intraspecific communication, underwater sounds may also be important for detecting soniferous prey. In fact, many fish species are known to produce a variety of sounds. For example, cod produce click sounds (Vester *et al.*, 2004) and Atlantic herring produce broadband pulses (1.7–22 kHz) lasting between 0.6 and 7.6 s (Wilson *et al.*, 2003), and these sounds are well within the detection capabilities of some marine predators. Utilization of prey sounds for hunting has been shown for several fish species (Myrberg, 1981), and some cetaceans have been documented to use passive listening during the search phase of the foraging process (Barrett-Lennard *et al.*, 1996; Gannon *et al.*, 2005; Tyack and Clark, 2000). Thus it seems plausible that foraging pinnipeds could make use of these sounds as well. On the other hand, pinnipeds may also acoustically detect, localize, and avoid predators such as killer whales (Deecke *et al.*, 2002) and sharks. Furthermore, usage of underwater sound for spatial orientation and navigation purposes in pinnipeds should be considered. As suggested by Norris (1967), it is conceivable that sounds emerging from abiotic sources, as for example the sound of breakers near a shore or a reef, could be used as acoustic landmarks while traveling along a coastline. A study concerning under-ice navigation showed that success in locating breathing holes by a blindfolded spotted seal (*Phoca largha*) was much improved by the addition of an acoustic cue at a designated hole (82% success versus 19% without acoustic cue) (Sonafrank *et al.*, 1983). For all these reasons, the ability of seals to detect, accurately determine, and finally attain the position of underwater sounds in the wild certainly has a significant impact on an individual's fitness. However, there is a gap in our knowledge concerning sound localization abilities as existing data cannot be simply applied to the case of moving subjects.

It is common practice to describe the sound localization acuity of a stationary subject in terms of the minimum audible angle (MAA), which is the smallest detectable differ-

a)Author to whom correspondence should be addressed.

ence between the azimuths of two identical sound sources (Mills, 1958), with smaller MAAs representing better localization ability. The MAA is one half of the threshold angular separation with the sound sources positioned symmetrically about the subject's midsaggital plane (Moore and Au, 1975). In air, the localization of sound involves mainly two types of binaural cues: interaural time and phase differences and interaural level differences (Mills, 1958; Yost, 2000). Interaural time and phase differences are suggested to dominate the localization of low-frequency tones (below 1.5 kHz), whereas, due to the shadowing effect of the head, interaural intensity differences are prevalent for the localization of higher frequencies (above 1.5 kHz) (Mills, 1958). The application of aerial localization models to an underwater context in research on humans (Feinstein, 1973; Hollien, 1973; Wells and Ross, 1980) revealed that both cues are greatly reduced in a liquid medium and that the 1.5-kHz crossover frequency separating the two types of interaural differences is higher underwater [about 6 kHz (Bovet *et al.*, 1998; Hollien *et al.*, 1986).

While auditory localization capabilities of terrestrial mammals have been studied in many species, little work has been done on marine mammals, especially underwater. Gentry (1967) determined the MAA of a California sea lion (*Zalophus californianus*) and obtained 15° for a 3.5-kHz and 10° for a 6.5-kHz pure tone (63% correct response level). For the same species, Moore (1975) determined the MAA using a click train signal with the peak energy of each click at 0.9 kHz. Depending on the threshold criterion (63% and 75% correct responses), MAAs were 6° and 9°, respectively. In a follow-up study, Moore and Au (1975) measured the MAA of a California sea lion as a function of frequency (0.5 to 16 kHz in 1-oct increments). The smallest MAA of 4° was obtained at 1 kHz. In a Northern fur seal (*Callorhinus ursinus*) Babushina and Poliakov (2004) determined an MAA of 6.5°–7.5° for pulsed pure tone signals and 3° for pulsed noise signals (75% correct response level). Møhl (1964) studied underwater sound localization in the harbor seal (*Phoca vitulina*) and reported an MAA of 3.1° for a 2-kHz pure tone signal at a 75% correct response level. However, using a click train, Terhune (1974) obtained a MAA for the same species of 9° (±4°). Presented with a 10-kHz pure tone, the sound localization of the animal tested by Terhune (1974) was very poor.

However, all studies on sound localization of pinnipeds were conducted with stationary animals, while in the wild the above-mentioned application scenarios such as acoustically guided localization of conspecifics or auditory detection of prey certainly occur mainly while seals are swimming. Furthermore, up to now tests on sound localization were confined to signals in the frontal portion of the horizontal plane. No study on pinnipeds has yet explored sound localization at posterior and lateral directions in the horizontal plane. Therefore, the main objective of the present study was to determine the acuity of sound localization of a harbor seal in the horizontal plane at anterior, posterior, and lateral positions while the animal was swimming a straight course (experiment I). Additionally, underwater MAAs for sounds presented in front (defined as "frontal MAA") as well as in the back of the head (defined as "backward MAA") of the stationary test animal (experiment II) were determined as a baseline.

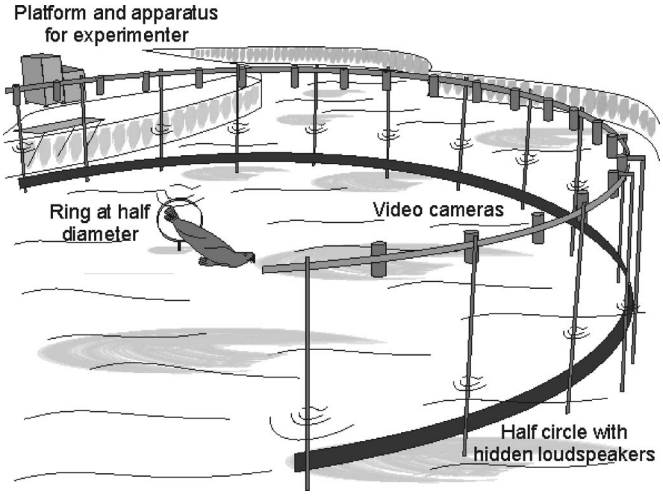


FIG. 1. Experimental setup used to test the underwater auditory localization of a swimming harbor seal.

II. MATERIALS AND METHODS

A. Subject

The experimental animal (Bill) was a 7-year-old male harbor seal. The subject was kept with seven other seals at our marine mammal research lab, Zoo Cologne (Germany), since 2001. It previously served as a subject in studies on olfaction (Kowalewsky *et al.*, 2006) and vision (Weiffen *et al.*, 2006). In the present study, the animal received approximately 80%–90% of its daily diet (3–4 kg of mixed herring and sprat) during experimental sessions. Tests were carried out once per day, on 5 to 6 days per week.

B. Experimental setup

In experiment I, the underwater auditory localization acuity in the swimming seal was determined. Testing was conducted in a 10-m-diam horizontal half circle arena (Fig. 1). The half circle consisted mainly of a black polyethylene board (25 cm high, 8 mm thick) installed 80 cm below the water surface on stainless steel feet fixed to the ground of the pool. Every 5° of the board's horizontal mid-line, small holes (2.5 cm in diameter) were drilled. A continuous ribbon of waterproof adhesive tape (5 cm wide) occluded all holes along the entire inner side of the board. At the upper edge of the board, 200 identical yellow marks indicated single degree steps (1° = 8.7 cm) of the half circle.

Six speakers (JSN 12 Neodymium Dome Tweeter) covered by a thin layer of silicone were used as signal transducers. They were installed at the outer side of the half circle, each behind one of the holes of the board. The transducers and their respective electric cables were hidden in a black tarpaulin at the outer side of the board in order to exclude any visual cues.

A stainless construction exactly following the underwater half circle was mounted 1.5 m above the upper edge of

the board (Fig. 1). Six small cameras (Conrad, $40 \times 40 \times 27$ mm^3, 20 g, 12 V, 291 000 pixels) were attached to the flat bar of the stainless construction, each of them above one of the six transducers. Each camera was housed in a PVC cup and covered an angle of 10° (87 cm) on both sides of the respective speaker. In order to exclude any association between camera position and sound source, 20 camera dummies, consisting of identical but empty PVC cups, were mounted together with the real camera housings uniformly distributed along the flat bar construction. This experimental setup allowed transducers and cameras to be mounted at any place between 0° and 180°, with 0° located in front of the subject's midline and 90° to the left of the subject's midline. The positions of transducers, cameras, and dummies were manually changed between experimental sessions.

Cameras and transducers were connected via cables to a selector switch and then to a computer. The computer-generated signal was routed to the selector switch that allowed one of the six connected speakers and its related camera to be manually selected by the experimenter. During data collection, a monitor showing the picture of the selected camera allowed the experimenter to observe and judge the animal's response. Additionally, video records of the experimental sessions were analyzed offline frame by frame. All electronic instruments (computer, monitor, keyboard, recorder, and speaker/camera selector) were housed in a waterproof cupboard ($0.9 \times 0.7 \times 0.4$ m^3) placed at the edge of the pool. During tests the experimenter was concealed by the open doors of the cupboard so that unintentional cuing was excluded.

At the edge of the experimental pool, right in front of the cupboard housing the electronics, a start station for the seal, consisting of a small plastic ball, was mounted 15 cm above the water surface. At the other side of the pool, opposite to the start station, a goal station was mounted 80 cm below the water surface. The distance between start and goal station was 13 m along the diameter of the half circle. On the way between start and goal a 1-m-diam hoop was mounted underwater on the bottom of the pool, 1 m in front of the center point of the half circle (Fig. 1). This guiding hoop ensured that the animal passed the exact center point of the half circle and that it swam at a depth of 80 cm, so that stimuli were projected at its ear level.

Finally, in order to measure the swimming speed of the animal and to analyze its complete trajectory during underwater sound localization, some experimental sessions were also recorded with a camera (Canon XL1) mounted at the top of a 6-m-high mast, thereby covering the entire area of the half circle.

After the experiment on sound localization with the swimming seal was completed, frontal and backward MAAs of the stationary test animal were determined underwater using a classical psychophysical procedure (experiment II). Testing occurred in a modified version of the experimental setup described above. Definitions concerning the half circle, the transducers, and the sound production are identical as above. However, the start and the goal stations as well as the guiding hoop were removed. A small hoop station exactly fitting the animal's muzzle was placed at the center point of the half circle (i.e., 5-m distance to the sound sources). When the seal placed its muzzle in the hoop station, its head was fixed underwater at a depth of 80 cm. Additionally, the seal was trained to press the tip of its lower jaw on a knob screwed on a jib that was welded to the lower edge of the hoop, so that head movements were restricted. Two response targets were mounted on both sides of the lower jaw station. For backward MAA tests, the stationing device together with the response targets was turned by 180°.

C. Stimuli

In experiment I, two different signals were used. The first signal, defined as double sound, consisted of two 6-kHz pure tones each lasting 0.5 s (50-ms rise and fall times) separated by an interval of 0.2 s (total signal duration 1.2 s). The second signal used (single sound) was a 6-kHz pure tone lasting 1.2 s (50-ms rise and fall times).

Using a calibrated hydrophone (Brüel & Kjaer, type 8104), the average sound pressure level of stimuli presented from different test angles was determined at the center of the half circle. Differences of up to 6 dB in sound pressure level were observed between stimuli projected from different transducer locations during acoustic mapping. Therefore, the position of all transducers was periodically switched between sessions, so that any differences in sound intensity between transducers was not systematically correlated with a position in the horizontal plane. Thus, the subject could not use intensity differences between speakers or speaker positions to improve performance.

In experiment II, the signal was the double sound already used during experiment I (6-kHz pure tone lasting 1.2 s including an interval of 0.2 s, 50-ms rise and fall times).

D. Procedure

In experiment I, a trial started with the seal waiting at the start station. Following a "go" command from the experimenter, the animal immediately swam along the diameter of the half circle at a depth of 80 cm in the direction to the guiding hoop. As soon as its head was in the center of the half circle, the signal was given. The animal was trained to change instantly its course towards the sound source and to indicate it by touching its assumed position on the inner surface of the board at the half circle. The seal's correct responses were reinforced differentially according to its accuracy in three predefined categories. Calculated from the midpoint of the half circle a response made with a deviation of 0° up to 2° to the right or left side of the actual sound location (i.e., 0 up to 17.4 cm from the respective transducer at the board) was reinforced by three pieces of cut herring, a deviation of 3° up to 6° was rewarded by two pieces of cut herring, and a deviation of 7° up to 10° was rewarded by one piece of cut herring. Responses deviating more than 10° from the actual sound location were defined as errors and were not reinforced. To maintain the seal's motivation to follow its linear course as long as there is no sound signal, five or six signal absent trials were conducted randomly within experimental sessions consisting of 36–41 trials each.

In these trials the animal had to continue on its straight course until it arrived at the goal station. Correct responses in signal absent trials were reinforced by two pieces of cut herring.

After a short training including only two sound sources, 13 new positions (between 0° and 180°, in 15° increments) were tested. The order of presentation was predetermined and pseudo-random (limit of three consecutive sounds from the same position). Thirty trials at each test azimuth (390 trials) were collected for each stimulus.

In experiment II, a left/right forced choice procedure was used to determine MAAs of the same subject, first for sounds presented in the front and then for sounds presented in the back. The seal entered the test area, swam to the hoop station, and placed its muzzle firmly at the lower jaw station. When the animal was stationed this way, the signal was given. The seal was required to respond to a sound from its subjective right or left by leaving the hoop station and pushing the right or left response target, respectively. Each correct response was rewarded by a piece of cut herring. Prior to testing, the seal was required to demonstrate its ability to perform the task by responding correctly on at least 90% of the trials at angles wider than 20°. The subject achieved this performance criterion by the fifth training session (each 40 trials per session).

The six underwater speakers were positioned at 5°, 10°, 15°, or 20° on either side of a point directly ahead of the animal's station. Within any one session, three speakers were placed to the right and three speakers were placed to the left of the subject's mid-line with a probability of left and right trial presentations of 0.5. The order of presentation was predetermined and pseudo-random (limit of three consecutive sounds from the same position). For the frontal MAA, 42 trials at each test azimuth were collected in 12 sessions, while for the backward MAA 50 trials at each test azimuth were collected in 10 sessions.

III. RESULTS

For experiment I the underwater auditory localization performance of the swimming harbor seal for the two different stimuli (double and single sound) is presented in Fig. 2.

Presented with the double sound, the swimming seal was able to localize the 13 tested positions with a mean deviation from the actual sound source of 2.8°. However, ANOVA analysis showed that precision of sound localization differed for the 13 positions tested [$F(12,372)=7.44$; $p<0.001$]. The seal's smallest deviation from the actual sound source at the half circle was 1.5° at position 150°. The seal's largest deviation from the actual sound source was 5.4° at position 180°. During the 390 trials with the double sound presented from the 13 positions at the half circle, the animal made only five errors, two at 0°, two at 30°, and one at 75°. This corresponds to an overall performance of 98% correct localizations.

Presented with the single sound, the swimming seal was able to localize the 13 tested positions with a mean deviation from the actual sound source of 4.5°. Again, ANOVA analysis showed that precision of sound localization differed for the 13 positions tested [$F(12,330)=12.44$; $p<0.001$]. The seal's smallest deviation from the actual sound source at the half circle was 2.3° at position 180°. The seal's largest deviation from the actual sound source was 7.5° at position 30°. During the 390 trials with the single sound presented from the 13 positions at the half circle, the animal made 47 errors, 13 at 60°, 11 at 120°, 10 at 30°, 5 at 75°, and 2 at 150°, 45°, 15°, and 0°, respectively. This corresponds to an overall performance of 87% correct localizations.

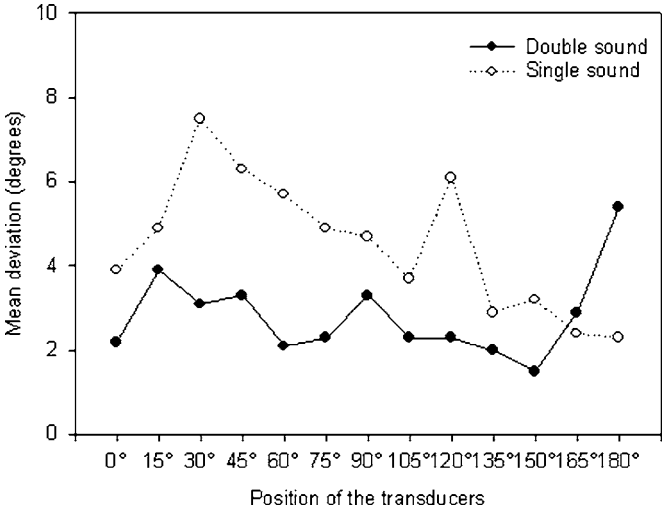


FIG. 2. Mean deviation of the animal's choice at the board of the half circle as a function of the position of the transducers.

With regard to the accuracy of sound localization, there was a significant effect of the type of stimulus used (double versus single sound) on the seal's mean deviation from the actual sound sources [$F(1,325)=89.364$; $p<0.001$]. From the 13 positions tested 11 were better localized with the double sound.

Analyses of the swimming speed (from video recordings) of the seal suggested that irrespective of the type of stimulus it accelerated right after stimulus presentation. The seal needed less time swimming the 5 m after the signal was given (2.01 s, swimming speed 2.48 m/s) than for the 5 m before signal presentation (2.85 s, swimming speed 1.75 m/s). Consequently, the signal was off 2 m before the seal arrived at the board of the half circle.

Frame-by-frame analysis of video recordings of 120 trials revealed that irrespective of the type of stimulus the seal did not approach the actual sound source on a straight course after presentation of a stimulus. Instead, it first swam for more than half of its course (3.5±0.5 m) heading towards an area located 15°–20° to the left side of the actual sound source (Fig. 3). Only when the signal was off, the animal corrected its course heading towards the actual sound source. To describe this swimming pattern, we suggest the term "double trajectory." It was observed in 70% of the trials, and it occurred at all positions tested but more often at those located in front of the seal.

In experiment II, 42 trials at each test azimuth (−20°, −15°, −10°, −5°, 5°, 10°, 15°, 20°) were conducted to determine the frontal MAA of our stationary harbor seal and 50 trials at each test azimuth were conducted to determine the

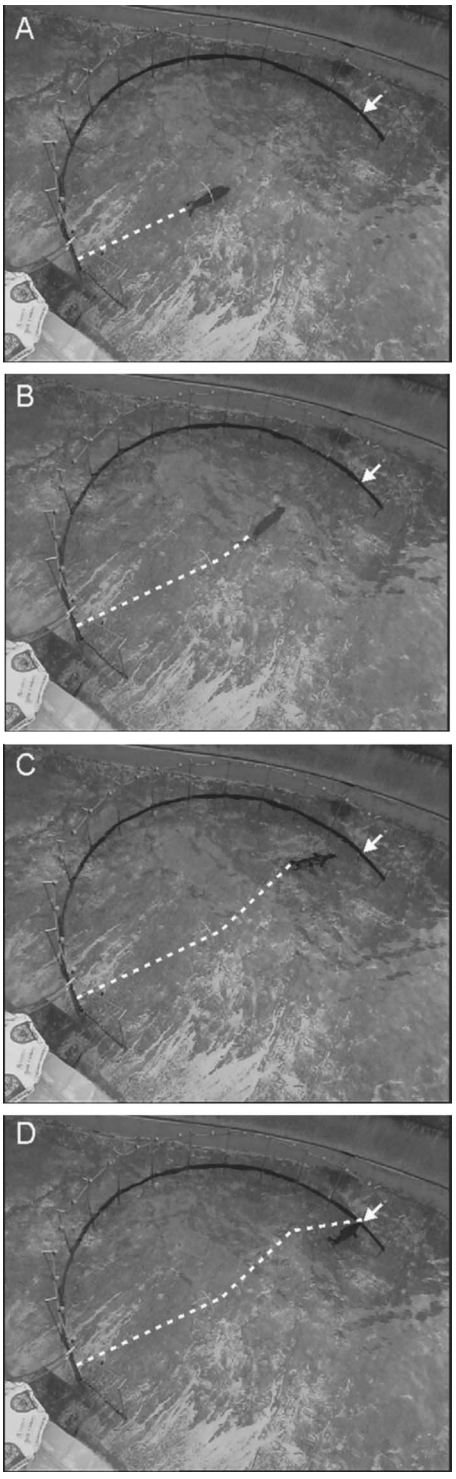


FIG. 3. Single frames from video recordings illustrating the double trajectory when the sound was presented at position 0° (white arrow). (a) The seal swims through the guiding hoop and the signal is given. (b) The seal changes its course heading towards a position located 20° to the left side of the actual sound source. (c) The seal changes its course a second time, now heading towards the actual sound source. (d) The seal touches the inner surface of the half circle at the exact position of the sound source.

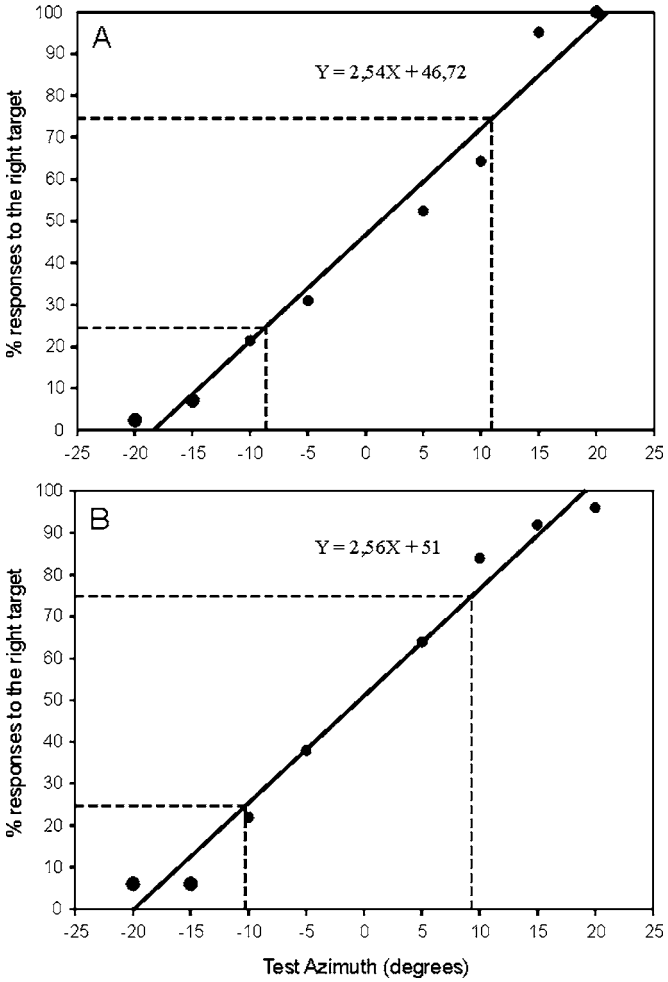


FIG. 4. The seal's auditory localization performance in the MAA experiments as a function of transducer azimuth: (a) frontal MAA and (b) backward MAA. For test angles presented to the right of the seal's midline, correct responses correspond directly to the percentage of right target responses while for left angles (negative angles on the x axis), correct responses are 100% minus the percentage of right target responses on the plots. Linear regression analyses on right target responses were used to find 75% correct decisions (i.e., 75% responses to the right target for right angles and 25% for left angles, indicated by the dashed lines), which were averaged to calculate the MAAs.

backward MAA. For each of the eight tested angles the percentage of correct choices was converted into percentage of responses to the right target (Holt *et al.*, 2004, 2005; Mills, 1958). This means that for the four test angles to the right side of the seal's mid-line the percentage of right target responses corresponds to the percentage of correct choices. For the four test angles to the left side of the seals mid-line the percentage of right target responses corresponds to 100% minus the percentage of correct choices (Fig. 4). The exact MAA was determined by linear regression. The threshold defined as 75% correct choices corresponds to 75% right target responses for sound presentations from the right side of the seals mid-line and 25% right target responses for sound presentations from the left side of the seals mid-line. Averaging these two threshold values resulted in a frontal MAA of 9.8° and a backward MAA of 9.7°.

IV. DISCUSSION

The present study provides first experimental evidence that a swimming harbor seal is able to localize and accurately attain hidden underwater sound sources. There is no indication that the seal was guided by other than acoustical information. There was a wide distribution of deviations from the

exact locations of actual sound sources, which would have not been expected if the seal had used, e.g., visual cues. Furthermore, it was never observed that the seal actively searched for the hidden loudspeakers, suggesting that the seal was simply not aware of the sound-emitting devices. Thus the seals impressive accuracy of sound localization has to be explained based merely on appropriate usage of acoustic information.

Several authors suggested that head movements should improve sound localization (Richardson *et al.*, 1995; Thurlow *et al.*, 1967; Yost and Dye 1997). A sound from the front or back of a receiver both produce no interaural time and level differences since the sound arrives at each ear at the same time and with the same level. By turning the head, interaural differences are changing and will be different for sounds in the front and in the rear (Thurlow *et al.*, 1967). Human divers, e.g., have been observed to perform frequent head movements when tested on an underwater navigational task (Wells and Ross, 1980). The swimming seal in our experiment I was also free to position its head and thus its ears in relation to the sound source, a tactic it could not use during the MAA experiment. In fact, the duration of the acoustic signal was intentionally chosen to be long enough to allow the seal performing head movements while swimming. However, instead of changing frequently its head position while approaching the sound source, video analyses showed that our test animal used a different behavioral tactic. It did not swim directly to the sound source but first swam heading towards an area located 15°–20° to the left side of the actual sound source. Compared to an animal swimming straight ahead towards a sound source this behavior resulted in an increased angle between the sound source and the subject's mid-line and, thus, led to more salient binaural cues. A signal duration of 1.2 s allowed the swimming seal to use more salient binaural cues to cover a distance of about 3 m towards the board. To finally attain the sound source the seal only had to orientate the last 2 m without any acoustic cue. Using a shorter signal, the seal would have had to cover a longer distance without the help of an acoustic cue, which may have resulted in a lower sound localization acuity. However, under natural conditions, potential prey fish like herring produce bursts of pulses lasting up to 7.6 s (Wilson *et al.*, 2003). These long-lasting sounds should permit a seal to navigate acoustically over considerable distances and thus probably represent prominent foraging cues.

During experiment I, the underwater sound localization of the free-ranging seal was more accurate and subject to fewer mistakes with the double sound (mean deviation of 2.8° and 5 mistakes) than with the single sound (mean deviation of 4.5° cm and 47 mistakes). Because testing was conducted with the double sound first, a possible learning effect can be excluded. Both stimuli used had the same total duration; however, it is feasible that the double sound was better localized by the animal because it functioned as a double cue. While swimming, the seal received the first part of the signal at the center point of the half circle, started to change its direction, and then received the second part of the signal while already approaching the sound source. It is plausible that pulsed sounds provide the opportunity to perceive interaural time differences repeatedly, thus the double sound might allow the animal to adjust its swimming pattern more precisely due to a second sound arrival. If in comparison to a single signal a repeated signal generally improves sound localization of a free ranging seal, the fast repetitive tick sounds recorded from herring (Wilson *et al.*, 2003) as well as the double/triplet clicks from cods recorded in the presence of seals (Vester *et al.*, 2004) can be considered to be highly salient cues.

The underwater MAAs obtained in this study (9.8° and 9.7°) are similar to the MAA of the harbor seal studied by Terhune (9° ± 4°), suggesting normal directional hearing of our experimental animal. Compared with the small MAA of a bottlenose dolphin [3.6° with a 6-kHz pure tone (Renaud and Popper, 1975)], underwater directional hearing of harbor seals and most other pinnipeds has been considered to be "mundane at best" (Schusterman *et al.*, 2004). However, it compares well to or is even better than suggested by aerial MAAs of terrestrial predators like cats [8° at 500 Hz and 25° at 4 kHz (Casseday and Neff, 1973)], European polecats [17° (Kavanagh and Kelly, 1987)], and least weasels [12° (Heffner and Heffner, 1987)]. Our experiments with the swimming seal suggest that together with appropriate behavioral tactics a MAA around 10° is sufficient to enable free-ranging animals to perform highly accurate sound localizations. Thus, the MAA as the sole physiological measure of directional hearing is not sufficient to fully describe the sound localization capability of a species.

The practical functions of underwater hearing in free-ranging pinnipeds are largely uninvestigated up to now (Schusterman *et al.*, 2004). However, for most marine mammals passive listening is suggested to be involved in prey location, predator avoidance, and navigation (Tyack and Clark, 2000). The study of Gannon *et al.* (2005) showed that bottlenose dolphins use passive listening extensively during the search phase of the foraging process. Once the dolphins discovered prey by passive listening, they appeared to use echolocation to track it during pursuit. Guinet (1992) hypothesized that killer whales also localize prey by passive listening. Considering the sounds recorded from different fish species belonging to the prey spectrum of seals (Vester *et al.*, 2004; Wilson *et al.*, 2003), passive listening can be considered to be a foraging strategy in pinnipeds as well. A herring swarm, e.g., 25 m away from a seal often cannot be localized visually because of darkness and/or turbidity of the water (Weiffen *et al.*, 2006). Furthermore, visual fields of the harbor seal do not provide circumferential visibility (Hanke *et al.*, 2006) so that potential prey fish simply might be out of the boundaries of the visual fields of the swimming seal. If in this situation herring start to produce sounds as long as about 7.6 s, a seal could swim in the direction of the sound source. For a distance of, e.g., 25 m and swimming speed of 2.5 m/s (as measured in our study) the seal could cover a distance of 19 m in that time, i.e., more than 75% of the initial distance separating it from the herring school. Arriving this way in the proximity of the prey fish, visibility might be sufficient to detect its exact position or the seal encounters the hydrody-

namic trail left by swimming fish by means of its vibrissae (Dehnhardt *et al.*, 2001), both allowing the seal to pursue and finally capture fish.

The method used in this study offers a new approach to investigate underwater auditory localization in aquatic species, focusing on the usage of sensory information. Current studies involve underwater playback of soniferous prey and vocalizations of conspecifics and will provide further insight into the ability of seals to localize sounds in their environment.

ACKNOWLEDGMENTS

We thank Gunther Nogge, Zoo Cologne, Germany, for supporting this study. This research was funded by grants of the VolkswagenStiftung to G.D.

Babushina, E. S., and Poliakov, M. A. (**2004**). "The underwater and airborne sound horizontal localization by the northern fur seal," (in Russian), Biofizika **49**, 723–726.

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.

Bartsh, S. S., Johnston, S. D., and Siniff, D. B. (**1992**). "Territorial behavior and breeding frequency of male Weddell seals (*Leptonychotes weddelli*) in relation to age, size, and concentrations of serum testosterone and cortisol," Can. J. Zool. **70**, 680–692.

Bjørgesaeter, A., Ugland, K. I., and Bjørge, A. (**2004**). "Geographic variation and acoustic structure of the underwater vocalization of harbor seal (*Phoca vitulina*) in Norway, Sweden and Scotland," J. Acoust. Soc. Am. **116**, 2459–2468.

Bovet, P., Drake, C., Bernaschina, F., and Savel, S. (**1998**). "Underwater sound localization: role of interaural differences," J. Acoust. Soc. Am. **103**, 2844.

Casseday, J. H., and Neff, W. D. (**1973**). "Localization of pure tones," J. Acoust. Soc. Am. **54**, 365–372.

Cleator, H. J., and Stirling, I. (**1990**). "Winter distribution of bearded seals (*Erignathus barbatus*) in the Penny Strait Area, Northwest Territories, as determined by underwater vocalisations," Can. J. Fish. Aquat. Sci. **47**, 1071–1076.

Deecke, V. B., Slater, P. J. B., and Ford, J. K. B. (**2002**). "Selective habituation shapes acoustic predator recognition in harbour seals," Nature (London) **420**, 171–173.

Dehnhardt, G., Mauck, B., Hanke, W., and Bleckmann, H. (**2001**). "Hydrodynamic trail-following in harbor seals (*Phoca vitulina*)," Science **293**, 102–104.

Feinstein, S. (**1973**). "Acuity of the human sound localisation response underwater," J. Acoust. Soc. Am. **53**, 393–399.

Gannon, D. P., Barros, N. B., Nowacek, D. P., Read, A. J., Waples, D. M., and Wells, R. S. (**2005**). "Prey detection by bottlenose dolphins, *Tursiops truncatus*: an experimental test of the passive listening hypothesis," Anim. Behav. **69**, 709–720.

Gentry, R. L. (**1967**). "Underwater auditory localization in the California sea lion (*Zalophus californianus*)," J. Aud Res. **7**, 187–193.

Guinet, C. (**1992**). "Hunting behavior in killer whales (*Orcinus orca*) around the Crozet Islands," Can. J. Zool. **70**, 1656–1667.

Hanggi, E. B., and Schusterman, R. J. (**1994**). "Underwater acoustic displays and individual variation in male harbour seals, *Phoca vitulina*," Anim. Behav. **48**, 1275–1283.

Hanke, W., Römer, R., and Dehnhardt, G. (**2006**). "Visual fields and eye movements in a harbor seal (*Phoca vitulina*)," Vision Res. **46**, 2804–2814.

Hayes, S. A., Kumar, A., Daniel, P. C., Mellinger, D. K., Harvey, J. T., Southall, B. L., and LeBoeuf, B. J. (**2004**). "Evaluating the function of the male harbour seal, *Phoca vitulina*, roar through playback experiments," Anim. Behav. **67**, 1133–1139.

Heffner, R. S., and Heffner, H. E. (**1987**). "Localization of noise, use of binaural cues, and a description of the superior olivary complex in the smallest carnivore, the least weasel (*Mustela nivalis*)," Behav. Neurosci. **101**, 701–708.

Hollien, H. (**1973**). "Underwater sound localization in humans," J. Acoust. Soc. Am. **53**, 1288–1293.

Hollien, H., Hicks, J. W., and Klepper, B. (**1986**). "An acoustic approach to diver navigation," Undersea Biomed. Res. **13**, 111–128.

Holt, M. M., Schusterman, R. J., Southall, B. L., and Kastak, D. (**2004**). "Localization of aerial broadband noise by pinnipeds," J. Acoust. Soc. Am. **115**, 2339–2345.

Holt, M. M., Schusterman, R. J., Kastak, D., and Southall, B. L. (**2005**). "Localization of aerial pure tones by pinnipeds," J. Acoust. Soc. Am. **118**, 3921–3926.

Kavanagh, G. L., and Kelly, J. B. (**1987**). "Contribution of auditory cortex to sound localization by the ferret (*Mustela putorius*)," J. Neurophysiol. **57**, 1747–1766.

Kowalewsky, S., Dambach, M., Mauck, B., and Dehnhardt, G. (**2006**). "High olfactory sensitivity for dimethyl sulphide in harbour seals," Biol. Lett. **2**, 106–109.

Mills, A. W. (**1958**). "On the minimum audible angle," J. Acoust. Soc. Am. **30**, 237–246.

Møhl, B. (**1964**). "Preliminary studies on hearing in seals," Vidensk. Medd. Dansk. Naturh. Foren. **127**, 283–294.

Moore, P. W. B. (**1975**). "Underwater localization of click and pulsed pure tone signals by the California sea lion (*Zalophus californianus*)," J. Acoust. Soc. Am. **57**, 406–410.

Moore, P. W. B., and Au, W. L. (**1975**). "Underwater localization of pulsed pure tones by the California sea lion (*Zalophus californianus*)," J. Acoust. Soc. Am. **58**, 721–727.

Myrberg, A. A. (**1981**). "Sound communication and interception in fishes," in *Hearing and Sound Communication in Fishes*, edited by W. N. Tavolga, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 395–425.

Norris, K. S. (**1967**). "Some observations on the migration and orientation of marine mammals," in *Animal Orientation and Navigation*, Proceedings of the 27th Annual Biology Colloquium (Oregon State U. P., Corvallis, OR), pp. 101–131.

Renaud, D. L., and Popper, A. N. (**1975**). "Sound localization by the bottlenose porpoise *Tursiops truncatus*," J. Exp. Biol. **63**, 569–585.

Richardson, W. J., Greene, C. R., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).

Rogers, T. L., Cato, D. H., and Bryden, M. M. (**1996**). "Behavioral significance of underwater vocalizations of captive leopard seals, *Hydrurga leptonyx*," Marine Mammal Sci. **12**(3), 414–427.

Schusterman, R. J., Kastak, D., Levenson, D. H., Reichmuth, C., and Southall, B. L. (**2004**). "Pinniped sensory systems and the echolocation issue," in *Echolocation in Bats and Dolphins*, edited by J. A. Thomas, C. Moss, and M. Vater (Univ. of Chicago, Chicago), pp. 531–535.

Serrano, A., and Terhune, J. M. (**2002**). "Antimasking aspects of harp seal (*Pagophilus groenlandicus*) underwater vocalizations," J. Acoust. Soc. Am. **112**, 3083–3090.

Sonafrank, N., Elsner, R., and Wartzok, D. (**1983**). "Under-ice navigation by the spotted seal, *Phoca largha*," Abstract. Fifth Biennial Conf. on the Biol. of Mar. Mammals, Boston, November 1983.

Terhune, J. M. (**1974**). "Directional hearing of a harbour seal in air and water," J. Acoust. Soc. Am. **56**, 1862–1865.

Thurlow, W. R., Mangels, J. W., and Runge, P. S. (**1967**). "Head movements during sound localization," J. Acoust. Soc. Am. **42**, 489–493.

Tyack, P. L., and Clark, C. W. (**2000**). "Communication and acoustic behavior of dolphins and whales," in *Springer Handbook of Auditory Research; Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York), pp. 156–224.

Van Parijs, S. M., Hastie, G. D., and Thompson, P. M. (**1999**). "Geographic variation in temporal and spatial vocalization patterns of male harbour seals in the mating season," Anim. Behav. **58**, 1231–1239.

Van Parijs, S. M., Hastie, G. D., and Thompson, P. M. (**2000a**). "Individual and geographical variation in display behaviour of male harbour seals in Scotland," Anim. Behav. **59**, 559–568.

Van Parijs, S. M., Janik, V. M., and Thompson, P. M. (**2000b**). "Display area size, tenure length, and site fidelity in the aquatically mating male harbour seal, *Phoca vitulina*," Can. J. Zool. **78**, 2209–2217.

Van Parijs, S. M., Lyderson, C., and Kovacs, K. M. (**2003**). "Vocalizations and movements suggest alternative mating tactics in male bearded seals," Anim. Behav. **65**, 273–283.

Vester, H. I., Folkow, L. P., and Blix, A. S. (**2004**). "Click sounds produced by cod (*Gadus morhua*)," J. Acoust. Soc. Am. **115**, 914–919.

Watkins, W. A., and Schevill, W. E., (**1979**). "Distinctive characteristics of underwater calls of the harp seal (*Phoca groenlandica*) during the breed-

ing season," J. Acoust. Soc. Am. **66**, 983–988.
Weiffen, M., Möller, B., Mauck, B., and Dehnhardt, G. (**2006**). "Effect of water turbidity on the visual acuity of harbour seals (*Phoca vitulina*)," Vision Res. **46**, 1777–1783.
Wells, M. J., and Ross, H. (**1980**). "Distortion and adaptation in underwater sound localization," Aviat., Space Environ. Med. **51**, 767–774.
Wilson, B., Batty, R. S., and Dill, L. M. (**2003**). "Pacific and Atlantic herring produce burst pulse sounds," Proc. R. Soc. London, Ser. B **7**, 95–97.
Yost, W. A. (**2000**). *Fundamentals of Hearing: An Introduction*, 4th ed., (Academic, San Diego, CA).
Yost, W. A., and Dye, R. H. (**1997**). "Fundamentals of directional hearing," Semin. Hear. **18**, 321–344.

The influence of subglottal acoustics on laboratory models of phonation

Zhaoyan Zhang,[a] Juergen Neubauer, and David A. Berry
UCLA School of Medicine, 31-24 Rehabilitation Center, 1000 Veteran Avenue, Los Angeles, California 90095-1794

(Received 13 May 2005; revised 20 June 2006; accepted 20 June 2006)

Many previous laboratory investigations of phonation involving physical models, excised larynges, and *in vivo* canine larynges have failed to fully specify the subglottal system. Many of these same studies have reported a variety of nonlinear phenomena, including bifurcations (e.g., various classes of phonation onset and offset, register changes, frequency jumps), subharmonics, and chaos, and attributed such phenomena to the biomechanical properties of the larynx. However, such nonlinear phenomena may also be indicative of strong coupling between the voice source and the subglottal tract. Consequently, in such studies, it has not been clear whether the underlying mechanisms of such nonlinear phenomena were acoustical, biomechanical, or a coupling of the acoustical and biomechanical systems. Using a physical model of vocal fold vibration, and tracheal tube lengths which have been commonly reported in the literature, it is hypothesized and subsequently shown that such nonlinear phenomena may be replicated solely on the basis of laryngeal interactions with the acoustical resonances of the subglottal system. Recommendations are given for ruling out acoustical resonances as the source of nonlinear phenomena in future laboratory studies of phonation. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2225682]

PACS number(s): 43.70.Aj, 43.70.Bk [AL] Pages: 1558–1569

I. INTRODUCTION

For years, it has been known that the vocal folds may exhibit a wide range of source-resonator coupling with the sub- and supra-glottal systems during different types of phonation. While strong source-resonator coupling is understood to be the norm for many musical instruments (Wilson and Beavers, 1974; Fletcher, 1993), weak source-resonator coupling is generally preferred for the vocal instrument (Titze, 1988a). This is because strong source-resonator coupling tends to be plagued by involuntary voice breaks, or abrupt jumps in fundamental frequency and register. Fortunately, the voice source is relatively independent of acoustic resonances for weak source-resonator coupling, which facilitates continuous, fine control of fundamental frequency and register.

Even for human phonation, source-resonator coupling tends to be strengthened at high frequencies, especially when the fundamental frequency approaches the first formant (Titze, 1988a; Joliveau *et al.*, 2004). Other variables which tend to strengthen source-resonator coupling include a stiff mucosal cover (Titze, 1988a) and a narrow epilarynx tube (Titze and Story, 1995; Mergell and Herzel, 1997). As a general rule, voice breaks and register transitions may also be indicative of strong source-tract coupling. For example, in the literature, a variety of investigators have proposed an intimate connection between subglottal acoustic resonances and vocal registers (Nadoleczny-Millioud and Zimmerman, 1938; Vennard, 1967; Van den Berg, 1968b; Large, 1972; Austin and Titze, 1997).

To reduce the risk of source-resonator coupling in laboratory models of phonation, relatively long subglottal tube lengths have often been employed. For example, in a physical model of vocal fold vibration, Titze *et al.* (1995) utilized a "10-ft-long (305 cm) helical soft-rubber subglottal tube…designed to place subglottal acoustic resonances well below any vibration frequency." Similarly, Alipour and Scherer (2001) used a subglottal tube length of 98 cm, and Thomson *et al.* (2005) used a subglottal tube length of about 180 cm. For computational models of phonation which allow source-tract coupling, a single tube length of about 17–20 cm is generally employed (Flanagan, 1958; Ishizaka *et al.*, 1976). However, it is even more common for numerical simulations to model the subglottal system as a constant pressure source (Ishizaka and Flanagan, 1972), thus precluding any interactions between the voice source and the subglottal system.

In many excised larynx studies, the subglottal tube length has simply been left unspecified (van den Berg and Tan, 1959; van den Berg, 1968a; Berry *et al.*, 1996; Švec *et al.*, 1999; Jiang *et al.*, 2003; Jiang and Zhang, 2005). Consequently, if interactions between the voice source and the subglottal system were to occur, the results could not necessarily be repeated by independent investigators. These same excised larynx studies reported various nonlinear phenomena, including bifurcations (e.g., various classes of phonation onset and offset, register changes, frequency jumps), subharmonics, and chaos—and such phenomena were attributed to the biomechanical properties of the laryngeal system. However, because such phenomena may also be indicative of strong source-resonator coupling, it may be that many of the nonlinear phenomena reported in these laboratory studies of phonation reflected laryngeal coupling with the acoustical

[a]Electronic mail: zyzhang@ucla.edu

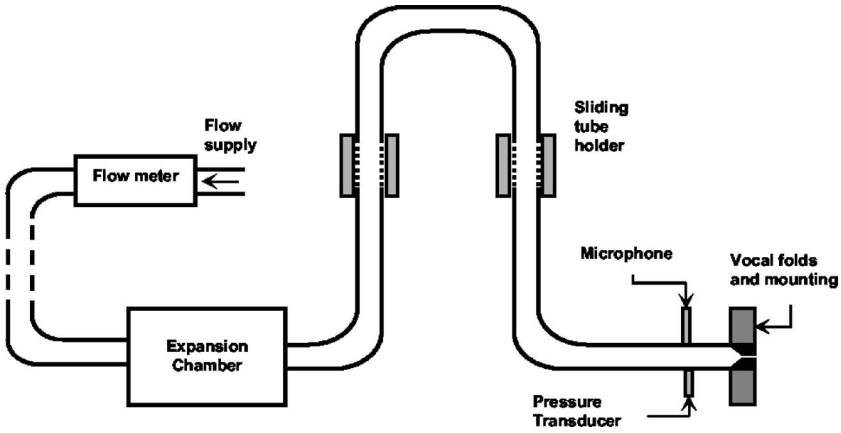


FIG. 1. Schematic of the experimental apparatus.

resonances of the subglottal system more than the biomechanical properties of the larynx. Consequently, in such studies, it was not clear whether the underlying mechanisms of the nonlinear phenomena were acoustical, biomechanical (e.g., cover or body stiffness, vocal fold geometry), or a coupling of the acoustical and biomechanical systems. Moreover, because such laboratory investigations have frequently (1) failed to report the length of the subglottal tube, or (2) utilized subglottal tube lengths considerably longer than that of the human trachea, if acoustical mechanisms were involved, one cannot be certain whether the nonlinear phenomena were reflective of human voice production, at least at the reported, vibrational frequencies. In order to investigate this concern, we hypothesize that many of the nonlinear phenomena reported in previous excised larynx experiments (van den Berg and Tan, 1959; van den Berg, 1968a; Berry *et al.*, 1996; Švec *et al.*, 1999; Jiang *et al.*, 2003; Jiang and Zhang, 2005) may be replicated solely on the basis of laryngeal interactions with the acoustical resonances of the subglottal tube. To explore this hypothesis, we will utilize the physical model of vocal fold vibration recently introduced by Thomson *et al.* (2005) over a range of tracheal tube lengths commonly used in laboratory investigations of phonation.

II. EXPERIMENTAL APPARATUS

A schematic of the experimental apparatus for the phonatory system is shown in Fig. 1. A self-oscillating physical model of the vocal folds (Thomson *et al.*, 2004; 2005) was built and used in this study. This physical model was chosen because of its repeatability and capability of sustaining phonation for long periods of time without significant changes in vibratory behavior. The vocal fold model was molded using a two-component liquid polymer solution mixed with a liquid flexibilizer solution. The stiffness of the physical model could be controlled by varying the mixing ratio of the compound solution. For further details regarding the fabrication and dynamic characteristics of the physical model, the reader is referred to several original papers (Thomson *et al.*, 2004; 2005). The vocal fold model used in this study measured approximately 1 cm in the superior-inferior direction, 1.7 cm in the anterior-posterior direction, and 0.8 cm in the medial-lateral direction. The inferior side of the vocal fold had an entrance convergence angle of 60° from the superior-inferior axis, yielding an inferior-superior thickness of the vocal fold approximately 5.4 mm. The stress-strain relationship of the physical model was measured using the 5544 Instron Testing System (Instron Corp., Norwood, MA). The Young's modulus was estimated to be approximately 11 kPa for strains in the range of 0–20%. The density was about 997 kg/m^3. Two acrylic plates were machined with a rectangular groove on the medial surface of each plate. The vocal fold model was glued into the rectangular grooves on the two plates. The medial surfaces of the two folds were positioned to be in contact so that the glottis was closed when no airflow was applied.

The vocal fold plates were connected downstream from the subglottal system. In humans, the cross-sectional area of the subglottal system stays fairly constant in the trachea and primary bronchi, and then increases abruptly (Weibel, 1963). Various models of the subglottal system have been developed (Ishizaka *et al.*, 1976; Harper *et al.*, 2001; Van de Plaats *et al.*, 2006). As far as the first subglottal resonance is concerned, the subglottal system can be modeled as a single tube about 17–20 cm long, equivalent to the trachea and the primary bronchi, terminated with the lossy compliance of the lungs (Flanagan, 1958; Ishizaka *et al.*, 1976). In this study, the subglottal system consisted of a uniform rigid tube connected to an expansion chamber, simulating the lung compliance. The expansion chamber was connected to the air supply on the other end. The length of the uniform tube was varied in the range between 17 and 325 cm, encompassing the range typical of humans and those used in previous laboratory experiments (e.g., Titze *et al.*, 1995; Vilain *et al.*, 2003; Ruty *et al.*, 2005). This large range of variation was achieved by using a U-shaped circular copper tube (inner diameter of 2.35 cm), which was connected to a 32-cm-long circular PVC tube (inner diameter of 2.54 cm) on the vocal fold end and the expansion chamber on the air supply side. Tube holders (Fig. 1) were designed to allow smooth sliding of the U-shaped tube while maintaining an airtight seal. The combination of the U-shaped tube and the PVC tube allowed the length of the pseudotrachea (and the subglottal acoustics) to be varied systematically over a large range (60–325 cm). Smaller tracheal lengths were achieved without the U-shaped tube by using a 15-cm-long PVC tube with segmental extensions in 5-cm steps. This extended the range of tracheal lengths to about 17 cm on the low end. The bends of the U-shaped tube may have slightly distorted the incoming flow

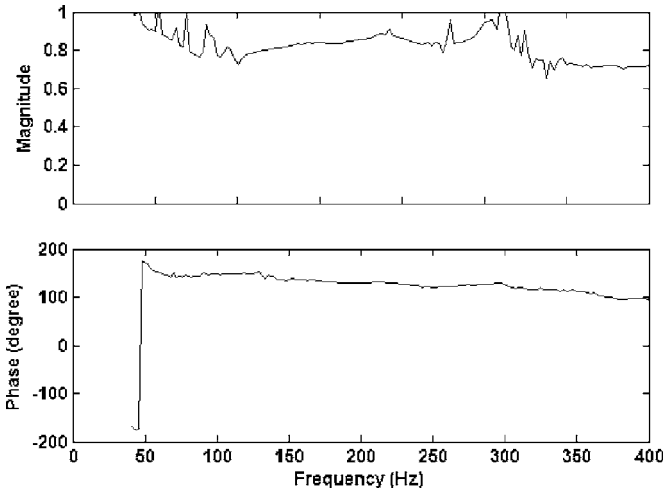


FIG. 2. Magnitude and phase of the reflection factor of the expansion chamber.

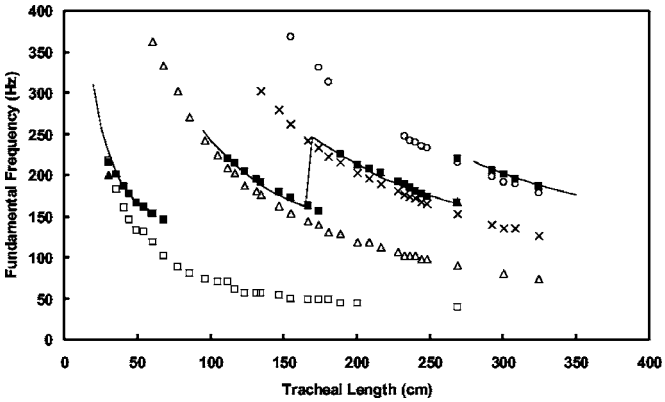


FIG. 3. Measured phonation fundamental frequency as a function of the tracheal length. Closed symbols ■: measured; ▲: second onset frequency measured if more than two frequencies are observed; open symbols: subglottal resonance frequency calculated using the measured reflection factor (□: first resonance; △: second resonance; ×; third resonance; ○; fourth resonance); —: prediction from theory.

distribution inside the tracheal tube. However, this effect was damped by the relatively long PVC tube (more than ten times the tube diameter) downstream, and considered to be negligible.

The expansion chamber was built to simulate the bronchi and the lungs. It had an inner cross-section of 23.5 $\times$25.4 cm and was 50.8 cm long. The inside of the expansion chamber was lined with a 2.54-cm-thick layer of Fiberglass. Cross-sectional area data of the subglottal airway (Ishizaka *et al.*, 1976) showed that the cross-sectional area function from the trachea and the primary bronchi to the lungs increased abruptly by 10–100 times over a distance of about 6 cm. In this setup, the cross-sectional area from the pseudotrachea section to the expansion chamber increased by a factor of about 117. The expansion chamber was connected to the airflow supply through a 15.2-m-long rubber hose, reducing possible flow noise from the air supply. The acoustic characteristics of the expansion chamber and the flow supply were evaluated using the two-microphone method (Seybert and Ross, 1977). Two microphones were mounted along the pseudotracheal tube 5 and 10 cm from the vocal fold plates, respectively. The measured equivalent termination reflection factor is shown in Fig. 2. The magnitude of the reflection factor was very close to one in the frequency range of measurement (40–400 Hz). The phase varied between 100° and 180°. The expansion chamber and upstream system behaved acoustically similar to an ideal open-ended termination to the tracheal tube.

The sound pressure in the tracheal tube was measured using a probe microphone (B&K 4182, Denmark), which was mounted flush with the inner wall of the tracheal tube, 5 cm upstream from the vocal fold plates. A pressure tap was also mounted flush with the inner wall of the tracheal tube, 2 cm upstream from the vocal fold plates. The time-averaged subglottal pressure was measured using a pressure transducer (Baratron type 220D, MKS Instruments, Inc., Andover, MA). The volumetric flow rate through the orifice was measured using a precision mass-flow meter (MKS type 558A, MKS Instruments, Inc., Andover, MA) at the inlet of the setup. Analog-to-digital conversion of the output signals was performed using a United Electronic Industries Powerdaq board (model No. PD2-MFS-8-500/16), with 16 bit resolution over a ±10 V measurement range at a sampling rate of 50 kHz.

During the experiments, the flow rate was increased from zero to a certain maximum value in discrete increments, and then decreased back to zero in discrete decrements. The maximum value of the flow rate varied, being slightly higher than the onset flow rate if phonation onset was observed or around 1800 ml/s when no obvious onset was observed. The increment in flow rate was comparatively large at low flow rates and became smaller when phonation onset was imminent. At each step, measurement was delayed for an interval (about 4–5 s) after the flow rate change, allowing the flow field to stabilize. Sound pressure inside the subglottal tube, flow rate, and subglottal pressure were recorded for a 2-s period. The procedure was repeated over a 17- to 325-cm range of tracheal tube lengths, as previously noted. For some cases, the vibration of the physical model was recorded from a superior view using a high-speed, digital camera (Fastcam-Ultima APX, Photron USA, Inc., San Diego, CA) at a frame rate of 2000 fps and at an image resolution of 1024$\times$1024 pixels.

III. RESULTS

A. Fundamental frequency

The fundamental frequency of vocal fold vibration at onset was measured from the spectra of the sound pressure signal and is shown in Fig. 3 as a function of the tracheal length. The variation of the fundamental frequency with the tracheal length showed a quasi-cyclic pattern, demonstrating a marked dependence on subglottal acoustics (see the Appendix), rather than an exclusive dependence on material properties of the physical model. The fundamental frequency decreased with increasing tracheal length within each cyclic range of the tracheal length in Fig. 3, and then jumped to the highest value of the next cyclic range. Also shown in Fig. 3 is the prediction from a simplified theoretical model of subglottal acoustics (see the Appendix) using the measured reflection factor and model constants [Eq. (A12)]. The theory

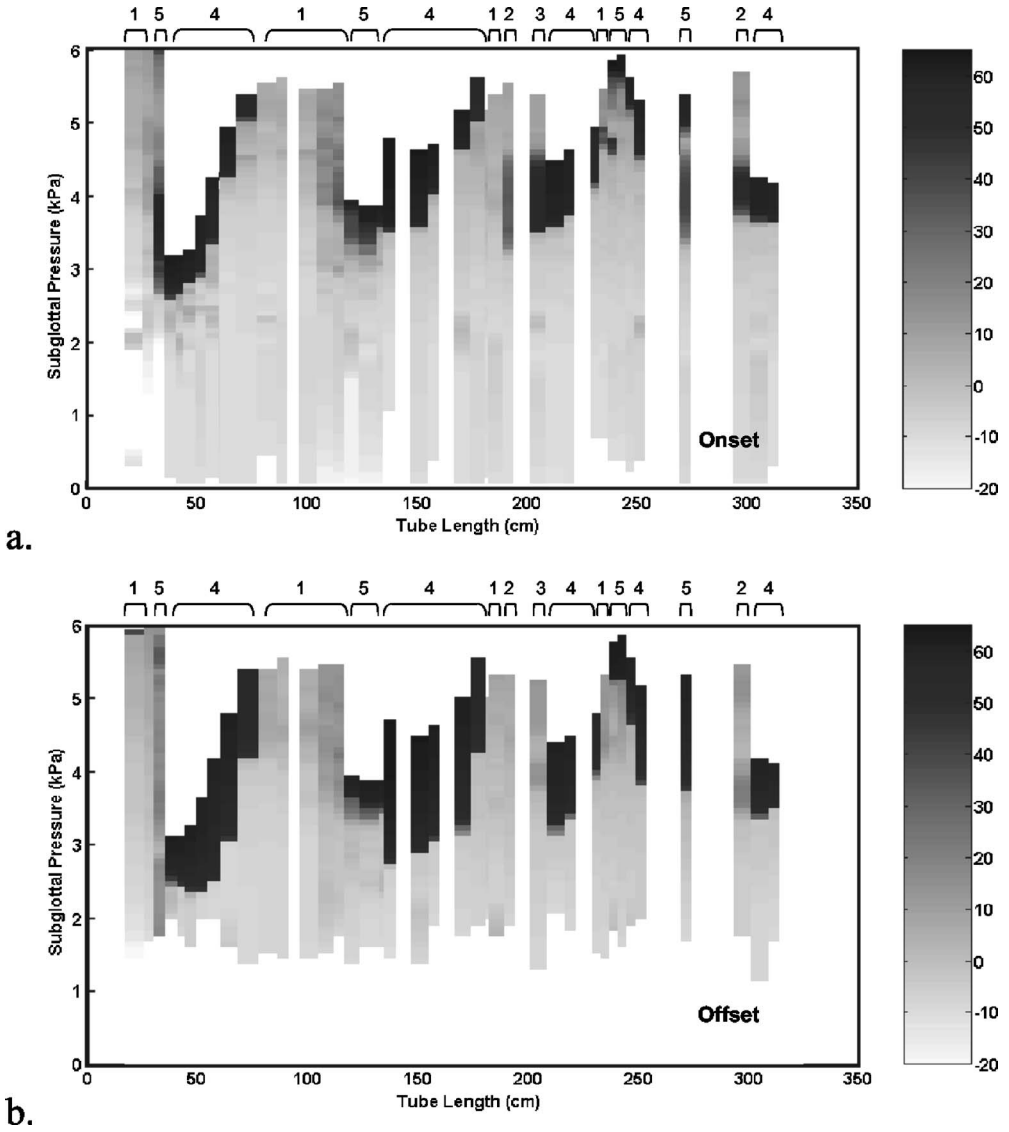


FIG. 4. Sound pressure amplitude (dB) in the subglottal tube as a function of the tracheal length and mean subglottal pressure in (a) increasing-flow and (b) decreasing-flow steps. Regions of various onset regimes are labeled for: 1. irregular phonation; 2. weak supercritical; 3 weak subcritical; 4. strong subcritical; 5. mixed types. The sound pressure amplitude in dB is represented by the gray scale.

and measurements were comparable, except for slight mismatches in the regions of frequency jumps. Note that no phonation was observed in the experiments (Figs. 3 and 9) over certain ranges of the tracheal length (e.g., 70–100 cm, and below 30 cm) in the flow rate range of the experiments. This is discussed further in Sec. IV. As the tracheal length increased, the variation in fundamental frequency over one cyclic range of the tracheal length decreased gradually from 72 Hz (over roughly a 30- to 70-cm range in tracheal length) to 55 Hz (over roughly a 190- to 270-cm range in tracheal length), which is consistent with the acoustic theory presented in the Appendix. The resonance frequencies of the subglottal system were calculated from the measured reflection factor of the expansion chamber and are also shown in Fig. 3. Clearly, the fundamental frequency closely approximated the subglottal acoustic resonance within each cyclic range of the tracheal length. Generally the fundamental frequency was slightly higher than the acoustic resonance frequency it followed. The difference gradually increased with tracheal length within each cyclic range of the tracheal length, but decreased with tracheal length between cyclic ranges. After onset, the fundamental frequency generally stayed constant and varied only slightly with the subglottal pressure.

B. Phonation onset quality

Figure 4 shows the sound pressure amplitudes for different values of the tracheal length and the mean subglottal pressure during increasing- (onset) and decreasing-flow (offset) experiments. The sound pressure amplitudes were estimated as the standard deviation of the recorded sound pressure data. Phonation onset and offset were identified by a significant change in the sound pressure amplitude between two consecutive steps (or a significant change of gray scale in Fig. 4).

The onset and offset behavior (related to the Hopf bifurcation from nonlinear systems theory [Berge *et al.*, 1984])

demonstrated a quasicyclic pattern based on the subglottal acoustics. In particular, alternating weak and strong onset regions appeared as the tracheal length increased. Four types of phonation onset were identified: irregular, weak supercritical, weak subcritical, and strong subcritical.

1. Irregular onset

Regions existed in Fig. 4 in which no apparent phonation onset was observed. An example is shown in Fig. 5(a), which was measured for a tracheal length of 182 cm. The left figure shows a bifurcation diagram in which the sound pressure amplitude is plotted against the mean subglottal pressure for both increasing- and decreasing-flow experiments. There was no clear phonation onset for this case. Only a very weak onset was observed in the increasing stage, as indicated by the small hump (around 4 kPa) in the bifurcation diagram. The right figure shows the corresponding power spectra of the measured sound pressure at each step, with subglottal pressure increased from zero to an upper limit and decreased back to zero. The power spectra showed that the energy inside the subglottal tube distributed mostly around the acoustic resonance frequency of the subglottal system. Time series data of the sound pressure showed a periodic component at 220 Hz, superimposed with intermittent low-frequency modulations, and chaos. The periodic component was so weak that the sound field was dominated by such instabilities.

2. Strong subcritical onset

Strong phonation onset observed in the experiments was normally subcritical (an abrupt increase in oscillation amplitude). One example is shown in Fig. 5(d) measured at a tracheal length of 208.4 cm. The sound pressure amplitude increased abruptly at a critical subglottal pressure [around 3.5 kPa in Fig. 5(d)]. After onset the sound pressure amplitude continued to increase with the subglottal pressure in the flow range of the experiments. In the decreasing-flow stage, the sound pressure amplitude decreased following a slight different curve from that in the increasing-flow stage. At a critical (offset) subglottal pressure slightly lower than the onset subglottal pressure, the sound pressure amplitude decreased abruptly to a very small value and phonation abruptly ceased. Power spectra during phonation showed that both the fundamental and harmonics were strongly excited. The onset pattern was repeatable whether immediately following a previous measurement, or after a period of rest.

3. Weak supercritical and subcritical onset

In between regions of irregular onset and strong onset in Fig. 4, two transitional types of onset existed. The onsets of these types were typically weak in strength and existed only for a certain limited range of subglottal pressure, i.e., there was a phonation offset even in the increasing-flow stage. Unlike the strong onset, the weak onset can be either supercritical (a gradual increase in glottis oscillation amplitude) or subcritical. Figure 5(b) shows one case of a weak and supercritical onset, obtained using a tube length of 188.7 cm. The difference between this type and the irregular onset is that there was a clear phonation onset above a critical subglottal pressure (or onset pressure). Time series sound pressure data at the onset showed a clear dominant periodic component with slight modulations and noise. The modulations disappeared as subglottal pressure continued to increase. Beyond the onset pressure, the sound pressure amplitude gradually increased with increasing subglottal pressure, indicating a supercritical (or soft) onset. The onset was weak and phonation was sustainable for only a narrow range of subglottal pressure. As subglottal pressure further increased, the sound pressure amplitude saturated and then gradually decreased to phonation offset. Although there was an apparent phonation onset, the onset strength was much weaker than that in the strong onset case. The maximum sound pressure was at least ten times smaller than that in the strong onset case. In the sound pressure spectra, only the fundamental frequency was strongly excited during the weak onset. The harmonics were hardly excited. It was observed in this study that the weak type of onset was always accompanied by weakly excited harmonics. In contrast to the strong onset case, there was no phonation in the decreasing-flow stage.

A case of weak subcritical onset is shown in Fig. 5(c), which was obtained for a tracheal length of 200.7 cm. This case was similar to the previous case [Fig. 5(b)], except that the onset was of the subcritical type. There was a jump in the sound pressure amplitude at phonation onset. As in the previous case, phonation was sustained only for a narrow range of subglottal pressure in the increasing-flow stage. As the subglottal pressure increased further, the sound pressure amplitude first saturated and then gradually decreased to phonation offset. The maximum sound pressure in this case was much larger than that in the previous case, but still smaller than that in the strong onset case. Sound pressure spectra during the onset showed that harmonics (particularly the first and third) were weakly excited. As in the weak supercritical case, no apparent phonation was observed in the decreasing-flow stage.

For both of the two weak onset cases, in an immediate repetition of the experiment, the bifurcation pattern (onset and offset) was qualitatively repeatable but with a much reduced saturation amplitude of the sound pressure. Phonation offset occurred at a lower value of subglottal pressure and the range of subglottal pressure for onset was reduced. The complete bifurcation pattern was repeatable only after a period of rest.

4. Coexistence of multiple onsets

More than one type of onset was observed at a same tracheal length. Figure 6 shows a case in which a supercritical onset was immediately followed by a subcritical onset. The tracheal length was 123.5 cm. The sound pressure spectra for these two onsets show that the difference lies in the

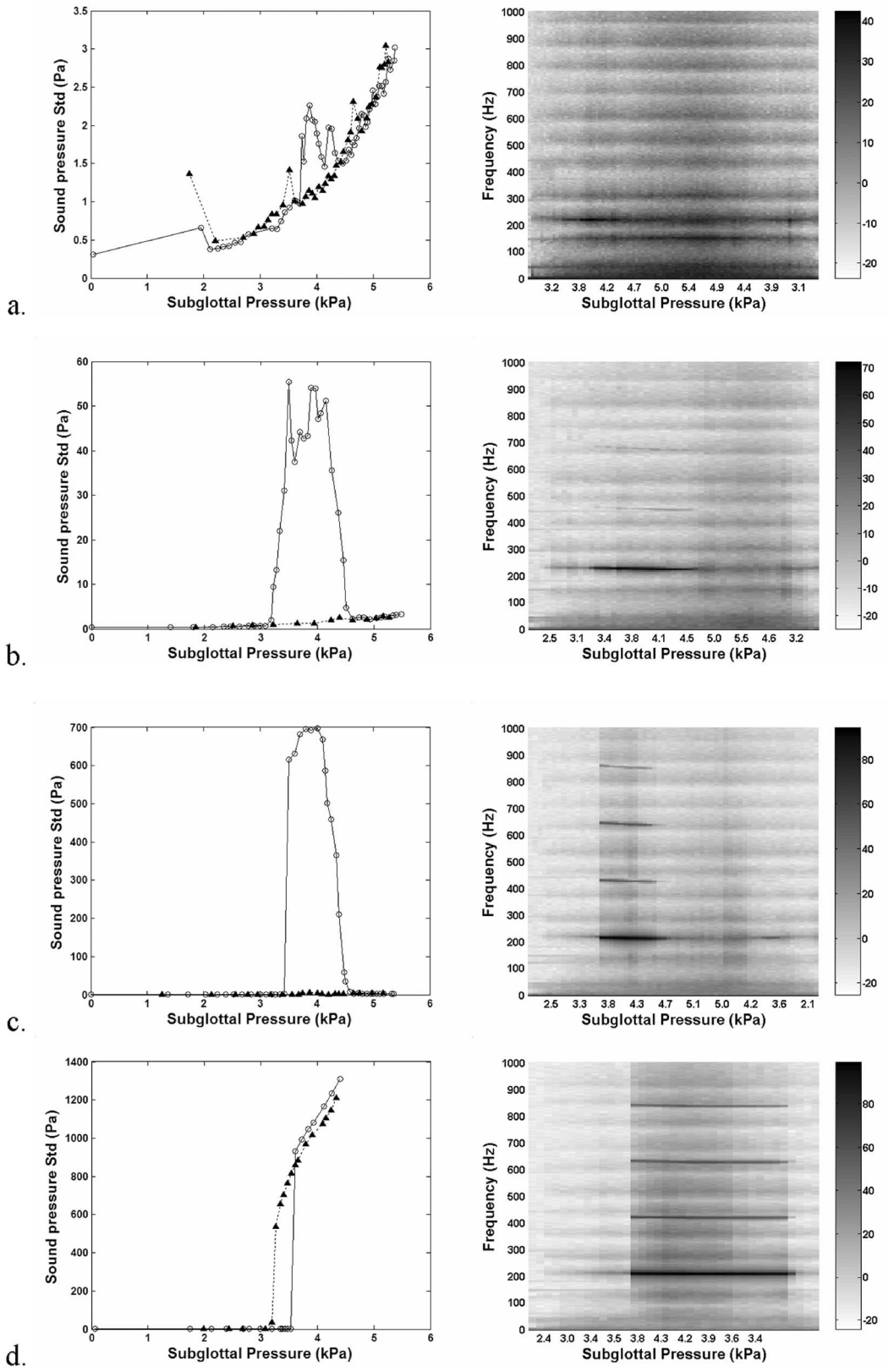


FIG. 5. Sound pressure amplitude (left) and power spectra (right) for increasing and decreasing subglottal pressure at four typical onset types: (a) irregular; (b) weak supercritical; (c) weak subcritical; and (d) strong subcritical. ○: increasing subglottal pressure; ▲: decreasing subglottal pressure. The power spectral amplitude in dB is represented by the gray scale.

excitation of the harmonics. The abrupt transition at the subcritical onset was accompanied by a boost in the strength of the harmonics, especially the second harmonic. A similar supercritical and subcritical behavior was observed in the decreasing-flow stage, in which the sound pressure amplitude first decreased abruptly to an intermediate value and then decreased gradually to phonation offset.

Figure 7 shows a case (with a tracheal length of 269 cm)

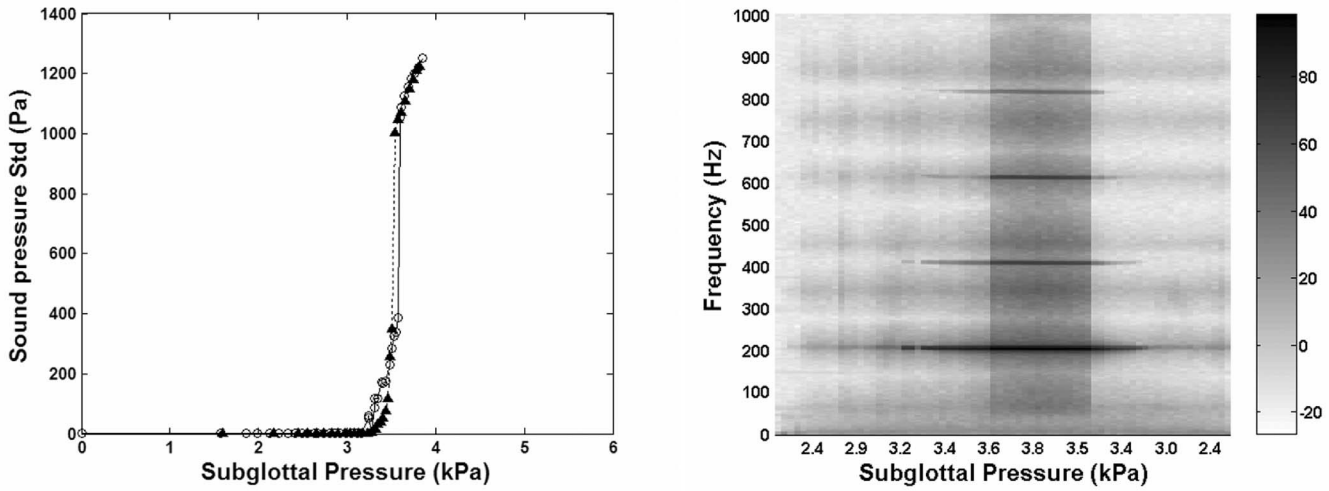


FIG. 6. Sound pressure amplitude (left) and power spectra (right) for increasing and decreasing subglottal pressure. Tracheal length $L=123.5$ cm. Two types of onset can be observed with a supercritical onset followed by a subcritical onset. ○: increasing subglottal pressure; ▲: decreasing subglottal pressure. The power spectral amplitude in dB is represented by the gray scale.

in which onsets of two different fundamental frequencies were observed. In this figure, the following phenomena were noted: a register change, a frequency jump, and a distinct entrainment to the subglottal acoustics. The onset at lower subglottal pressure (around 3.3 kPa) was weak and supercritical, with a fundamental frequency around 220 Hz. The sound pressure amplitudes increased slowly with the subglottal pressure after onset and then decreased gradually to an offset (around 4.5 kPa). The harmonics were weakly excited, characteristic of a weak onset. As subglottal pressure continued to increase, a second onset occurred at a subglottal pressure of about 4.9 kPa. The second onset was strong and subcritical, with a lower fundamental frequency around 168 Hz. Harmonics were strongly excited during the second phonation. The two fundamental frequencies corresponded roughly to the third and fourth resonance frequency of the subglottal system (Fig. 3). Superior views of the physical vocal fold model were recorded using a high-speed camera during the two onsets. The time history of a medial-lateral line taken from the center of the vocal folds (known as a spatio-temporal plot) is shown in Fig. 8 for a few oscillation cycles. The vocal folds during the first phonation demonstrated a falsetto-like, high-frequency, and low-amplitude vibration pattern, while the second phonation showed a more chest-like vibration with much larger amplitude at a lower frequency.

As discussed in Sec. V of the Appendix, more than one solution may exist for the fundamental frequency of the coupled laryngeal-subglottal system. However, only the solution with the largest negative damping is singled out and amplified. The two-onset case shown in Fig. 7 was in the region where frequency jump occurred in Fig. 3. In this region, several competing solutions were likely to have similar total damping, therefore similar onset threshold pressure. Even a slight increase in the subglottal pressure could trigger the system to favor one over another. It is possible that the

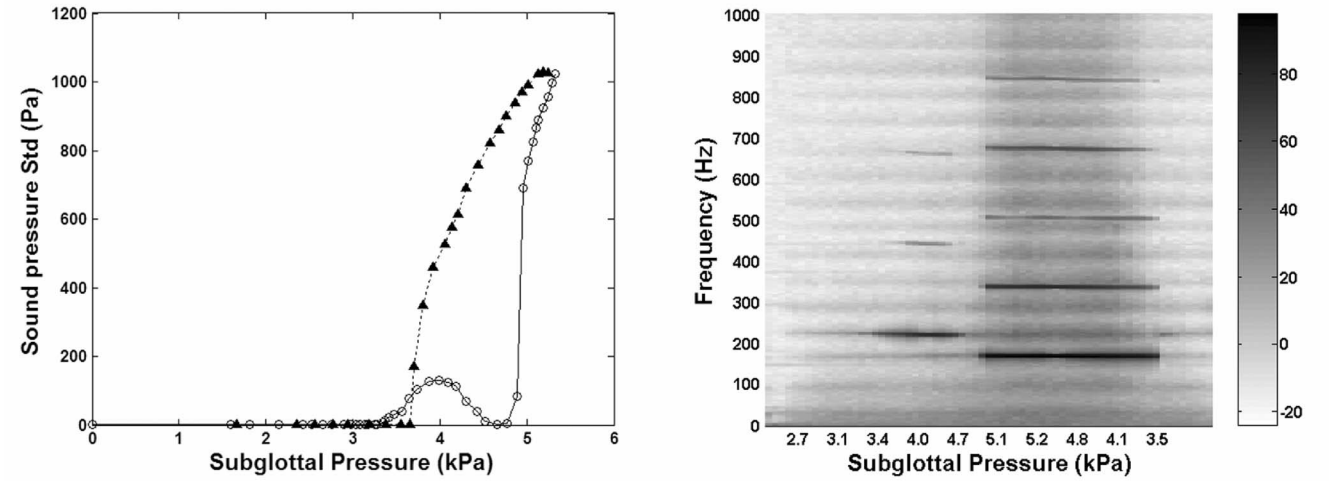


FIG. 7. Sound pressure amplitude (left) and power spectra (right) for increasing and decreasing subglottal pressure. Tracheal length $L=269$ cm. Two types of onset with different fundamental frequencies can be observed with a supercritical onset followed by a subcritical onset, ○: increasing subglottal pressure; ▲: decreasing subglottal pressure. The power spectral amplitude in dB is represented by the gray scale.

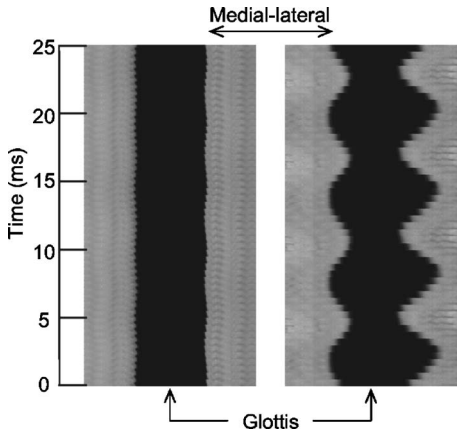


FIG. 8. Time history of a medial-lateral strip of the vocal folds at the center during the two types of onset shown in Fig. 7. Left figure: onset with high F_0: right figure: onset with low F_0. Note that there is a phase difference between the left and right vocal folds.

first onset in Fig. 7 had a lower threshold subglottal pressure but a weaker strength. As subglottal pressure further increased the first one was less preferred. Above the threshold pressure for the second type, the second onset became the favored and dominated with a much stronger onset.

C. Subglottal pressure at onset and offset

Onset and offset subglottal pressures were measured for cases with clear onset and offset. The onset and offset pressures were defined as the subglottal pressure at which sudden change in sound pressure amplitude first occurred. For supercritical onset, it was difficult to identify the exact conditions at which phonation onset and offset occurred. In these cases, the onset (offset) subglottal pressure was estimated as the subglottal pressure at which there was a considerable increase (decrease) in the slope in the plot of the sound pressure amplitude versus subglottal pressure. No offset pressure was measured for weak onsets as there was normally no phonation in the decreasing-flow stages.

The measured onset and offset pressures are shown in Fig. 9. Both the onset and offset subglottal pressures showed cyclic behavior in their variation with the tracheal length.

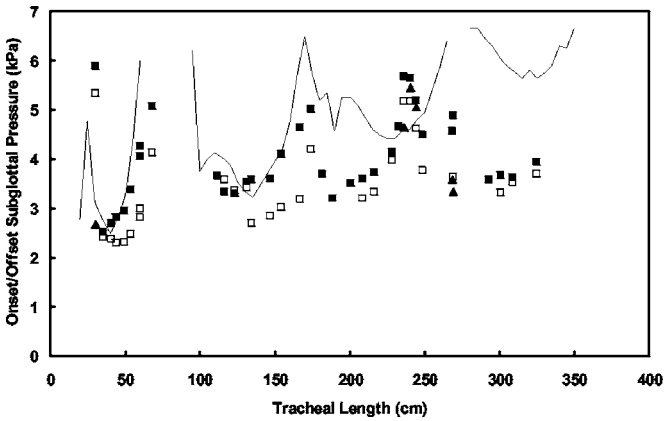


FIG. 9. Measured phonation onset (closed symbols) and offset (open symbols) threshold subglottal pressure as a function of the tracheal length.—: predicted onset threshold pressure from theory. ■: major onset; ▲: weak or supercritical onset for cases where more than one onset is observed.

Within each cyclic range of the tracheal length, the onset pressure slightly decreased and then increased with the tracheal length. The minimum value of the onset subglottal pressure within each cyclic range slightly increased with tracheal length. Theoretically, this behavior could be reproduced by including acoustic losses inside the tracheal tube in the model (see the Appendix). The offset pressure followed a similar trend as the onset pressure. Generally the offset subglottal pressure was lower than the onset subglottal pressure. Within each cyclic range of the tracheal length, the difference between onset and offset pressures was largest at the local minimum of the onset subglottal pressure, and gradually decreased as the tube length increased.

Also shown in Fig. 9 is the predicted onset pressure from the theory in the Appendix. The predicted onset pressure was overall slightly higher than the measured values. The theoretical model in this study only focused on the effects of the subglottal system. Accurate prediction of phonation onset pressure would require other factors to be included, which are beyond the scope of this study. The predicted and measured values exhibited the same general trends over the first two cyclic ranges of the tracheal length (up to a tracheal length of about 170 cm). There was a slight mismatch between measurement and prediction in the tracheal length range of the third cyclic range of the onset pressure. The highest onset pressure within the third cyclic range in experiments occurred at a tracheal length around 240 cm, while the predicted tracheal length was around 270 cm. In this region (230–270 cm), experiments exhibited mixed types of onsets and onsets of distinct frequencies. It is possible that two neighboring cyclic ranges overlapped for large values of the tracheal length, which may have distorted the onset pattern.

The variation of the phonation onset pressure with tracheal length was related to that of the onset strength. Generally low onset pressure corresponded to the regions of strong onset, while high onset pressure occurred at regions of weak and irregular onset types. Comparison of Figs. 3 and 4 shows that regions of irregular onset occurred roughly at fundamental frequencies above a threshold value (about 215 Hz), while the strong onset occurred at fundamental frequencies below this threshold value. As discussed in the Appendix, the phonation onset condition [Eq. (A11)] tended to restrict the fundamental frequency to a limited range. In this range the total system damping was negative. This may correspond to the strong onset case. For F_0 above this range, the total system damping was positive and no phonation or irregular phonation would occur. At the upper bound of this range, the total system damping was very small, either positive or negative, and the system is at a critical state. This seems to correspond to the regions of weak onset. As the theoretical model in the Appendixwas based on linear theory, it could not predict all the onset types encompassed by nonlinear systems theory. Nevertheless, it is reasonable to suppose that total system damping may have played a role in determining onset quality.

IV. DISCUSSION

Many previous laboratory investigations of phonation involving physical models, excised larynges, and *in vivo* canine larynges have failed to fully specify the subglottal system. Many of these same studies have reported a variety of nonlinear phenomena, including bifurcations (e.g., various classes of phonation onset and offset, register changes, frequency jumps), subharmonics, and chaos—and attributed such phenomena to the biomechanical properties of the larynx. However, such nonlinear phenomena may also be indicative of strong coupling between the voice source and the subglottal tract. Consequently, in such studies, it is not clear whether the underlying mechanisms of such nonlinear phenomena are acoustical, biomechanical, or a coupling of the acoustical and biomechanical systems. Moreover, because such laboratory investigations have frequently failed to report the length of the subglottal tube, or utilized subglottal tube lengths considerably longer than that of the human trachea, if acoustical mechanisms were involved, one cannot be certain whether the nonlinear phenomena were reflective of human voice production, at least at the reported vibrational frequencies. Using a physical model of vocal fold vibration, it is shown that such nonlinear phenomena may be replicated solely on the basis of laryngeal interactions with the acoustical resonances of the subglottal tube, using tracheal tube lengths which have been commonly reported in the literature, and no changes in the biomechanical properties of the laryngeal model.

In these experiments, no phonation was observed for short tracheal lengths (17–30 cm). In subsequent studies, however, phonation was achieved with the same physical model for short tracheal lengths when the flow rate was increased beyond the 1800 ml/s levels of flow rate utilized in this study. As might be expected, lower phonation threshold pressures have also been achieved at these short tracheal tube lengths when the vocal fold stiffness was lowered. Nevertheless, for longer tracheal tube lengths, phonation onset was strongly influenced by the subglottal acoustics in a cyclic manner, as previously shown. Moreover, for these longer tracheal tube lengths, after onset occurred, the secondary instabilities such as period-doubling, biphonation, and frequency jumps appeared more frequently.

In addition, this study has confirmed that the constructive or destructive interference of the subglottal acoustics can result in changes of the vibratory modes of the vocal folds, which Titze (1988b) suggested as a possible explanation for register change. For singing, in which the subglottal resonances are close to the pitch frequency, the subglottal influence is likely to play a more important role. Furthermore, the vocal tract, another acoustic resonator above the glottis, is expected to have similar influence on the vocal fold vibration. It has been shown that singers can tune the vocal tract to benefit from the pitch frequency being close to a formant (Joliveau *et al.*, 2004). The interplay between the subglottal system and the vocal tract will be addressed in a future study.

The results presented in this paper indicate that laboratory experiments involving excised larynges or physical models should be performed with a detailed understanding of the subglottal system. Previous excised-larynx experiments have reported a variety of onset conditions and instabilities, without including a detailed description of the subglottal system utilized. Accurate interpretation of the results would be difficult without a good understanding of the characteristics of the subglottal system. Indeed, without appropriate details regarding the subglottal system, it would also be impossible to compare results across laboratories. For example, consider the experimental setup used in this study. With a slight change of the tracheal length from 170 to 220 cm, the onset pattern changed significantly from irregular, weak supercritical, to strong subcritical. Comparison of phonation threshold pressure across these different cases would be meaningless.

The following conclusions are reached: (1) the only way to rule out source-tract interactions in laboratory experiments of phonation is through the construction of an anechoic termination, which is a time-consuming and difficult task (see Zhang *et al.*, 2002); (2) if an anechoic termination is not utilized, source-tract interactions are always a possibility, so the subglottal system must be fully specified, and subglottal resonances identified; (3) when register jumps and other nonlinear phenomena occur, one must rule out the possibility of source-tract interactions before concluding that such phenomena are solely a reflection of the laryngeal biomechanics, (4) in case investigators want to study source-tract interactions, to help ensure their possible relevance to human phonation, the tracheal tube length should include physiological lengths (approximately 12–17 cm).

V. CONCLUSIONS

Many previous laboratory studies of human phonation have utilized tracheal lengths considerably longer than found in the human system. Other studies have simply not reported the tracheal dimensions. However, this study has shown the potential for strong interactions between laryngeal dynamics and subglottal acoustics. This concern becomes amplified when one is studying nonlinear vocal phenomena and attempting to relate such phenomena to laryngeal biomechanics. As illustrated in this study, this task cannot be performed without ruling out the possible influence of subglottal acoustics.

ACKNOWLEDGMENTS

This investigation was supported by Research Grant Nos. R01 DC03072 and R01 DC004688 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health. The authors also acknowledge the assistance of Dr. Scott L. Thomson and Dr. Luc G. Mongeau in providing a prototype vocal fold model and for their suggestions in constructing the vocal fold models.

APPENDIX: DETAILS OF THE SIMPLIFIED THEORETICAL MODEL

Patterned after the models of Gupta *et al.* (1973) and Wilson and Beavers (1974), in the following a one-dimensional, theoretical model of the subglottal acoustics is presented which captures the vocal phenomena in this laboratory investigation. We assumed a steady laminar flow

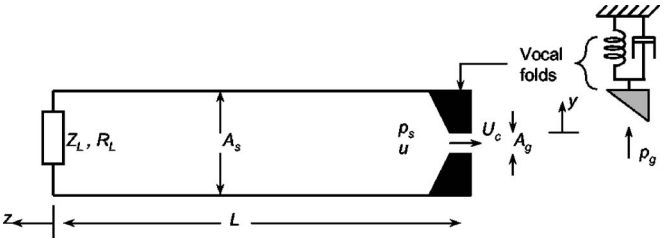


FIG. 10. Simplified model of the vocal folds and subglottal system. The vocal folds are modeled as a symmetric spring-mass-damping system.

through the glottis prior to phonation onset. Stability of the system under small-amplitude, single-frequency disturbances in the subglottal system was considered.

1. The vocal fold model

The vocal fold was modeled as a symmetric one-dimensional spring-mass-damping system (Fig. 10). Consider a small-amplitude simple harmonic oscillation of the vocal folds $y=Ye^{i\omega t}$, where y is the displacement of the vocal fold mass from its equilibrium position, Y is the complex displacement amplitude, and ω is the complex angular frequency. For small-amplitude oscillations the nonlinearities in the coupled system can be neglected so that all dynamic variables are simple harmonic functions of the same frequency.

The driving pressure applied on the vocal fold surface, p_g, can be obtained by integrating the pressure profile within the glottis. It can be related to the subglottal pressure p_s as follows:

$$p_g = C_g p_s, \tag{A1}$$

where the coupling factor C_g accounts for the dependence of the mean pressure p_g on the resting glottal geometry. The equation of motion for vocal fold oscillation is

$$M\ddot{y} + B\dot{y} + Ky = C_g p_s', \tag{A2}$$

where p_s' is the fluctuating component of the subglottal pressure p_s, M, K, and B are the effective lumped mass, stiffness, and damping per unit area of the vocal folds, respectively.

2. Pressure-flow relationship through the glottis

For an open prephonatory glottis and small-amplitude oscillation, the unsteady flow through the glottis can be accurately modeled using a quasisteady approximation (Zhang *et al.*, 2002). The jet centerline velocity u_c is related to the subglottal pressure p_s by

$$p_s = \tfrac{1}{2}\rho u_c^2. \tag{A3}$$

Equation (A3) can be linearized by decomposing the pressure and flow fields into a mean and a fluctuating component and neglecting higher-order terms:

$$p_{s0} = \tfrac{1}{2}\rho u_{c0}^2, \quad p_s' = \rho u_{c0} u_c', \tag{A4}$$

where (p_{s0}, u_{c0}) denote the mean components, and (p_s', u_c') denote the fluctuating components. Similarly, linearization of the continuity equation gives, neglecting density fluctuations,

$$u_s' A_s = C_d(A_{g0} u_c' + u_{c0} A_g'), \tag{A5}$$

where C_d is the orifice discharge coefficient, and A_g is the minimum cross-sectional area of the glottal channel. The fluctuating glottal area A_g' is related to the vocal fold displacement by

$$A_g' = 2l_g y = 2l_g Y e^{i\omega t}, \tag{A6}$$

where l_g is the length of the glottis.

3. Subglottal system

The subglottal system consists of a uniform tube of length L and cross-sectional area A_s simulating the trachea (Fig. 10). The bronchi and the lungs are modeled as an equivalent acoustic impedance Z_L (or reflection factor R_L). The input impedance of the subglottal system to the glottis, Z_s, is (Kinsler *et al.*, 2000):

$$Z_s = \frac{P_{z=-L}}{U_{z=-L}} = \rho c \frac{e^{-ik(-L)} + R_L e^{ik(-L)}}{e^{-ik(-L)} - R_L e^{ik(-L)}}, \tag{A7}$$

where ρ is the air density, c is the speed of sound, and P and U are the complex acoustic pressure and velocity inside the subglottal tube, respectively. The complex wave number $k = 2\pi f/c - i\alpha$, where α is the absorption coefficient which accounts for the acoustic loss inside the trachea.

4. Equation of motion of the coupled system

The subglottal acoustic pressure and velocity are related to the fluctuating subglottal pressure and flow velocity by

$$p_s' = P_{z=-L} e^{i\omega t}, \quad u_s' = U_{z=-L} e^{i\omega t}. \tag{A8}$$

Combining Eqs. (A4)–(A8) yields

$$p_s' = Z_a y = \frac{2Z_s C_d u_{c0} l_g}{\left(1 - \dfrac{Z_s C_d A_{g0}}{\rho u_{c0} A_s}\right) A_s} y, \tag{A9}$$

where Z_a is the effective loading of the subglottal system to the vocal folds. Equation (A9) gives an explicit relationship between the driving oscillating pressure in the subglottal tube and the resulting oscillation of the vocal folds. In general, the effective loading Z_a is a function of tracheal length, oscillating frequency, subglottal pressure, and coupling between the tracheal and the glottis A_{g0}/A_s. The characteristic equation of the coupled system can be obtained from Eqs. (A2) and (A9):

$$[-\omega^2 M + i\omega(B - C_g X_a/\omega) + K - C_g R_a] Y = 0, \tag{A10}$$

where R_a and X_a are the real and imaginary part of Z_a, respectively.

Equation (A10) shows that the influence of the subglottal acoustics on the vocal fold oscillation is twofold. First, the subglottal acoustics introduces an effective damping $-C_g X_a/\omega$ into the coupled system. This effective damping may be positive or negative, depending on the oscillating frequency and tracheal length, and may increase or decrease the phonation threshold pressure. The subglottal acoustics

also introduces an effective stiffness $-C_g R_a$. This will lead to cyclic variations of the fundamental frequency with the tracheal length.

Equation (A10) can be solved numerically for the complex angular frequency $\omega=\omega_r+i\omega_i$ as a function of the tracheal length and the subglottal pressure. The fundamental frequency F_0 can then be determined from the real part of ω. For sustained oscillations of the vocal folds, the fundamental frequency also has to satisfy the net energy transfer condition, i.e., the total system damping needs to be zero or negative:

$$B - C_g X_a/\omega_r \leqslant 0, \tag{A11}$$

where the equality is reached at phonation onset. Phonation threshold pressure can be determined to be the minimum subglottal pressure for which Eq. (A11) was satisfied. Note that when the inequality in Eq. (A11) is satisfied, the vocal fold displacement will increase exponentially. The small-amplitude oscillation assumption will not be valid and this model is not applicable anymore.

5. Open-ended ideal tracheal tube

As an example, Eq. (A10) was solved for the case of an ideal open-ended and lossless trachea as a function of kL, where k is the wave number and L the tracheal length. Typical results are shown in Fig. 11 for a mean subglottal pressure 1.6 kPa. We used the following values for the model constants, which were adapted from the two-mass model (Ishizaka and Flanagan, 1972):

$$M = 2.656 \text{ kg m}^{-2}, \quad K = 1.904 \times 10^6 \text{ N m}^{-3},$$

$$B = 1600 \text{ N s m}^{-3},$$

$$l_g = 1.4 \times 10^{-2} \text{ m}, \quad A_{g0} = 6 \times 10^{-6} \text{ m}^2,$$

$$A_t = 6.45 \times 10^{-4} \text{ m}^2,$$

$$C_d = 1, \quad C_g = 0.5. \tag{A12}$$

The fundamental frequency showed a periodic variation with kL. Over most of the range of kL, the fundamental frequency stayed close to the natural frequency of the lumped vocal fold model

$$\omega_0 = \sqrt{K/M - (B/2M)^2}, \tag{A13}$$

which was obtained when kL was integer multiples of π. The subglottal influence on the fundamental frequency dominated only at values of kL close to the quarter-wavelength resonances of the subglottal system. In this range the fundamental frequency closely approximated the quarter-wavelength resonance of the subglottal system. The maximum fundamental frequency was obtained when kL was odd multiples of $\pi/2$:

$$\omega_{\max} = \sqrt{K/M - (B/2M)^2 + C_g l_g P_{s0}/MA_{g0}}, \tag{A14}$$

which increases with the mean subglottal pressure. In between the two extreme cases, the fundamental frequency varied smoothly between ω_0 and $\omega_{\max}$.

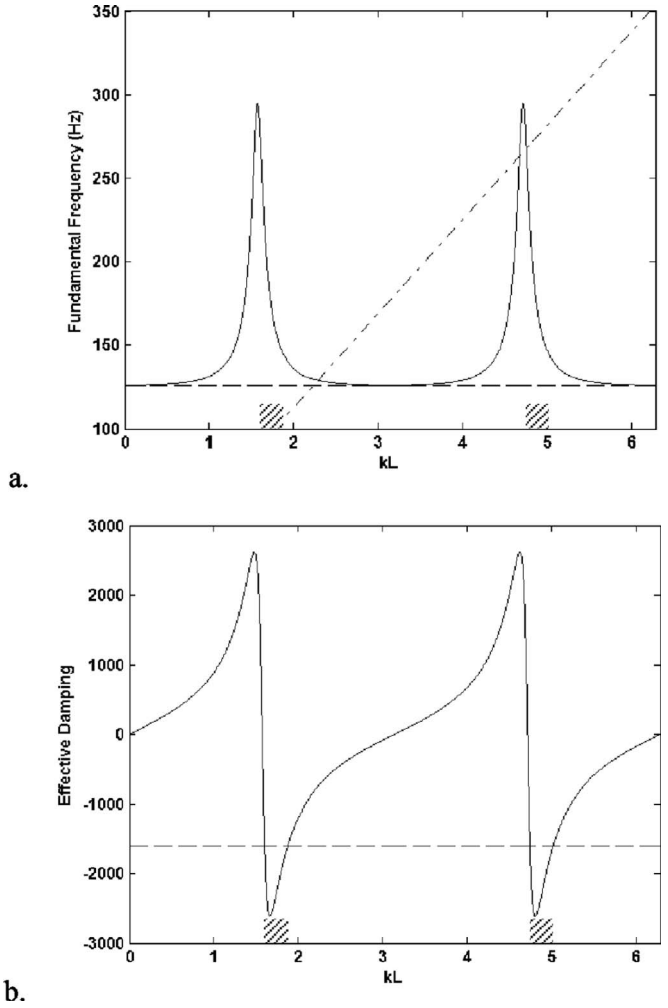


FIG. 11. (a) Fundamental frequency of vocal fold vibration and (b) corresponding effective damping induced by the subglottal system as a function of the product of wave number and tracheal length for an open-ended and lossless trachea tube. —: solution curve; – - –: (a) natural frequency of the vocal folds and (b) critical effective damping at which the equality in Eq. (A11) is reached; –·–: $f=(c/2\pi L)\cdot kL$ for $L=100$ cm; shaded area: regions possible for phonation onset.

For self-sustained oscillations of the vocal folds to occur, Eq. (A11) had to be satisfied, i.e., the total damping of the coupled system had to be zero (at onset) or negative. This requirement limited eligible kL values to series of narrow bands at kL values slightly higher than the quarter-wavelength resonances of the subglottal system, as indicated by the shaded areas in Fig. 11(b). The fundamental frequency at a certain tracheal length L could be obtained by drawing a straight line from the origin with a slope of $c/(2\pi L)$ in Fig. 11(a). The intersection of this line with the solution curve gave the fundamental frequency. For a certain length of the trachea, one or multiple solutions for the fundamental frequency were possible [e.g., three solutions are possible for $L=100$ cm, as shown in Fig. 11(a)]. A unique set of solutions of the fundamental frequency may be obtained by choosing the one with the largest negative value of the effective damping.

Alipour, F., and Scherer, R. C. (**2001**). "Effects of oscillation of a mechanical hemilarynx model on mean transglottal pressures and flows," J.

Acoust. Soc. Am. **110**, 1562–1569.
Austin, S. F., and Titze, I. R. (**1997**). "The effect of subglottal resonance upon vocal fold vibration," J. Voice **11**, 391–402.
Berge, P., Pomeau, Y., and Vidal, C. (**1984**). *Order Within Chaos* (Hermann and Wiley, Paris, France).
Berry, D. A., Herzel, H., Titze, I. R., and Story, B. H. (**1996**). "Bifurcations in excised larynx experiments," J. Voice **10**, 129–138.
Flanagan, J. L. (**1958**). "Some properties of the glottal sound source," J. Speech Hear. Res. **1**, 99–116.
Fletcher, N. H. (**1993**). "Autonomous vibration of simple pressure-controlled valves in gas flows," J. Acoust. Soc. Am. **93**, 2172–2180.
Gupta, V., Wilson, T. A., and Beavers, G. S. (**1973**). "A model for vocal chord excitation," J. Acoust. Soc. Am. **54**, 1607–1617.
Harper, P., Kraman, S. S., Pasterkamp, H., and Wodicka, G. R. (**2001**). "An acoustic model of the respiratory tract," IEEE Trans. Biomed. Eng. **48**, 543–550.
Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal chords," Bell Syst. Tech. J. **51**, 1233–1267.
Ishizaka, K., Matsudaira, M., and Kaneko, T. (**1976**). "Input acoustic-impedance measurement of the subglottal system," J. Acoust. Soc. Am. **60**, 190–197.
Jiang, J. J., and Zhang, Y. (**2005**). "Spatiotemporal chaos in excised larynx experiments," Phys. Rev. E **72**, 035201–1:4.
Jiang, J. J., Zhang, Y., and Ford, C. N. (**2003**). "Nonlinear dynamics of phonations in excised larynx experiments," J. Acoust. Soc. Am. **114**, 2198–2205.
Joliveau, E., Smith, J., and Wolfe, J. (**2004**). "Vocal tract resonances in singing: The soprano voice," J. Acoust. Soc. Am. **116**, 2434–2439.
Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**2000**). *Fundamentals of Acoustics* (Wiley, New York).
Large, J. (**1972**). "Towards an integrated physiologic-acoustic theory of vocal registers," NATS Bull. **29**, 18–25.
Mergell, P., and Herzel, H. (**1997**). "Modeling biphonation – The role of the vocal tract," Speech Commun. **22**, 141–154.
Nadoleczny-Millioud, M., and Zimmerman, R. (**1938**). "Categories et registres de la voix," Names and Registers of the Voice **23**, 21–31.
Ruty, N., Van Hirtum, A., Pelorson, X., Lopez, I., and Hirschberg, A. (**2005**). "A mechanical experimental setup to simulate vocal folds vibrations. Preliminary results," ZAS Papers in Linguistics **40**, 161–175.
Seybert, A. F., and Ross, D. F. (**1977**). "Experimental determination of acoustic properties using a two-microphone random-excitation technique," J. Acoust. Soc. Am. **61**, 1362–1370.
Švec, J. G., Schutte, H. K., and Miller, D. G. (**1999**). "On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges," J. Acoust. Soc. Am. **106**, 1523–1531.
Thomson, S. L., Mongeau, L., and Frankel, S. H. (**2005**). "Aerodynamic transfer of energy to the vocal folds," J. Acoust. Soc. Am. **118**, 1689–1700.
Thomson, S. L., Mongeau, L., Frankel, S. H., Neubauer, J., and Berry, D. A. (**2004**). "Self-oscillating laryngeal models for vocal fold research," *Proceedings of the Eighth International Conference on Flow-Induced Vibrations*, Paris, France, Vol. **2**, pp. 137–142.
Titze, I. R. (**1988a**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**(4), 1536–1552.
Titze, I. R. (**1988b**). "A framework for the study of vocal registers," J. Voice **2**, 183–194.
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. (**1995**). "Acoustic interactions of the voice source with the lower vocal tract," J. Acoust. Soc. Am. **101**, 2234–2243.
van den Berg, J. W. (**1968a**). "Sound production in isolated human larynges," Ann. N.Y. Acad. Sci. **155**, 18–27.
van den Berg, J. W. (**1968b**). "Register problems," Ann. N.Y. Acad. Sci. **155**, 129–134.
van den Berg, J. W., and Tan, T. S. (**1959**). "Results of experiments with human larynxes," Pract. Otorhinolaryngol. (Basel) **21**, 425–450.
Van der Plaats, A., Schutte, H. K., van der Eerden, F. J. M., de Vries, M. P., Mahieu, H. F., and Verkerke, G. J. (**2006**). "An *in-vitro* test set-up for evaluation of a voice- producing element under physiologic acoustic conditions," Ann. Biomed. Eng. **34**, 893–900.
Vennard, W. (**1967**). "*Singing…the Mechanisms and the Technique*," Revised 2nd ed. (Fisher, New York).
Vilain, C. E., Pelorson, X., Hirschberg, A., Le Marrec, L., Op't Root, W., and Willems, J. (**2003**). "Contribution to the physical modeling of the lips. Influence of the mechanical boundary conditions," Acta Acoustica United with Acustica **89**, 882–887.
Weibel, E. R. (**1963**). *Morphometry of the Human Lung* (Springer, Berlin).
Wilson, T. A., and Beavers, G. S. (**1974**). "Operating modes of the clarinet," J. Acoust. Soc. Am. **56**(2), 653–658.
Zhang, Z., Mongeau, L., and Frankel, S. H. (**2002**). "Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes," J. Acoust. Soc. Am. **112**(4), 1652–1663.

Anterior-posterior biphonation in a finite element model of vocal fold vibration

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 8 July 2005; revised 9 June 2006; accepted 15 June 2006)

In this paper, a finite-element model is used to simulate anterior-posterior biphonation [Neubauer *et al.*, J. Acoust. Soc. Am. **110**(6), 3179–3192 (2001)]. The anterior-posterior stiffness asymmetric factor and the anterior-posterior shape asymmetric factor describe the asymmetry properties of vocal folds. Spatiotemporal plot, spectral analysis, anterior-posterior fundamental frequency ratio, cross covariation function, and correlation length quantitatively estimate the spatial asymmetry of vocal fold oscillations. Calculation results show that the anterior-posterior stiffness asymmetry decreases the spatial coherence of vocal fold vibration. When the stiffness asymmetry reaches a certain level, the drop in spatial coherence desynchronizes the vibration modes. The anterior and posterior sides of the vocal fold oscillate with two independent fundamental frequencies (f_a and f_p). The complex spectral characteristics of vocal fold vibration under biphonation conditions can be explained by the linear combination of f_a and f_p. Empirical orthogonal eigenfunctions prove the existence of higher-order anterior-posterior modes when anterior-posterior biphonation occurs. Then, it is found that the anterior-posterior shape asymmetry also decreases the spatial coherence of vocal fold vibration, and shape asymmetry is a possible reason for anterior-posterior biphonation.
 [DOI: 10.1121/1.2221546]

PACS number(s): 43.70.Aj, 43.70.Bk, 43.70.Jt [AL] Pages: 1570–1577

I. INTRODUCTION

There exists substantial evidence that vocal fold vibration is a highly nonlinear process.[1–3] Vibration of the vocal folds shows complex behaviors in both the time and spatial domain such as bifurcation,[4–6] chaos,[7–10] biphonation,[11–17] spatial-temporal chaos,[18] and so on. Because irregular vocal fold vibration is associated with phonation quality and laryngeal pathology development,[8–10,19–24] many aspects of irregular vibration have been widely studied in the last years from both the clinical and theoretical point of view. Excised larynx experiments provide methods to directly observe the irregular spatiotemporal behavior of vocal fold vibration.[25,26] Nonlinear dynamics analysis methods such as phase space, Poincare map, Lyapunov exponent, correlation dimension, spatiotemporal correlation, and empirical orthogonal functions quantitatively classify vocal instabilities.[7,18,27–32] Mathematical models aid in understanding the mechanism of vocal fold vibration during phonation.[8–10,33–43]

The lumped mass model is firstly used to simulate the irregular vibration of the vocal folds due to the fact that it maintains a good balance between physiological completeness and computational costs.[33–37] One disadvantage of a lumped mass model's simplicity is that it cannot provide enough spatial information. However, several experimental and theoretical indications show that the desynchronization of the spatial modes plays an important role in the irregular vocal fold vibration pattern.[6,14,16,18] Therefore, it is necessary to employ a continuum model to study vocal fold vibration.[31–36] The continuum model can deal with complex boundaries and provide more spatial information than lumped mass models. Moreover, a continuum model can also define particular material properties. Consequently, a continuum model allows us to model the spatiotemporal characteristics of the vocal folds and reveal their inherent mechanism; this is valuable for improving speech quality and preventing laryngeal pathology development.

Biphonation is the simultaneous occurrence of two independent fundamental frequencies during phonation. Biphonation usually corresponds to a unique spatiotemporal pattern of vocal fold vibration.[16] Riede *et al.* and Brown *et al.* found biphonation in the natural howling of a dog-wolf mix and in excised larynx experiments of squirrel monkeys.[15,17] Experimental observations of Mergel *et al.* found that the left-right asymmetric vocal fold oscillations induce biphonation.[14] Recently, Neubauer *et al.* observed two independent fundament frequencies and higher-order modes in anterior-posterior asymmetric vocal fold vibration.[16] They introduced this biphonation mode as "AP (anterior-posterior) biphonation" while referring to the biphonation mode due to left-right asymmetric vocal fold vibration as "LR (left-right) biphonation."[16] Studies using the lumped mass model proved that stiffness imbalance of the left vocal fold and right vocal fold is an important factor that causes left-right oscillation desynchronization and thus left-right biphonation.[5,12,14] However, the mechanism of anterior-posterior biphonation has seldom been studied because of the limitations of the lumped mass model.

In this paper, we will try to simulate the anterior-posterior biphonation phenomenon of vocal fold vibration. First, a finite-element model of anterior-posterior asymmetric

[a)]Author to whom correspondence should be addressed. Electronic mail: jiang@surgery.wisc.edu

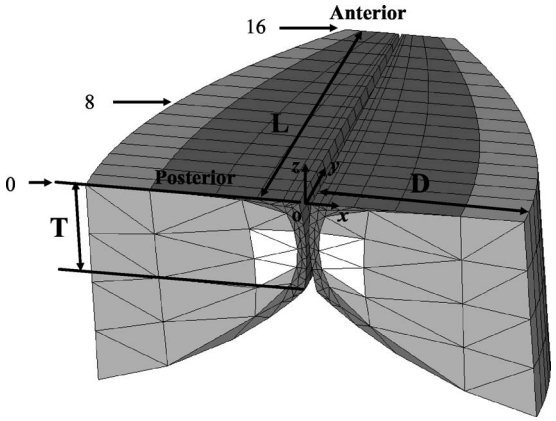


FIG. 1. A 3D view of the finite-element model, where the white region, light gray region, and dark gray region correspond to ligament, body, and cover, respectively. The length L of the vocal fold is 1.6 cm, the thickness T of the vocal fold is 0.45 cm, and the depth D of vocal folds is defined in Eq.(1). This finite element model contains 17 longitudinal layers. These layers, from posterior to anterior, are numbered from 0 to 16. Layers 0, 8, and 16 are labeled in this figure.

vocal fold vibration was introduced to simulate the spatiotemporal characteristics of vocal fold vibration. In this finite-element model, the anterior-posterior stiffness and shape asymmetry of the vocal fold are taken into account. The stiffness asymmetric factor (Q_{stf}) and shape asymmetric factor (Q_{shp}) quantitatively describe the anterior-posterior asymmetry of the finite-element model. We then introduced the anterior-posterior fundamental frequency ratio (F_{ap}), correlation function ($C_{\max}$), and correlation length (L_{corr}) to measure the asymmetry of vocal fold vibration. Finally, we studied the influence of stiffness asymmetry and shape asymmetry on vibration asymmetry and biphonation using this asymmetry finite-element model.

II. MODEL

In this section, we will introduce our finite-element model. The finite-element model was built and mathematically solved using the finite-element software ANSYS. Figure 1 gives a 3D view of our finite-element model where the white region represents the vocal fold ligament, the dark gray region represents the vocal fold cover, and the light gray region represents the vocal fold body. This biomechanical model of the vocal fold contains 17 longitudinal layers. These layers, from posterior to anterior, are numbered from 0 to 16. The two folds contain 2432 elements and 1734 nodes; there are 1216 elements and 867 nodes in each fold. In this study, we assume that the left and right folds are symmetric. Therefore only the right fold is simulated, and the impact between the two folds is represented by the interaction between the right fold and a rigid mid plane.[43,44] The collision process between the vocal fold and the rigid mid-plane is modeled by the augmented Lagrangian method.[44] In order to create vibration in the finite-element model, airflow pressure is applied on the vocal fold surface as a boundary condition under the assumption that there is no airflow in the anterior-posterior direction and that airflow pressure distribution along the inferior-superior direction follows Bernoulli's law,[36] that is, $P(z)=P_L[1-G_0^2/G^2(z)]$, where G_0 is the glottal width at the glottal exit, $G(z)$ is the width in the glottis, and $P_L=1$ kPa is lung pressure. When the vocal fold collision occurs, the pressure applied on the inferior side of the collision point is equal to lung pressure (P_L), and the pressure applied on the superior side of the collision point is set to zero. Dirichlet conditions are applied to all vocal fold surfaces not in contact with airflow, that is, the displacement on these surfaces is always equal to zero. With the driving airflow pressure, the motion of the vocal fold is solved using the Newmark method.

TABLE I. The input parameters used in the finite-element model of the vocal fold.[7,45]

Transverse Young's modulus of the body	4 kPa
Transverse Young's modulus of the ligament	2 kPa
Longitudinal shear modulus of the body	12 kPa
Longitudinal shear modulus of the ligament	40 kPa
Viscosity of the body and ligament	5 Poise
Transverse Poisson's ratio for all tissue layers	0.3

Clinical observations show that some vocal fold diseases can cause the asymmetry of the vocal fold shape and stiffness in the AP direction.[45,46] Inflammation and localized atrophy could change the vocal fold shape. Vocal scarring due to tissue trauma increases the local tension in the vocal folds. In order to simulate the vibration of these asymmetrical vocal folds, we introduce stiffness asymmetry and shape asymmetry in this FEA model.

The vocal fold tissue is modeled by a linear elastic model with linear damping. Furthermore, we also assume that each tissue layer in the vocal folds is transversely isotropic and that the vibration takes place in a single plane.[47] The movement in each layer can be reduced to lateral and vertical motions. Therefore, with such assumptions, the longitudinal Poisson's ratio is zero and the tissue properties can be specified by four constants: the transverse Young's modulus, the longitudinal shear modulus, viscosity, and the transverse Possion's ratio. The input parameters of the finite model's ligament layer and body layer are given in Table I.[7,47] Clinical observation has shown that many vocal diseases are often accompanied with atrophy and dehydration,[46,48] which could cause the change of the tissue compressibility. Therefore, a relative small transverse Poisson's ratio was chosen to represent the possible compressible tissue under pathological condition. The input parameters of the cover layer are anterior-posterior asymmetric. For the sake of convenience, we assume that the tissue property of the cover layer is half split. In the posterior half, the transverse Young's modulus, longitudinal shear modulus, and viscosity are 1 kPa, 2 kPa, and 3 poise, respectively. In the anterior half, the longitudinal shear modulus and viscosity are 2 kPa and 3 poise, which are equal to those in the posterior side. However, the transverse Young's modulus is $Q_{\text{stf}}\times 1$ kPa in the anterior half. Here, Q_{stf} is the anterior-posterior stiffness asymmetric parameter, which describes the anterior-posterior stiffness asymmetric property. For $Q_{\text{stf}}=1$, the anterior and posterior tensions are symmetric. For $Q_{\text{stf}}>1$, the anterior stiffness is larger than the posterior stiffness.

In this finite-element vocal-fold model, the length of vocal fold (L) is 1.6 cm, and the thickness of the vocal folds (T) is 0.45 cm.[47] The depth of the vocal folds (D) decreases from a maximum at the posterior region ($D_{\max}$) to a minimum at the anterior region ($D_{\min}$). The depth of vocal fold follows the equation[49]

$$D(y) = D_{\max} - (D_{\max} - D_{\min}) \times \frac{y^2}{L^2}, \quad D_{\max} \geqslant D_{\min} \qquad (1)$$

In order to describe the anterior-posterior shape asymmetry of the vocal fold, we introduce the anterior-posterior shape asymmetric factor (Q_{shp}), which is defined as the ratio of the maximum depth ($D_{\max}$) at the posterior side to the minimum depth ($D_{\min}$) at the anterior side:

$$Q_{\text{shp}} = D_{\max}/D_{\min}. \qquad (2)$$

In the following study, we set $D_{\max}$ equivalent to 1.0 cm and vary the value of Q_{shp}. Clearly, $Q_{\text{shp}}=1$ indicates a vocal fold with anterior-posterior symmetric shape. $Q_{\text{shp}}>1$ indicates a vocal fold with an anterior-posterior asymmetric shape, and that the depth at the posterior is larger than that of the anterior. As the value of Q_{shp} increases, the AP shape asymmetry becomes more significant.

III. RESULTS

In this section, we will use the above finite-element model to simulate vocal fold vibration. In order to quantitatively estimate the AP asymmetry of vocal fold vibration, the AP fundamental frequency ratio,[16] cross covariation function,[18,50] and correlation length[18] are employed to analyze the simulation results of the finite-element model. The AP fundamental frequency ratio (F_{ap}) is defined as: $F_{\text{ap}} = f_a/f_p$, where f_a is the fundamental frequency of the anterior side of the vocal folds, and f_p is the fundamental frequency of the posterior side of the vocal folds. The cross covariation is defined as[50]

$$C(i,j,\tau) = \frac{\langle \delta u(i,t)\, \delta u(j,t+\tau)\rangle_T}{\sqrt{\langle \delta u(i,t)^2\rangle_T \langle \delta u(j,t+\tau)^2\rangle_T}}, \qquad (3)$$

where $\delta u(i,t) = u(i,t) - \langle u(i,t)\rangle_T$, i is spatial position, t is time, and $\langle \bullet \rangle_T$ denotes the average value over time. For all the values of the delay time (τ), the maximum value [$(C_{\max}(i,j))$] of $C(i,j,\tau)$ estimates the spatial coherence of $u(i,t)$ and $u(j,t)$. In the following discussion, the glottal width signal [$G_w(i,t)$] of each layer i along the anterior-posterior direction is used as the spatiotemporal signal $u(i,t)$. $G_w(i,t)$ is the minimum distance between the mid plane and the node on the surface of the ith layer at time t. The middle layer (layer 8) is set as a spatial reference point. Figure 2 presents a typical waveform of $G_w(8,t)$ with $Q_{\text{stf}}=1$ and $Q_{\text{shp}}=1$. The $C_{\max}$ between the middle layer and the other layers is calculated. A large $C_{\max}$ indicates a strong spatiotemporal correlation, whereas a small $C_{\max}$ suggests a small spatiotemporal correlation. A small spatiotemporal correlation usually correlates with the spatial asymmetry, irregular spatiotemporal vibration pattern, and even spatiotemporal chaos.[18,50] The correlation length (L_{corr}) is defined as

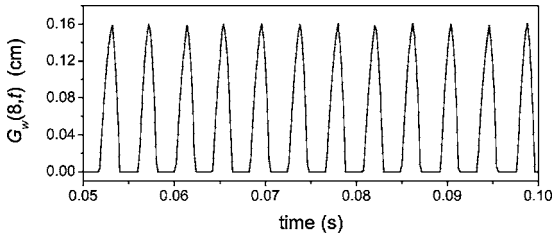


FIG. 2. The waveform of glottal width signal $G_w(8,t)$, where the vocal fold is symmetric, i.e., $Q_{\text{stf}}=1$ and $Q_{\text{shp}}=1$.

$$L_{\text{corr}} = \frac{1}{L} \| y_a|_{C_{\max}(y_a)=0.9} - y_p|_{C_{\max}(y_p)=0.9} \| \times 100\,\%, \qquad (4)$$

where y_a and y_p represent the two spatial points close to the anterior and posterior ends at which $C_{\max}(y_a)$ and $C_{\max}(y_p)$ are decreased to 0.9 with respect to the reference layer (middle layer). A larger correlation length value corresponds to a larger size of spatially symmetric pattern.

The rich information provided by a finite-element model allows us to reproduce the spatiotemporal glottal pattern of vocal fold vibration. Figure 3 illustrates spatiotemporal plots of the glottal width signal for AP shape symmetric vocal folds ($Q_{\text{shp}}=1$). The vertical axis is spatial position and the horizontal axis is time. i denotes the layer index in the AP direction (see Fig. 1). The gray level denotes the value of the glottal width $G_w(i,t)$. The bright region corresponds to a large G_w value, which indicates a wider glottis, whereas the dark region corresponds to a small G_w value, which indicates a narrow or closed glottis. In Fig. 3(a), the stiffness of the anterior and posterior sides is symmetric, i.e., $Q_{\text{stf}}=1$. We can see that the spatiotemporal pattern of a symmetric vocal fold is also symmetric. The fundamental frequencies of the anterior side and posterior side are identical ($F_{\text{ap}}=1$), and the AP biphonation phenomenon does not appear. However, when the anterior-posterior stiffness becomes asymmetric,

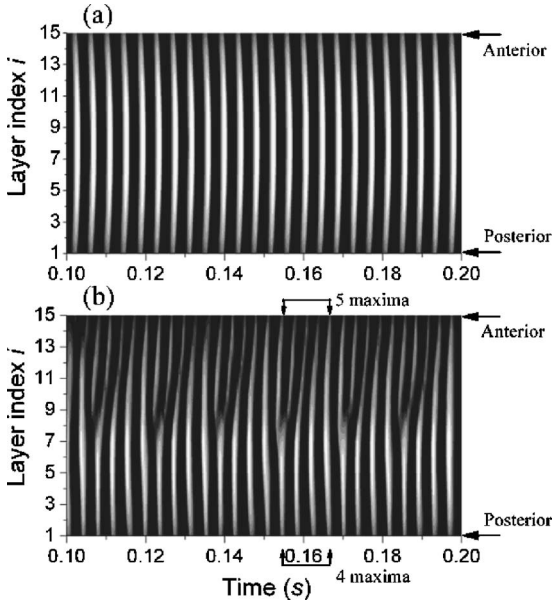


FIG. 3. The spatiotemporal plots of the glottal width signal $G_w(i,t)$ for AP shape symmetric vocal folds ($Q_{\text{shp}}=1$). (a) An AP stiffness symmetric vocal fold ($Q_{\text{stf}}=1$) and (b) an AP stiffness asymmetric vocal fold ($Q_{\text{stf}}=2$).

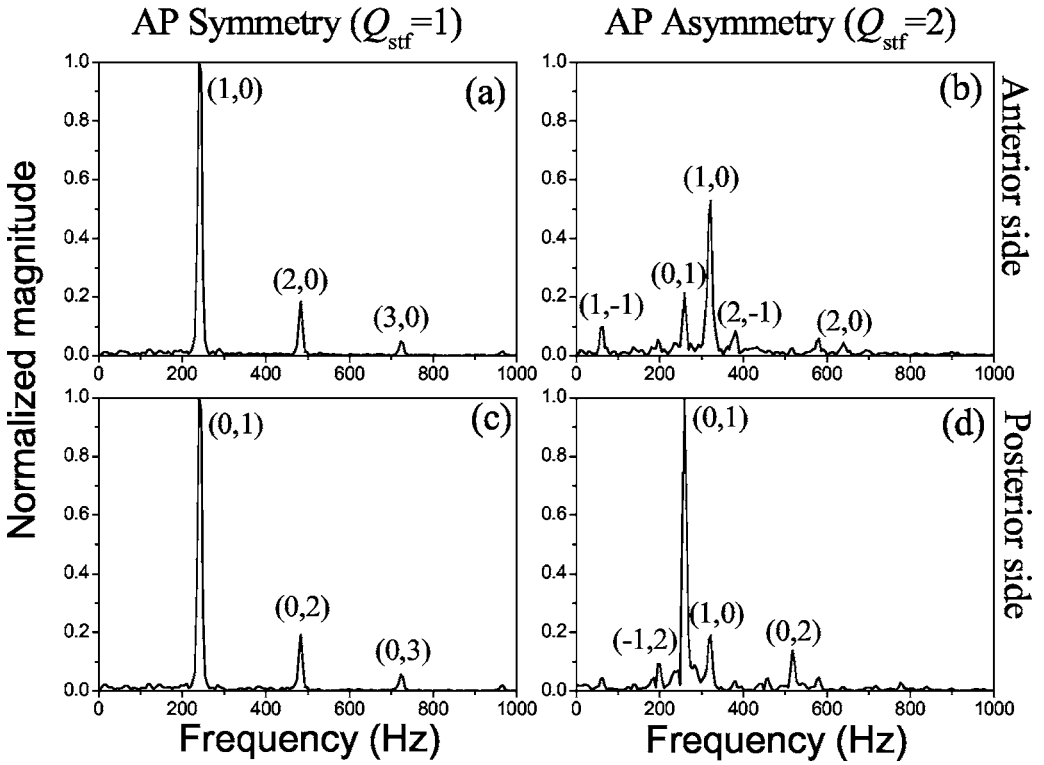


FIG. 4. The spectrum of surface nodal oscillation of the vocal fold. The left column and right column correspond to an AP stiffness symmetric vocal fold (Q_{stf}=1) and an AP stiffness asymmetric vocal fold (Q_{stf}=2), respectively. The upper row and the lower row correspond to the anterior (layer index i=11) and posterior sides (layer index i=4) of vocal folds, respectively. (m,n) labeled on the peak frequencies denotes that its frequency can be expressed as the linear superposition of two independent frequencies: $f_{(m,n)}=mf_a+nf_p$, where f_a indicates the fundamental frequency of anterior side and f_p indicates the fundamental frequency of posterior side. For the stiffness asymmetric vocal fold [(b) and (d)], f_a=322 Hz and f_p=259 Hz, which denotes a fundamental frequency ratio $F_{ap}\approx 1.24$.

the spatial symmetry of vocal fold vibration will be broken. Figure 3(b) gives the spatiotemporal plots with an asymmetric stiffness (Q_{stf}=2). It is seen that the anterior side and the posterior side of the vocal fold oscillate with different fundamental frequencies. While the maximum number of oscillation on the anterior side reaches 5, there is only an oscillation maximum of 4 on the posterior side. Moreover, these numbers do not depend on the temporal window used for counting them, which indicates an AP frequency asymmetry ratio of $F_{ap}\approx 5/4$. Thus anterior-posterior biphonation occurs.

Figure 4 presents the spectrum of surface nodal oscillation. The left column of Fig. 4 [Figs. 4(a)–4(c)] corresponds to AP stiffness symmetric vocal folds (Q_{stf}=1), and the right column [Figs. 4(b) and 4(d)] corresponds to AP stiffness asymmetric vocal folds (Q_{stf}=2). The upper row and the lower row correspond to the anterior (layer index i=11) and posterior side (layer index i=4), respectively. For a symmetric vocal fold [Figs. 4(a) and 4(c)], the spectrum structures for posterior and anterior sides are almost identical. The anterior and posterior sides have the same fundamental frequency of approximately 259 Hz. The other peak frequencies are the higher-order harmonics of the fundamental frequency. However, for an asymmetric vocal fold [Figs. 4(b) and 4(d)], the spectrum structures become complex, and the spectra of anterior side [Fig. 4(b)] and posterior side [Fig. 4(d)] are significantly different. Two independent fundamental frequencies appear in the spectrum at 259 and 322 Hz. The dominant frequency of the anterior side is f_a=322 Hz, but the dominant frequency of the posterior side is f_p=259 Hz. The other peak frequencies that appear in Figs. 3(b) and 3(d) can be interpreted as linear combinations of f_a and f_p:

TABLE II. Cumulative sum $\Sigma\lambda_m^2$ for the first five values of the relative EOF weights.

Eigenfunction index	Cumulative sum $\Sigma\lambda_m^2$ (%)		
	Symmetry	Stiffness asymmetry (Q_{stf}=2)	Shape asymmetry (Q_{shp}=4)
1	94.8	68.5	80.3
2	99.5	84.2	92.0
3	99.8	94.3	97.8
4	99.9	96.5	99.2
5	99.9	98.2	99.6

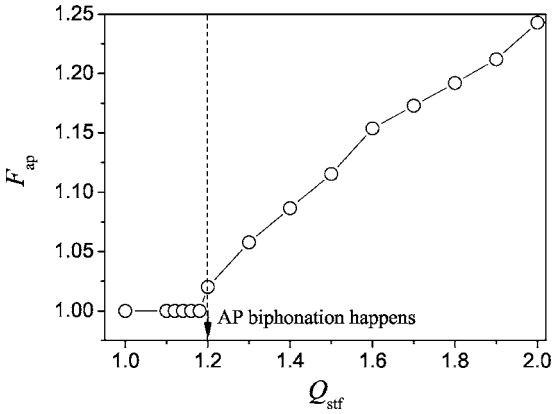


FIG. 5. The relationship between the AP fundamental frequency ratio F_{ap} and the AP stiffness asymmetric factor Q_{stf}.

$$f_{(m,n)} = mf_a + nf_p, \quad m,n \in Z = \{0, \pm 1, \pm 2, \ldots\}. \tag{5}$$

These simulation results are in accord with the experimental observations of Neubauer *et al.*[16] Furthermore, the technique of empirical orthogonal functions (EOF) is used to decompose the glottal width signal $G_w(i,t)$ into principal modes of vibration.[7,16] The normalized eigenvalue λ_m^2 represents the contribution of the mth modes to the complete spatial dynamics. Table II presents the cumulative sum $\Sigma\lambda_m^2$ of the first five normalized eigenvalues. For a symmetrical vocal fold, the first eigenmode has almost 95% of the total energy. However, for a tension asymmetrical vocal fold, the first eigenmode has only 68.5% of the total energy because of the existence of the higher-order AP modes. The above results show that the occurrence of AP biphonation and higher-order AP modes is associated with AP stiffness asymmetry.

Figure 5 illustrates the relationship between the AP fundamental frequency ratio F_{ap} and the AP stiffness asymmetric factor Q_{stf}. It is seen that F_{ap} is strongly dependent on Q_{stf}. For an AP symmetric vocal fold, F_{ap} is 1. For a vocal fold with slight AP stiffness asymmetry $Q_{stf} < 1.2$ for our model, F_{ap} still maintains a value of 1 and biphonation does not appear. However, when the AP stiffness asymmetry factor Q_{stf} exceeds a threshold value, F_{ap} is no longer equivalent to 1 and biphonation occurs. When Q_{stf} continues to increase, the fundamental frequency difference between the anterior and posterior sides also increases, thus indicating a more significant AP asymmetry.

Figures 6 and 7 give the relationship between stiffness asymmetry and the spatial coherence of vocal fold vibration, where $Q_{shp}=1$ indicates that the shape of vocal fold is symmetric. Figure 6 presents the curves of $C_{max}(i)$ versus layer index i, where the square, circle, and triangle symbols correspond to $Q_{stf}=1$, 1.14, and 2, respectively. Figure 7 presents the relationship between correlation length and the stiffness asymmetric factor Q_{stf}. The symmetric vocal fold vibrates in a high spatial coherence and its correlation length almost approaches 100%. With the increase in AP stiffness asymmetry (for example, $Q_{stf}=1.14$), the spatial coherence of vocal fold vibration decreases, and L_{corr} reduces to about 60%. However, the spatial inhomogeneity of vocal fold vibration is not serious enough to desynchronize the modes between anterior side and posterior side. Therefore, in such a condition, the anterior and posterior sides vibrate with the same fundamental frequency, and biphonation has not yet occurred, as shown in Fig. 5. With a greater increase of Q_{stf} (for example, $Q_{stf}=2$), the correlation length decreases to 10%, and serious spatial inhomogeneity of vocal fold vibration results in AP biphonation. In Figs. 6 and 7, because the shape of the vocal fold was symmetric, the influence of shape asymmetry on vocal fold vibration has been artificially excluded. The AP stiffness asymmetry is the only asymmetric factor existing in this model. Therefore, the decrease of correlation length shown in Figs. 6 and 7 is purely due to the stiffness asymmetry, which demonstrates that the AP stiffness imbalance is one important reason for the drop in the spatial coherence of vocal fold vibration which causes AP biphonation.

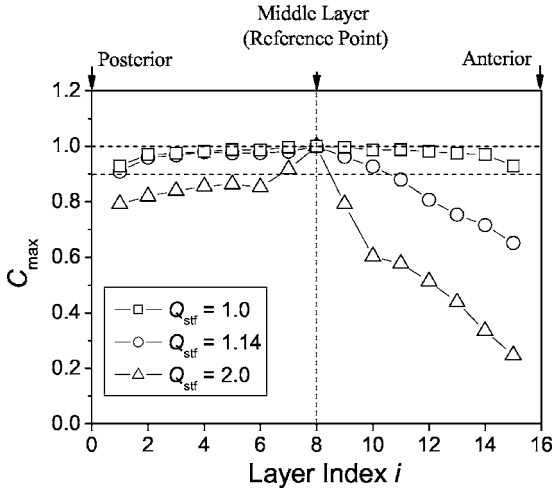


FIG. 6. Spatial correlation function $C_{max}(i)$ versus layer index i for different AP stiffness asymmetric factors, where square, circle, and triangle marks correspond to $Q_{stf}=1$, 1.14, and 2, respectively.

Shape asymmetry is another important asymmetric factor of the vocal folds. In our study, the parameter Q_{shp} is used to quantitatively describe the shape asymmetry of the vocal folds. A Q_{shp} value of 1 indicates an AP shape symmetric vocal fold. $Q_{shp} > 1$ indicates AP shape asymmetry by showing that the depth of the anterior side of the vocal fold is smaller than that of the posterior side. Figure 8 presents the curve of $C_{max}(i)$ versus layer index i with different shape

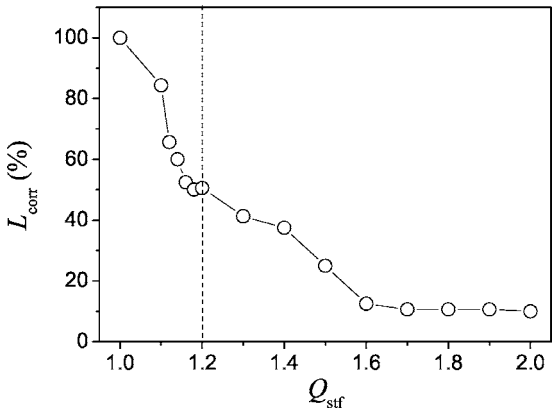


FIG. 7. The relationship between correlation length L_{corr} and stiffness asymmetric factor Q_{stf}.

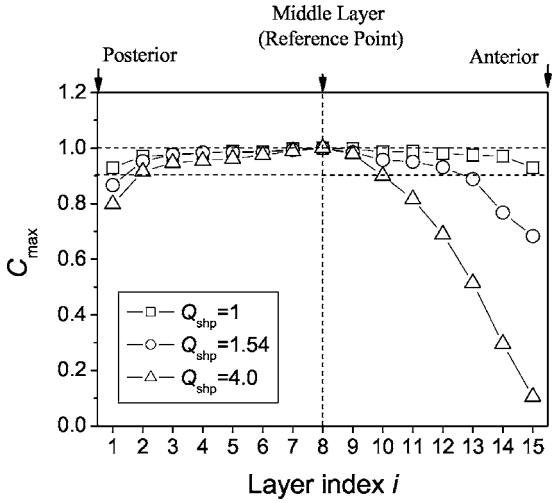


FIG. 8. Spatial correlation function $C_{max}(i)$ versus layer index i for different AP shape asymmetric factors, where square, circle, and triangle marks correspond to Q_{stf}=1, 1.54, and 4, respectively.

asymmetric factors Q_{shp}, where the square, circle, and triangle symbols correspond to Q_{shp}=1, 1.54, and 4, respectively. Figure 9 presents the relationship between correlation length and the asymmetric shape factor. It is seen that the anterior-posterior shape asymmetry of the vocal fold also potentially decreases the spatial coherence of vocal fold vibration. For a symmetric vocal fold, vocal fold vibration is symmetric as well, and the correlation length almost approaches 100%, where the slight C_{max} difference between the layers in a symmetric model can be explained by the weak high-order mode due to vocal fold collision, tissue interaction in the vocal folds, strong airflow-tissue interaction, and so on. When Q_{shp}=1.54, the correlation length is about 70% and AP biphonation cannot be found in such an asymmetric shape condition. Therefore, although it was clinically observed that the depth of the human vocal fold usually decreases from the posterior to the anterior part, i.e., the vocal fold is AP asymmetric, AP biphonation usually cannot occur in a normal vocal fold with slight shape asymmetry. When Q_{shp}=4, the correlation length further decreases to about 47%. The seriously broken spatial coherence of vocal fold vibration produces AP biphonation. Its spatiotemporal plot is

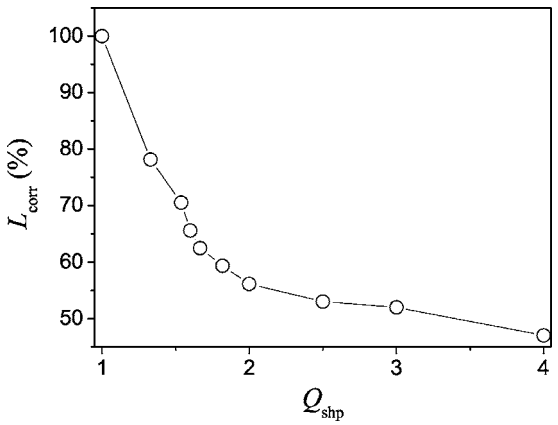


FIG. 9. The relationship between correlation length L_{corr} and shape asymmetric factor Q_{shp}.

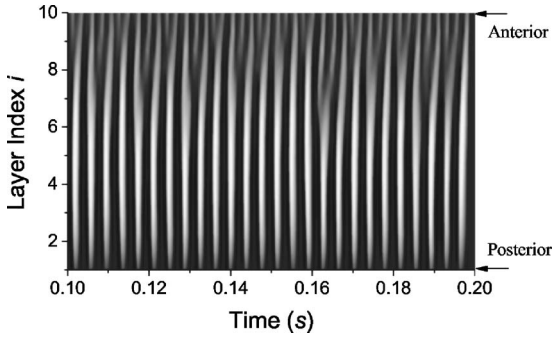


FIG. 10. The spatiotemporal plots of the glottal width signal $G_w(i,t)$ for the vocal fold with an AP asymmetry shape of Q_{shp}=4, but an AP symmetric stiffness of Q_{stf}=1.

presented in Fig. 10 and the corresponding cumulative sum $\Sigma\lambda_m^2$ of the first five normalized eigenvalues is given in Table II. It is seen that the fundamental frequencies of the posterior and anterior sides are no longer identical, and higher-order AP modes exist in the vibration. In Figs. 8–10, because the stiffness of the vocal fold was symmetric, stiffness asymmetry did not contribute to the decrease of the spatial coherence of vocal fold vibration. Therefore, shape asymmetry makes the contribution to the decrease of correlation length and AP biphonation in this example.

IV. DISCUSSION AND CONCLUSION

The goal of this paper is to analyze the spatiotemporal characteristics and explore the potential reasons for anterior-posterior biphonation. In this paper, we use a finite-element model to simulate vocal fold vibration and reproduce the spatiotemporal glottal pattern. A spatiotemporal plot, spectral analysis, AP fundamental frequencies ratios, correlation function, and correlation length are used to describe the spatiotemporal characteristics of the glottal pattern.

Our simulation found that anterior-posterior stiffness asymmetry can possibly cause anterior-posterior biphonation. Spectral analysis shows that there are two independent fundamental frequencies for an anterior-posterior stiffness asymmetric vocal fold. The anterior side and posterior side of the vocal folds oscillate with independent fundamental frequencies (f_a and f_p). The two independent fundamental frequencies indicate two independent vibration modes. The interaction between the two independent modes generates a complex spectrum structure. The apparent peak frequencies in the spectrum can be explained as the linear combination of the two independent fundamental frequencies (f_a and f_p). This is the so-called "AP biphonation." The calculation of the empirical orthogonal functions proves the existence of higher-order AP modes in the vibration of the asymmetrical vocal fold.

We then studied the influence of stiffness asymmetry on fundamental frequency ratios and correlation length. The results show that correlation length has a close relationship with stiffness asymmetry. The correlation length decreases with an increase in stiffness asymmetry. When AP stiffness asymmetry reaches a certain level ($Q_{stf}>1.2$ in our model), AP biphonation occurs, and the fundamental frequency ratio increases simultaneously with AP stiffness asymmetry. The

stiffness asymmetry describes the inhomogeneity of vocal fold tissue. The correlation length and fundamental frequency ratios describe the decrease in the spatial coherence of vocal fold vibrations. We can therefore consider that AP stiffness asymmetry is one possible reason for the spatial inhomogeneity of vocal fold vibration. Serious spatial inhomogeneity of vocal fold vibration results in mode desynchronization between the anterior side and the posterior side and generates AP biphonation. These results strongly support the hypothesis that "inhomogeneities of the vocal fold tissue properties like local, hidden morphological changes in deeper tissue layers could induce and support the independent vibration of modes."[16] In other words, we can also consider that the independent vibration of modes and the occurrence of AP biphonation sometimes may imply that the inhomogeneities of vocal fold tissue properties are due to inflammation, localized edema, or scarring, which are commonly seen clinically. Therefore, the independent vibration of modes could provide useful clinical information in assessing laryngeal function.

The vocal fold shape can be affected in many pathological conditions such as the occurrence of tumors and infection. It is important to know how the shape of the vocal folds changes the vibration. In this study, besides stiffness asymmetry, we found that shape asymmetry is another potential reason for spatial inhomogeneities of vocal fold vibration. Shape asymmetry could also induce AP biphonation. The anterior side, with a smaller vocal fold depth, has a higher fundamental frequency. Conversely, the posterior side, with a larger vocal fold depth, has a lower fundamental frequency. Usually, greater stiffness and smaller size correspond to a higher fundamental frequency vibration mode. Lower tension and larger size correspond to a lower fundamental frequency vibration mode.[49,51,52] Therefore, the smaller depth in the anterior side of the vocal fold plays a role similar to greater stiffness. The AP asymmetry of the vocal fold provides the potential for a multi-mode coexistence in the vocal folds and breaks down the spatial coherence of vocal fold vibration. When the asymmetry of the vocal fold reaches a certain level, the vibration modes of the anterior and posterior sides will be desynchronized.

Therefore, we can draw the following conclusion: just as the reported left-right asymmetry of the vocal folds leads to a desynchronized left-right vibration and thus left-right biphonation,[5,12,14] the anterior-posterior asymmetry of vocal folds is one potential reason for a desynchronized anterior-posterior vibration of the vocal folds and anterior-posterior biphonation.

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.

[1]W. Wende, H. Herzel, and K. Wermke, "Bifurcations and chaos in newborn cries," Phys. Lett. A **145**, 418–424 (1990).
[2]I. R. Titze, R. Baken, and H. Herzel, "Evidence of chaos in vocal fold vibration," in *Vocal Fold Physiology: New Frontier in Basic Science*, edited by I. R. Titze (Singular, San Diego, 1993), pp. 143–188.
[3]J. C. Lucero, "Dynamics of the two-mass model of the vocal folds: Equilibria, bifurcations, and oscillation region," J. Acoust. Soc. Am. **94**, 3104–3111 (1993).
[4]H. Herzel and C. Knudsen, "Bifurcation in a vocal fold model," Nonlinear Dyn. **7**, 53–64 (1995).
[5]I. Steinecke and H. Herzel, "Bifurcations in an asymmetric vocal-fold model," J. Acoust. Soc. Am. **97**, 1874–1884 (1995).
[6]D. A. Berry, H. Herzel, I. R. Titze, and B. H. Story, "Bifurcations in excised larynx experiments," J. Voice **10**, 129–138 (1996).
[7]D. A. Berry, H. Herzel, I. R. Titze, and K. Krischer, "Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions," J. Acoust. Soc. Am. **95**, 3595–3604 (1994).
[8]J. J. Jiang, Y. Zhang, and J. Stern, "Modeling of chaotic vibrations in symmetric vocal folds," J. Acoust. Soc. Am. **110**, 2120–2128 (2001).
[9]J. J. Jiang and Y. Zhang, "Chaotic vibration induced by turbulent noise in a two-mass model of vocal folds," J. Acoust. Soc. Am. **112**, 2127–2133 (2002).
[10]Y. Zhang and J. J. Jiang, "Chaotic vibrations of a vocal fold model with a unilateral polyp," J. Acoust. Soc. Am. **115**, 1266–1269 (2004).
[11]M. E. Smith, G. S. Berke, B. R. Gerrat, and J. Kreiman, "Laryngeal paralyses: Theoretical considerations and effects on laryngeal vibration," J. Speech Hear. Res. **35**, 545–554 (1992).
[12]P. Mergell and H. Herzel, "Modeling biphonation—the role of the vocal tract," Speech Commun. **22**, 141–154 (1997).
[13]I. Wilden, H. Herzel, G. Peters, and G. Tembrock, "Subharmonics biphonation, and deterministic chaos in mammal vocalization," Bioacoustics 9, 171–196 (1998).
[14]P. Mergell, H. Herzel, and I. R. Titze, "Irregular vocal-fold vibration—High-speed observation and modeling," J. Acoust. Soc. Am. **108**, 2996–3002 (2000).
[15]T. Riede, H. Herzel, D. Mehwald, W. Seidner, E. Trumler, G. Böhme, and G. Tembrock, "Nonlinear phenomena in the natural howling of a dog-wolf mix," J. Acoust. Soc. Am. **108**, 1435–1442 (2000).
[16]J. Neubauer, P. Mergell, U. Eysholdt, and H. Herzel, "Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial modes," J. Acoust. Soc. Am. **110**, 3179–3192 (2001).
[17]C. H. Brown, F. Alipour, and D. A. Berry, "Laryngeal biomechanics and vocal communication in the squirrel monkey (*Saimiri boliviensis*)," J. Acoust. Soc. Am. **113**, 2114–2126 (2003).
[18]Y. Zhang and J. J. Jiang, "Spatiotemporal chaos in excised larynx vibrations," Phys. Rev. E **72**, 035201(R) (2005).
[19]D. Wong, M.R. Ito, and N. B. Cox, "Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases," J. Acoust. Soc. Am. **89**, 383–394 (1991).
[20]I. R. Titze, "Mechanical stress in phonation," J. Voice **8**, 99–105 (1994).
[21]S. M. Zeitels, "Phonosurgery—past, present, and future," Operative Tech. Otolaryngol. Head Neck Surg. **9**, 179 (1998).
[22]N. Isshiki, "Mechanical and dynamic aspects of voice production as related to voice therapy and phonosurgery," J. Voice **12**, 125–137 (1998).
[23]Y. Zhang, C. McGilligan, L. Zhou, M. Vig, and J. J. Jiang, "Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps," J. Acoust. Soc. Am. **115**, 2270–2277 (2004).
[24]Y. Zhang and J. J. Jiang, "Nonlinear dynamic analysis in signal typing of pathological human voices," Electron. Lett. **39**, 1021–1023 (2003).
[25]J. J. Jiangand I. R. Titze, "A methodological study of hemilaryngeal phonation," Laryngoscope **103**, 872–882 (1993).
[26]D. A. Berry, D.W. Montequin, and N. Tayama, "High-speed digital imaging of the medial surface of the vocal folds," J. Acoust. Soc. Am. **110**, 2539–2547 (2001).
[27]H. Herzel, D. Berry, I. R. Titze, and M. Saleh, "Analysis of vocal disorders with methods from nonlinear dynamics," J. Speech Hear. Res. **37**, 1008–1019 (1994).
[28]A. Kumar and S. K. Mullick, "Nonlinear dynamical analysis of speech," J. Acoust. Soc. Am. **100**, 615 (1996).
[29]I. Hertrich, W. Lutzenberger, S. Spieker, and H. Ackermann, "Fractal dimension of sustained vowel productions in neurological dysphonias: An acoustic and electroglottographic analysis," J. Acoust. Soc. Am. **102**, 652–654 (1997).
[30]A. Behrman and R. J. Baken, "Correlation dimension of electroglottographic data from healthy and pathologic subjects," J. Acoust. Soc. Am. **102**, 2371–2379 (1997).

[31]A. Behrman, "Global and local dimensions of vocal dynamics," J. Acoust. Soc. Am. **105**, 432–443 (1999).
[32]I. Tokuda, T. Riede, J. Neubauer, M. J. Owren, and H. Herzel, "Nonlinear analysis of irregular animal vocalizations," J. Acoust. Soc. Am. **111**, 2908–2919 (2002).
[33]K. Ishizaka and J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**, 1233–1268 (1972).
[34]K. Ishizaka and N. Isshiki, "Computer simulation of pathological vocal-cord vibration," J. Acoust. Soc. Am. **60**, 1193–1198 (1976).
[35]N. Isshiki, M. Tanabe, K. Ishizaka, and D. Broad, "Clinical significance of asymmetrical vocal cord tension," Ann. Otol. Rhinol. Laryngol. **86**, 58–66 (1977).
[36]I. R. Titze, "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536–1552 (1988).
[37]B. H. Story and I. R. Titze, "Voice simulation with a body-cover model of the vocal folds," J. Acoust. Soc. Am. **97**, 1249–1260 (1995).
[38]F. Alipour and I. R. Titze, "Simulation of particle trajectories of vocal fold tissue," in *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control*, edited by I. R. Titze and R. C. Scherer (Denver Center for the Performing Arts, Denver, 1985), pp. 183–190.
[39]F. Alipour and I. R. Titze, "A finite element simulation of vocal fold vibration," in *Proceedings of the Fourteenth Annual Northeast Bioengineering Conference*, Durham, NH, edited by J. R. LaCourse, pp. 186–189 (1988).
[40]F. Alipour and I. R. Titze, "Combined simulation of airflow and vocal fold vibrations," in *Vocal Fold Physiology, Controlling Complexity & Chaos*, edited by P. Davis and N. Fletcher (Singular, San Diego, 1996), pp. 17–29.
[41]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**, 603–610 (1998).
[42]F. Alipour and R. C. Scherer, "Vocal fold bulging effects on phonation using a biophysical computer model," J. Voice **14**, 470–483 (2000).
[43]H. E. Gunter, "A mechanical model of vocal-fold collision with high spatial and temporal resolution," J. Acoust. Soc. Am. **113**, 994–1000 (2003).
[44]J. C. Simo and T. A. Laursen, "An Augmented Lagrangian Treatment of Contact Problems Involving Friction," Comput. Struct. **42**(1), 97–116 (1992).
[45]R. H. Colton and J. K. Casper, *Understanding Voice Problems: A Physiological Perspective for Diagnosis and Treatment*, 2nd ed. (Lippincott Williams & Wilkins, New York, 1996).
[46]J. C. Stemple, L.E. Glaze, and B. G. Klaben, *Clinical Voice Pathology: Theory and Management*, 3rd ed. (Singular, Thomson Learning, San Diego, 2000).
[47]F. Alipour, D.A. Berry, and I. R. Titze, "A finite-element model of vocal-fold vibration," J. Acoust. Soc. Am. **108**, 3003–3012 (2000).
[48]R. T. Sataloff, *Professional Voice—The science and art of clinical care* (Singular, San Diego, 1997).
[49]I. R. Titze and D. T. Talkin, "A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation," J. Acoust. Soc. Am. **66**, 60–74 (1979).
[50]J. Loenardy, F. Kaiser, M. R. Belic, and O. Hess, "Running transverse waves in optical phase conjugation," Phys. Rev. A **53**, 4519–4527 (1996).
[51]I. R. Titze, "On the mechanics of vocal-fold vibration," J. Acoust. Soc. Am. **60**, 1366–1380 (1976).
[52]I. R. Titze and W. J. Strong, "Normal modes in vocal cord tissues," J. Acoust. Soc. Am. **53**, 736–744 (1975).

Theoretical assessment of unsteady aerodynamic effects in phonation

Michael H. Krane[a)]
Center for Advanced Information Processing, Rutgers University, Piscataway, New Jersey 08854

Timothy Wei[b)]
Department of Mechanical and Aerospace Engineering, Rutgers University, Piscataway, New Jersey 08854

(Received 17 March 2005; revised 24 May 2006; accepted 26 May 2006)

This paper ranks the importance of unsteady aerodynamic mechanisms in glottal flow. Particular emphasis is given to separation point motion, acceleration of glottal airflow by vocal fold motion, and viscous blockage. How nondimensional parameters such as the Reynolds, Strouhal, and Womersley numbers help in this ranking is also addressed. An equation of motion is derived which includes terms explicitly describing the effects of interest, assuming (1) a symmetrical glottis, (2) zero pressure recovery downstream of the vocal folds, and (3) a quasisteady glottal jet. Estimating the order of magnitude of the terms in this equation, it is shown that the flow is characterized by two temporal regimes: (1) a flow initiation/shutoff regime where local unsteady acceleration and wall motion dominate, and (2) a "quasisteady" regime where the flow is dominated by convective acceleration. In the latter case, separation point motion and viscous blockage are shown to be out of phase with motion of the vocal folds, thereby impacting the shape of the glottal volume flow waveform. The analysis suggests that glottal flow may be considered quasisteady only insofar as traditional assumptions concerning glottal jet behavior can be confirmed. [DOI: 10.1121/1.2215408]

PACS number(s): 43.70.Bk [AL] Pages: 1578–1588

I. INTRODUCTION

This paper presents a theoretical approach to determine the dynamic relevance of unsteady aerodynamic effects during phonation. The particular focus is on how time variations in (1) inertia, (2) viscous blockage, (3) vocal fold wall motion, and (4) separation point motion influence the shape of the glottal volume flow waveform, the physical determinant of voice power and quality. In doing so, how these effects might cause departures from the commonly assumed quasisteady model of glottal airflow is also addressed. While these effects have previously received attention in the literature, their relative importance and influence on the glottal volume flow waveform have never been assessed directly using the equations of motion.

Consider the flow through the vocal fold during a single cycle of vibration. A schematic of the flow is shown in Fig. 1. Upon opening, air is forced through the glottis by the transglottal pressure gradient. Near the glottal exit, the flow separates, forming the glottal jet. The diffusion of this jet causes the pressure loss which the transglottal pressure gradient must overcome. This airflow continues until the vocal folds close at the end of the cycle. The simplest description of this flow is that the glottal volume flow has a simple dependence on only the transglottal pressure and glottal area. This description, which has been used widely in glottal aerodynamic modeling (see, e.g., Pelorson *et al.*, 1994; Story and Titze, 1995; Lous *et al.*, 1998), implicitly assumes a quasisteady glottal jet with minimal pressure recovery (Hirschberg, 1992). In this model, because the glottal volume flow changes instantaneously in response to changes in either transglottal pressure or glottal area, airflow in the glottis is quasisteady. Several recent studies (Mongeau *et al.*, 1997; Zhang *et al.*, 2002; Vilain *et al.*, 2004) seem to support this approximation.

Several other factors also mediate the relationship between the driving pressure gradient, glottal area, and volume flow. Due to its inertia, a finite length of time is needed for the air in the glottis to accelerate from rest [this was accounted for in the early glottal flow models of Ishizaka and Flanagan (1972)]. Similarly, at the end of the cycle, when the folds close, the flow is shut off over a finite duration during which the flow decelerates to rest. Along the vocal fold walls, in the direction of the flow, there exist regions, the boundary layers, in which friction is important (see Fig. 1). These regions grow in thickness through diffusion and in response to time variations of glottal flow. Through the displacement effect (see, e.g., Rosenhead, 1963) the boundary layers can decrease the effective aerodynamic area of the glottis. This *blockage effect*, if large enough, increases the transglottal pressure difference needed to push air through the glottis at a given rate [see, e.g., discussion of entrance flow in a straight pipe in Bejan (1984)]. Near the glottal exit,

[a)]Current address: Applied Research Laboratory, Penn State University, University Park, PA 16804. Electronic mail: mhk5@only.arl.psu.edu
[b)]Current address: Department of Mechanical, Aerospace, and Nuclear Engineering, Rensselaer Polytechnic Institute, Troy, NY 121808.

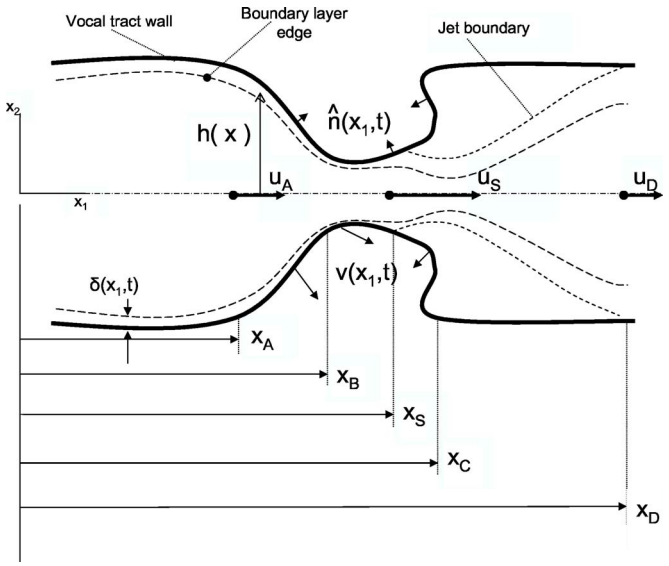


FIG. 1. Geometry of glottal flow passage. Trachea ends at $x_1=x_A$. x_B is the location of the glottis entrance, x_C of the glottis exit. The location of the separation point is $x_1=x_S(t)$, where $x_B<x_S<x_C$. The wall shape is described by the function $x_2=h(x_1,t)$. The boundary layer thickness is given by $x_2=\delta(x_1,t)$. The flow fills the passage until $x_1=x_S$, at which point the flow separates and forms a jet, whose boundary is shown. At the separation point, the boundary layer thickness is δ_S. The location where the jet reattaches to the wall is designated $x_1=x_D$.

the boundary layers separate from the vocal fold walls, forming the glottal jet. Separation is affected by the glottis shape, volume flow, and boundary layer thickness, as well as the rate at which these factors change in time (Hirschberg, 1992; McGowan, 1993). Finally, the motion of the vocal fold walls displaces air in the glottis, exchanging momentum with it (Flanagan and Ishizaka, 1976; Deverge *et al.*, 2003). Previous consideration of these phenomena has made use of nondimensional parameters such as the Reynolds number or the Strouhal number. Bejan (1984) has pointed out that the interpretation of these parameters is highly dependent on the flow being studied. A careful consideration of the meaning of nondimensional parameters in phonatory airflows has not been done.

In this article, a theoretical approach is taken to address these issues. An equation of motion is derived which explicitly includes the phenomena in question. Using the scales of motion observed in phonation, the order of magnitude of each of the terms in the equation is estimated. Assessing the order of magnitude of each term relative to transglottal pressure provides a clear basis for ranking each effect in terms of its dynamic importance. Using these estimates, the relevant time scales governing each effect are also revealed. The analysis also allows interpretation of nondimensional numbers such as the Reynolds, Strouhal, and Womersley numbers. Neglecting terms two orders of magnitude smaller than transglottal pressure leads to a simplified equation of motion. The solution of this equation is used to express the glottal volume flow in terms of the leading order effects.

II. APPROXIMATE DESCRIPTION OF GLOTTAL AERODYNAMICS

A. Problem definition

The geometry of the airflow through the vocal folds is shown in Fig. 1. The airflow proceeds from left to right through the constriction leading from the cylindrical trachea through the glottis, a rectangular slit, and into the cylindrical pharynx as shown. The coordinate axes show x_1 as the axial coordinate and x_2 the transverse coordinate. The vocal folds lie between $x_A<x_1<x_C$, and their surfaces are defined by the function $x_2=h(x_1,t)$ and the components $n_i(x_1,t)$ of the wall unit normal vector $\boldsymbol{n}$.

The local cross-sectional area $S(x_1)$ is defined as $S(x_1)=2L_Gh(x_1)$, where L_G is the length of the glottis in the direction perpendicular to the page in Fig. 1. The presence of boundary layers along the walls means the flow "sees" a smaller cross-sectional area, $S'(x_1)$. The local flow cross-sectional area, $S'(x_1)$, corrected for viscous blockage, is defined as $S'(x_1)=2(1-\delta^*(x_1)/h(x_1))L_Gh(x_1)$. Here, δ^* is the boundary layer displacement thickness, which depends on the shape of the local axial velocity profile (see, e.g., Rosenhead, 1963; Bejan, 1984), and which scales on the local boundary layer thickness δ.

The locations shown in the sketch denote the regions of the airflow: For $x_1<x_A$, the air particle velocity is small enough that the motion is linear and essentially acoustic. The locations x_B and x_C denote the inlet and outlet, respectively, of the glottis. The axial location $x_1=x_S$, downstream of the point of minimum constriction, is where the airflow separates from the wall and a planar jet is formed. In the present work separation point behavior will not be computed, but its behavior will assumed to be consistent with that described previously (Hirschberg, 1992; Pelorson *et al.*, 1994; Liljencrants, 1996; Lous *et al.*, 1998; Alipour and Scherer, 2004), in which the separation point moves rapidly during glottal opening and closure. The separation point location is important because it indicates not only the distribution of aerodynamic forces driving vocal fold vibration, but also the volume of air allowed to pass through the glottis. In other words, glottal volume flow $U_g=u_SS'_S$. Thus the movement of the separation point helps determine the temporal behavior of U_g. At some point further downstream, $x_1=x_D$, the jet has diffused enough, through the action of friction and turbulent mixing, to reattach to the vocal tract walls.

The airflow in the region $x_A<x_1<x_D$ is analyzed using the following assumptions:

(1) *The tracheal and pharyngeal cross-sectional areas* (S_A *and* S_D, *respectively*) *are assumed approximately equal, and much larger than the glottal area*. In other words, $S_A\approx S_D\gg S_S$.

(2) *The airflow is incompressible*, equivalent to saying that the airflow fluctuations are slow enough that they are communicated to the rest of the region in question during a time interval much shorter than the period of the fluctuation. In other words, the region of interest is acoustically compact—see, e.g., Hirschberg (1992).

(3) *Vocal fold shape, its motion, and the airflow are symmetrical about the mid-saggital plane*. Although there are indications (see, e.g., Hofmans *et al.*, 2003) that the glottal jet becomes asymmetrical, for purposes of this work, where orders of magnitude are sought, symmetry is assumed. Note that this previous work does not indicate asymmetric flow upstream of the separation point for a symmetrical glottis.

(4) *No pressure recovery downstream of the separation point* (*in the region* $x_S<x_1<x_D$). *In addition, this process is quasisteady*. This assumption has been widely used in phonation modeling (Ishizaka and Flanagan, 1972; Peloroson *et al.*, 1994; Story and Titze, 1995). This assumption approximates the air pressure downstream of the point of separation to be uniform and equal to ambient pressure p_D. In other words, the pressure $p_S=p_C=p_D$, so that the transglottal pressure difference $p_A-p_D\approx p_A-p_S$. This approximation implicitly assumes that the jet flow is established very rapidly compared to the other time scales of the flow and of vocal fold motion. Note that if this assumption is correct, any asymmetric jet motions that may occur (see, e.g., Hofmans *et al.*, 2003) are not dynamically relevant.

(5) *The flow in the glottal flow boundary layers is laminar*. While this assumption finds support from the experiments of Hofmans *et al.* (2003), Deverge *et al.* (2003), and Vilain *et al.* (2004), it appears to be unresolved at present. For present purposes, this assumption is sufficient—a turbulent boundary layer will exhibit a higher rate of transverse diffusion and a different separation point location, but qualitatively the boundary layer behaves in the same manner (Bejan, 1984). Note that the boundary layers being laminar does not preclude the glottal jet from transitioning to a fully turbulent flow.

(6) *Separation point motion occurs during brief intervals (compared to the time the glottis is open), and moves a distance of order of the axial length of the glottis*. This assumption has been used previously (Hirschberg, 1992; Liljencrants, 1996; Lous *et al.*, 1998). Although the work of Pelorson *et al.* (1994) and Alipour and Scherer (2004) conflict in terms of the direction of separation point motion, both support this assumption. In the present analysis the effect of separation point motion in both directions is discussed.

(7) *Convective acceleration in the converging glottis, between the trachea and the separation point, is the dominant mechanism by which transglottal pressure gradient accelerates the airflow*. This is equivalent to saying that the quasisteady Bernoulli equation, in which the axial acceleration of the flow in the glottis inlet is balanced by transglottal pressure, is the lowest-order model of phonatory airflow, as described in the Introduction. As a result, all other phenomena are expected to be corrections to this model. This assumption is consistent with the success of past efforts to model vocal fold vibration.

B. Derivation of equation describing glottal flow

The glottal flow equation is derived in a manner similar to the derivation of Bernoulli's equation, but for a more general situation. First, the axial momentum equation is integrated along the axial direction in the region $x_A<x<x_S$ extending from the glottis entrance to the moving separation point. Then, mass conservation is used to express all appearances of velocity in terms of the velocity at the separation point.

The momentum equation along the symmetry streamline ($x_2=0$) is given by

$$\rho\left(\frac{\partial u}{\partial t}+\frac{1}{2}\frac{\partial u^2}{\partial x_1}\right)+\frac{\partial p}{\partial x_1}=0, \tag{1}$$

where ρ is the fluid density, u is the axial component of velocity, p is the static pressure, and x_1 is the coordinate in the axial direction. Integrating this equation along the symmetry streamline from $x_1=x_A$ to $x_1=x_S(t)$, the following equation is obtained:

$$\rho\frac{d}{dt}\int_{x_A}^{x_S(t)}u(x_1)\,dx_1-\rho u_S\dot{x}_S+p_S-p_A+\frac{\rho}{2}(u_S^2-u_A^2)=\frac{\int_{x_A}^{x_S(t)}\tau_W(x_1)\,dx_1}{S_{\text{wall}}} \tag{2}$$

where Liebnitz's rule (see Panton, 1994) was used to move the time derivative outside the integral in the first term, resulting in the appearance of a new term which explicitly includes the speed of motion of the separation point. An additional term containing τ_W, the wall shear stress, also appears on the right-hand side, representing the head loss due to friction on the glottis walls. The mass conservation equation (see the Appendix) yields

$$u(x)=\frac{\dot{V}_{xS}}{S'(x)}+(u_S-\dot{x}_S)\frac{S'_S}{S'(x)}. \tag{3}$$

Here, $\dot{V}_{x_S}$ is the time rate of change of glottal volume from $x_1=x$ to x_S, and $S'(x)$ is the cross-sectional area of the glottis at x_1 (corrected for viscous blockage), to rewrite the integrand. Using this result, the unsteady acceleration term becomes

$$\rho\frac{d}{dt}\int_{x_A}^{x_S(t)}u(x)\,dx=\rho\frac{d}{dt}\int_{x_A}^{x_S(t)}\frac{\dot{V}_{xS}}{S(x)}\,dx+\rho\frac{d}{dt}\left((u_S-\dot{x}_S)\int_{x_A}^{x_S(t)}\frac{S'_S}{S(x)}\,dx\right). \tag{4}$$

Now the mass conservation equation is used to rewrite the u_A^2 term in Eq. (2) in terms of u_S and the speed of the separation point. After some rearrangement, the following equation results:

$$\underset{\text{I}}{\rho\frac{d}{dt}(u_S L_{\text{eff}})} \quad \underset{\text{II}}{-\rho\frac{d}{dt}(\dot{x}_S L_{\text{eff}})} \quad \underset{\text{III}}{+\rho\frac{d}{dt}\int_{x_A}^{x_S(t)} \frac{\dot{V}_{Ax}}{S(x,t)}\,dx} \quad \underset{\text{IV}}{-\rho u_S \dot{x}_S} \quad \underset{\text{V}}{+\rho\frac{u_S^2}{2}\left(1-\left(\frac{S_S'}{S_A}\right)^2\right)} \quad \underset{\text{VI}}{-\rho u_S \frac{2S_S'}{S_A^2}\dot{V}_{AS}} \quad \underset{\text{VII}}{-\rho u_S \frac{2S_s'^2}{S_A^2}\dot{x}_S} \quad \underset{\text{VIII}}{-\rho\frac{\dot{V}_{AS}^2}{S_A^2}} \quad \underset{\text{IX}}{-2\frac{\dot{V}_{AS}}{S_A^2}S_{S'}\dot{x}_S} \quad \underset{\text{X}}{+\frac{S_S'^2}{S_A^2}\dot{x}_S^2} \quad \underset{\text{XI}}{+p_S - P_A} = \underset{\text{XII}}{\frac{\int_{x_A}^{x_S(t)} \tau_W(x)\,dx}{S_{\text{wall}}}} \tag{5}$$

where

$$L_{\text{eff}} = \int_{x_A}^{x_S(t)} \frac{S_S}{S(x)}\,dx \sim L \tag{6}$$

(see, e.g., Pelorson *et al.*, 1994). The first four terms in Eq. (5) represent unsteady acceleration of the air in the glottis. From left to right, these terms represent (I) unsteady acceleration of glottal volume flow, (II) a correction term due to the unsteady motion of the separation point, (III) acceleration due to the motion of the vocal fold walls, and, finally, (IV) a second correction due to the motion of the separation point. The first term (I) is the familiar unsteady acceleration term that appears even when vocal fold wall and separation point motion is neglected. Terms V–X represent convective acceleration of the air in the glottis, including contributions from movement of the walls (VI, VIII, and IX) and the separation point (VII, IX, and X). The last two terms are the two axial forces acting on the flow: (XI) the transglottal pressure, where assumption 3 has been invoked so that $p_A - p_D = p_A - p_S$, and (XII) the net friction force. Note that vocal fold wall motion, separation point motion, and blockage ratio contribute not only to the unsteady acceleration, but also to the convective acceleration.

If the effects of wall motion and separation point movement are neglected, then Eq. (5) reduces to the familiar unsteady Bernoulli equation, as expressed in Hofmans *et al.* (2003), supplemented by friction:

$$\rho\frac{d}{dt}(u_S L_{\text{eff}}) + \rho\frac{u_S^2}{2}\left[1-\left(\frac{S_S'}{S_A}\right)^2\right] + p_S - p_A = \frac{\int_{x_A}^{x_S(t)} \tau_W(x)\,dx}{S_{\text{wall}}}, \tag{7}$$

III. SIMPLIFICATION BY SCALE ANALYSIS

Having derived Eq. (5), which describes the motion of air in a time-varying glottis, the next step is to determine the hierarchy of effects, by ranking each term according to its order of magnitude. By comparing these estimates to those for the dominant terms, three results are obtained: (1) the dynamic relevance of the effect represented by each term, (2) the duration over which a given term might be dynamically relevant, and (3) the nondimensional parameters describing the flow and their interpretation.

In this section, the scales of motion are first identified in Sec. III A. In Sec. III B, the technique just described is applied to Eq. (7) in order to demonstrate the approach. Section III C then studies the behavior of viscous blockage from boundary layers, and under what conditions this effect is unsteady. In Secs. III D and III E, the effects of wall motion and movement of the separation point, respectively, are then determined using the same approach. Section III F then uses the results of Secs. III A–III E to develop a simplified equation of motion, and in Sec. III G the implications for the behavior of the glottal jet speed and the glottal volume flow waveform are discussed.

A. Scales of motion

The length scales describing glottal flow are those which describe the volume of the time-varying glottis itself. The length of the volume under consideration is $L \sim x_C - x_A \sim L_{\text{eff}}$, which is approximately 1 cm for adult humans. The glottal volume is approximately this length times the cross-sectional area, $S_S'(t)$, which is approximately $h_S(t)L_G(1 - \delta_S/h_S)$. The length scale describing the maximum glottal width, h_S, is of order 1 mm.

There are several time scales describing glottal flow. The first two characterize the motion of the vocal fold walls: the period of vocal fold vibration, T, and the length of time, T_o, that the glottis is open. Because the interest is the order of magnitude of terms, rather than their exact value, these two quantities will be used interchangeably: $T_o \sim T$. Here, motions of order 100 Hz (speaking voice for adult humans) are considered, so T is approximately 0.01 s. The unsteady acceleration of glottal flow during flow initiation and shutoff has its own time scale, T_a. A second time scale of acceleration, T_{us}, relates the time during which the acceleration of the air in the glottis is dominated by motion of the vocal fold walls. Another relevant time scale, T_c, describes separation point motion. Finally, there is a time scale, T_v, governing the rate of diffusion of the boundary layers. This time scale controls the behavior of viscous blockage.

The scales which characterize the velocities involved in glottal flow are the flow speed at the separation point u_S, which is approximately 30 m/s, and the transverse speed of the vocal fold walls $v_W \approx \Delta h_S/T \sim O(0.1)$ m/s. Here, and in what follows the notation $f \sim O(x)$ means that f has the same order of magnitude as x (see, e.g., Bejan, 1984).

B. Example: Traditional model—wall motion, separation point neglected

Before considering Eq. (5) in full, the "traditional" case, where acceleration due to wall motion and separation point motion are neglected, will be addressed. This will demon-

strate how the current approach applies in this familiar case, in which the only effects considered are the local unsteady flow acceleration, the convective acceleration, the driving transglottal pressure gradient, and friction, which acts to resist the flow.

1. Unsteady acceleration versus convective acceleration

In the traditional case, it is clear that the driving force (transglottal pressure difference) will accelerate the air using two mechanisms, (1) unsteady acceleration, which emphasizes time variations in the flow speed, and (2) convective acceleration, in which the flow is accelerated spatially by squeezing the air through the narrowing glottal inlet. Traditionally, the unsteady acceleration has been neglected. The most concise statement of how this neglect is justified is given by Hirschberg (1992). The magnitude of the two acceleration terms are estimated as follows using scales of motion:

$$\rho\frac{d}{dt}(u_S L_{\text{eff}}) \sim \rho\frac{u_S L}{T}, \quad \rho\frac{u_S^2}{2}\left(1-\left(\frac{S_S'}{S_A}\right)^2\right) \sim \rho u_S^2, \tag{8}$$

where T is the period of vibration and $S_S^2/S_A^2 \sim (10^{-2})$, by assumption. The ratio of the unsteady to the convective acceleration terms thus has order of magnitude St $=(L/u_S T)$, where St is the Strouhal number. Because St is the ratio of the unsteady and the convective acceleration, and because St$\sim O(10^{-2})$, the argument goes, unsteady acceleration may be safely neglected. The soundness of this argument will now be considered.

Note that the estimate of the unsteady acceleration uses the vibration cycle period T as its time scale. But does that time scale really characterize the motion of the air? Measurements of glottal jet speed (see, e.g., Berke *et al.*, 1989; Alipour *et al.*, 1995; Mongeau *et al.*, 1997) suggest that in fact the glottal jet speed u_S changes much more rapidly, especially during flow initiation and shutoff. During initiation, the flow speed is small, but rapidly changing. As a result, the nonlinear convective acceleration terms (proportional to u_S^2) are small compared to the local unsteady acceleration term (proportional to u_S). Once the flow is initiated, however, the rate of change of flow speed decreases, but the flow is still being accelerated spatially (i.e., convectively) through the glottal inlet. This suggests that two temporal regimes exist for the flow: (1) a regime at the beginning and end of the cycle, of duration T_a, in which the unsteady acceleration dominates, and (2) a regime in the middle of the cycle, of approximate duration $T_o - 2T_a$, in which the convective acceleration may dominate. This model has support from Mongeau *et al.* (1997). What fraction of the cycle do each of these regimes occupy? In other words, what is T_a/T?

T_a is the time it takes for the flow to initiate, so it is also a measure of the duration over which the flow has accelerated sufficiently that the two acceleration terms are the same order of magnitude:

$$\rho\frac{d}{dt}(u_S L_{\text{eff}}) \sim \rho\frac{u_S^2}{2}\left(1-\left(\frac{S_S'}{S_A}\right)^2\right). \tag{9}$$

The order of magnitude of the unsteady acceleration term is then estimated as

$$\rho\frac{d}{dt}(u_S L_{\text{eff}}) \sim \rho\frac{u_S L}{T_a} \sim \rho\frac{u_S L}{T}\frac{T}{T_a}. \tag{10}$$

Combining (8) and (9),

$$\rho\frac{u_S L}{T}\frac{T}{T_a} \sim \rho\frac{u_S^2}{2}. \tag{11}$$

Rearranging these, the following relation is obtained:

$$\frac{L}{u_S T}\frac{T}{T_a} \sim \text{St}\,\frac{T}{T_a} \sim O(1). \tag{12}$$

From this result, the fraction of the cycle occupied by the unsteady regime is given by

$$\frac{T_a}{T} \sim \text{St}. \tag{13}$$

Thus the Strouhal number can be interpreted as the fraction of the vibration period that it takes for the flow to accelerate from, or decelerate to, rest. This fraction is a few percent of the vocal fold vibration period for adult speech, but may reach 10% for infants, and an even greater percentage for high-frequency sung notes.

If Eq. (7) is sufficient to describe glottal flow, then in the second, middle temporal regime, it reduces to a quasisteady Bernoulli equation, corrected by wall shear stress,

$$\rho\frac{u_S^2}{2} - \Delta p_{AD} = \frac{\int_{x_A}^{x_S(t)} \tau_W(x)\,dx}{S_{\text{wall}}} = \bar{F}_f, \tag{14}$$

which is quadratic in u_S, so that u_S may be expressed as

$$u_S = \sqrt{\frac{2(\Delta p_{AD} - \bar{F}_f)}{\rho}}, \tag{15}$$

Note that friction acts to oppose the effect of the transglottal pressure difference, reducing the jet speed u_S.

2. Skin friction behavior

Friction is often neglected in glottal flow modeling (see, e.g., Story and Titze, 1995; Pelorson *et al.*, 1994; Lous *et al.*, 1998). The accuracy of this approximation may be found by comparing the relative orders of magnitude of the friction and convection terms. The friction term may be rewritten as

$$\bar{F}_f = \frac{\int_{x_A}^{x_S(t)} \tau_W(x) L_G\,dx}{\int_{S_A}^{S_S(t)} dS_{\text{wall}}} = \frac{\int_{x_A}^{x_S(t)} \frac{1}{2}\rho\,u(x)^2 c_L\sqrt{\nu/u(x)x}\,L_G\,dx}{\int_{S_A}^{S_S(t)} dS_{\text{wall}}}, \tag{16}$$

where (see, e.g., Bejan, 1984) τ_W, for a laminar boundary

layer, is proportional to $\rho u(x_1)^2 c_L \mathrm{Re}^{1/2}(x_1/h_S)^{1/2}$. In Eq. (15), ν is the kinematic viscosity, the ratio μ/ρ of the dynamic viscosity and the fluid density. Here, c_L is in general a factor of $O(1)$ (e.g., 5.0 for a laminar boundary layer on a flat plate in a uniform flow), which depends on the instantaneous glottal shape, and $\mathrm{Re}=u(x_1)h_S/\nu$ is the Reynolds number based on glottal width, h_S. By using the mass conservation equation (neglecting displacement of air by wall motion), the friction term may be further rewritten as

$$\bar{F}_f = \frac{L_G}{S_{\mathrm{wall}}} \int_{x_A}^{x_S(t)} \tau_W(x)\, dx = \frac{L_G}{S_{\mathrm{wall}}} \frac{1}{2} \rho\, u_S^2 h_S^2 c_L \sqrt{\frac{\nu}{u_S h_S}} \int_{x_A}^{x_S(t)} \frac{dx}{\sqrt{h(x,t)^3 x}}. \tag{17}$$

The integral is dominated by its behavior where $h(x,t)$ is minimum, which can be approximated as a $h_S \times h_S$ region, where $h(x,t) \approx h_S$. As a result, the order of magnitude of friction force term is

$$\bar{F}_f \sim \frac{L_G h_S}{S_{\mathrm{wall}}} \rho\, u_S^2 c_L \sqrt{\frac{\nu}{u_S h_S}}. \tag{18}$$

Using $S_{\mathrm{wall}} \sim L_G L$, the order of magnitude of the friction term relative to the convective acceleration is then

$$\frac{\bar{F}_f}{\rho_\infty (u_S^2/2)} \sim \frac{h_S}{L} c_L \sqrt{\frac{\nu}{u_S h_S}}, \tag{19}$$

so that the relative contribution of friction to the dynamics of glottal flow is proportional to the ratio of the glottal aspect ratio h_S/L and the square root of the Reynolds number $\mathrm{Re}=u_S h_S/\nu$. The traditional argument states that the Reynolds number, the ratio of inertia to friction, is sufficient to determine whether friction is negligible. Here, it is seen that the glottal aspect ratio also plays a role.

3. Summary of analysis of traditional case

The preceding analysis shows how, for the traditional case, estimation of the order of magnitude of the terms in the equation of motion may be used to determine not only which terms are dominant, but also the duration of each term's importance. It was shown that two temporal regimes exist, one in which the unsteady acceleration dominates, and another in which the convective acceleration dominates. In addition, the interpretation of nondimensional parameters is clarified. First, the Strouhal number is seen to set the fraction of the vibration period that the unsteady acceleration will dominate. Second, the Reynolds number alone is not sufficient to ascertain the importance of friction.

C. Unsteadiness of viscous blockage

As described above, viscous blockage reduces the effective glottal aerodynamic cross-sectional area S_S', relative to the true cross-sectional area S_S. Mongeau *et al.* (1997) and Zhang *et al.* (2002) have quantified this effect using discharge coefficients, assuming that during phonation, the boundary layers are quasisteady. In this case, S_S' and S_S are in phase. Deverge *et al.* (2003) and Vilain *et al.* (2004) have also studied the behavior of viscous blockage, intuitively deriving a scaling law for the blockage ratio and using it to interpret measurements from a model glottal flow. Neither study provided direct evidence that the boundary layers are in fact quasisteady. In fact, Deverge *et al.* (2003) showed that friction is likely greater during glottal closure, which suggests that the boundary layers are not quasisteady. In this section, a relation for the viscous blockage ratio is derived from the equation of motion. In addition, the possibility that the boundary layers are unsteady (S_S' and S_S are *not* in phase) is considered.

Using

$$\frac{\delta_S}{L} \sim \frac{\delta_S}{h_S}\frac{h_S}{L} \sim c_L \sqrt{\frac{\nu}{u_S h_S}} \sqrt{\frac{h_S}{L}} \tag{20}$$

(see, e.g., Bejan, 1984), where δ_S is the boundary layer thickness, Eq. (19) can be rewritten as:

$$\frac{\bar{F}_f}{\rho(u_S^2/2)} \sim \frac{\delta_S}{L}\sqrt{\frac{h_S}{L}} \sim \frac{\delta_S}{h_S}\left(\frac{h_S}{L}\right)^{3/2}, \tag{21}$$

which shows that the contribution of friction is directly proportional to the blockage ratio. Equating the rightmost term of (21) to the right-hand side of relation (19), the blockage ratio δ_S/h_S is shown to be

$$\frac{\delta_S}{h_S} \sim \frac{c_L}{\sqrt{\mathrm{Re}_{h_S}}} \sqrt{\frac{L}{h_S}}, \tag{22}$$

supporting the arguments of Deverge *et al.* (2003) and Vilain *et al.* (2004) that the controlling parameter for the importance of blockage is $\mathrm{Re}\, h_S/L$, where friction and viscous blockage become most important as $\mathrm{Re}\, h_S/L \to 0$.

Under what conditions is a boundary layer and the resulting viscous blockage unsteady? The temporal response of a boundary to a change in the flow is governed by the time it takes for the change in the core flow to diffuse through the boundary layer. Thus, for a laminar boundary layer, where diffusion is governed by molecular processes, the viscous diffusion time $T_v \sim \delta_S^2/\nu$. The character of the time response of the boundary layer is governed by the ratio of T_v to the period of vibration. If $T_v/T \ll 1$, then changes are communicated through the boundary layer instantaneously, the boundary layer grows very rapidly, and the flow is quasisteady. If $T_v/T \gg 1$, then the boundary layer growth rate is so slow that the boundary layers are thin for the whole vibration cycle and the blockage ratio is negligible. If, however, $T_v/T \sim O(1)$, then the boundary layers grow appreciably during the cycle and may thus be thicker during glottal closure than during opening. In this case, viscous blockage depends not only on glottal area and Reynolds number, but also on the time since the glottis opened. The result is that S_S' experiences a phase lag relative to S_S, a lag of order T_v.

The ratio of diffusion time to glottal period is given by

$$\frac{T_v}{T} \sim \frac{\delta_S^2}{\nu T} \sim \frac{h_S^2}{\nu T}\frac{\delta_S^2}{h_S^2} \sim \mathrm{Wo}^2\frac{\delta_S^2}{h_S^2}, \tag{23}$$

where the Womersley number, Wo (based upon glottal width), is given by $\mathrm{Wo}=(h_S^2/(\nu T))^{1/2}$. In this relation, Wo^2 appears as a proportionality between T_v/T and the blockage ratio (squared). In other words, (1) the higher the blockage ratio, the more likely the boundary layer will be unsteady, and (2) the Womersley number gives the intensity of this effect. Wo in glottal flow is 2.6, using the scales from Sec. III A. Thus, for full blockage ($\delta_S/h_S=0.5$), $T_v/T=1.69$, so it appears that the boundary layers will be unsteady when the blockage ratio is high. When the blockage is small, e.g., $\delta_S/h_S=0.01$, then $T_v/T \sim O(10^{-4})$. The crossover, where $T_v/T=1$, occurs for a blockage ratio of 0.43. During the closure of the glottis, the boundary layers are thicker than during opening, and decreasing glottal area will likely result in an increase in the instantaneous blockage ratio during that interval. As a result, viscous blockage depends explicitly on time, so that S_S' and S_S are out of phase, especially during closure.

D. Acceleration by wall motion

Now the importance of the terms of Eq. (5), which describe the effect of wall motion, is determined, for the interval $T_a<t<T_o-T_a$, in which the convective acceleration dominates relative to the local unsteady acceleration, as described in Sec. III B above. In this section, the terms relating to separation point motion are neglected for simplicity. The equation of motion for glottal flow in this case is given by

$$\rho_\infty\frac{u_S^2}{2} + p_A - p_D + \bar{F}_f = \rho_\infty\frac{\dot{V}_{AS}^2}{S_A^2} + \rho_\infty u_S\frac{2S_S'}{S_A^2}\dot{V}_{AS} - \rho_\infty\frac{d}{dt}\int_{x_A}^{x_S(t)}\frac{\dot{V}_{xS}}{S(x,t)}\,dx \tag{24}$$

The order of magnitude of the left-hand terms has already been discussed in the previous section. In order to estimate the right-hand side terms, the proper scaling for the rate of air volume displacement by wall motion, dV_{AS}/dt, and the integral term need to be determined. The volume V_{xS}, between $x_1=x>x_A$ and $x_1=x_S$, is given by

$$V_{xS} = \int_x^{x_S} S(x_1,t)\,dx_1.$$

The time rate of change of this volume is

$$\dot{V}_{xS} = \int_x^{x_S}\dot{S}(x_1,t)\,dx_1 = 2\int_x^{x_S} v(x_1,t)L_G\,dx_1.$$

For $x_1=x_A$, this quantity has the following order of magnitude:

$$\dot{V}_{AS} \sim v_w L_G L. \tag{25}$$

For the first term on the right-hand side of Eq. (24), its magnitude relative to convection is

$$\frac{\rho(\dot{V}_{AS}^2/S_A^2)}{\rho(u_S^2/2)} \sim \frac{\rho\dot{V}_{AS}^2}{\rho(u_S^2S_S'^2/2)}\left(\frac{S_S'}{S_A}\right)^2 \sim \left(\frac{v_w S_{\mathrm{wall}}}{u_S S_S'}\right)^2\left(\frac{S_S'}{S_A}\right)^2. \tag{26}$$

Before estimating the magnitude of the other two wall motion terms, some attention will be given to interpretation of (26). It shows that the importance of wall motion is governed by the ratio of two volume fluxes, the air volume flux due to wall displacement, $v_W\ S_{\mathrm{wall}}$, and the glottal throughflow volume flux $u_S S_S'$. Because $v_W \sim h_S/T_o$ and $S_S' \sim h_S L_G$, the volume flux ratio may also be written as

$$\frac{v_w S_{\mathrm{wall}}}{u_S S_S'} \sim \frac{v_w L_G L}{u_S h_S L_G} \sim \frac{v_w L}{u_S h_S} \sim \frac{L}{u_S T} \sim \mathrm{St}, \tag{27}$$

supporting an interpretation of Deverge *et al.* (2003) that the Strouhal number indicates the relative strength of volume flux due to wall displacement and the transglottal volume flux. Deverge *et al.* (2003) also noted that the volume flux ratio is likely small when the glottis is wide open, but it should become larger as the glottis closes and the air velocity decreases, resulting in an essentially unsteady flow. In this case,

$$\frac{v_w L_G L}{u_S S_S'} \sim \frac{v_w L}{u_S h_S} \sim O(1), \tag{28}$$

which means that the wall motion is important when

$$\frac{v_w}{u_S} \sim \frac{h_S}{L} \sim \frac{S_S'}{S_{\mathrm{wall}}}. \tag{29}$$

When does this occur? The wall speed is never an appreciable fraction of the flow speed, except for the instants just upon opening and closure. During these intervals the flow has either just been initiated, or is being extinguished, but the walls are moving. However, if the glottis is long and narrow, the area over which displacement by wall motion takes place is larger than the glottal area, so that the volume fluxes may then be commensurate. However, this is rarely the case. Clearly, the shape of the glottis is relevant in determining the importance of wall displacement, as shown by the experiments of Deverge *et al.* (2003).

Having estimated the order of magnitude of the rate of wall volume flux, the remaining wall motion terms in Eq. (24) may be likewise estimated. The magnitude of the right-hand side terms, relative to convection, are given by

$$\frac{\rho(\dot{V}_{AS}^2/S_A^2)}{\rho(u_S^2/2)} \sim \mathrm{St}^2\left(\frac{S_S'}{S_A}\right)^2 \sim O(10^{-6}), \tag{30}$$

$$\frac{\rho u_S(2S_S'/S_A^2)\dot{V}_{AS}}{\rho(u_S^2/2)} \sim \mathrm{St}\left(\frac{S_S'}{S_A}\right)^2 \sim O(10^4), \tag{31}$$

$$\frac{\rho\dfrac{d}{dt}\displaystyle\int_{x_A}^{x_S(t)}[\dot{V}_{xS}/S(x,t)]dx}{\rho u_S^2/2} \sim \frac{L^2}{u_S^2T^2} \sim \mathrm{St}^2 \sim O(10^{-4}). \tag{32}$$

Over what time scale are these terms important? The time interval T_{us} during which the wall volume displacement

and glottal volume displacement are commensurate may be estimated by using $v_w \sim \Delta h_S / T_{us}$, where Δh_S is the change in glottal width during the time T_{us} during which wall motion is relevant. Then,

$$\frac{v_w L}{u_S h_S} \sim O(1) \sim \frac{\Delta h_S L}{u_S h_S T_{us}} \sim \frac{\Delta h_S L}{u_S h_S T} \frac{T}{T_{us}} \sim \mathrm{St} \frac{T}{T_{us}} \frac{\Delta h_S}{h_S}. \tag{33}$$

T_{us} has the fraction of a vibration period:

$$\frac{T_{us}}{T} \sim \mathrm{St} \frac{\Delta h_S}{h_S} \sim \frac{T_a}{T} \frac{\Delta h_S}{h_S}, \tag{34}$$

where (13) has been used. Thus the time interval during which wall displacement is important is of the same magnitude as the interval T_a (see Sec. III B 1, above) the flow takes to accelerate from, or decelerate to, zero. This suggests that glottal airflow acceleration by wall motion is negligible except during the initial and final few percent of the period the glottis is open, and is thus confined to that temporal regime.

E. Importance of separation point motion

Equation (5), after neglecting wall motion terms, may be rewritten in the following manner:

$$\rho \frac{u_S^2}{2} - \Delta p_{AD} + \bar{F}_f = -\rho_\infty \frac{d}{dt}(\dot{x}_S L_{\mathrm{eff}}) + \rho_\infty u_S \dot{x}_S + \rho_\infty u_S \frac{2 S_S'^2}{S_A^2} \dot{x}_S + 2 \frac{\dot{V}_{AS}}{S_A^2} S_S' \dot{x}_S - \frac{S_S'^2}{S_A^2} \dot{x}_S^2, \tag{35}$$

where the terms under consideration in this section have been moved to the right-hand side. If the separation point is assumed to move the length of the glottis in a time T_c, the separation point speed $dx_S/dt \sim L/T_c \sim (L/T)(T/T_c)$.

The magnitude of these separation point motion terms, relative to the convective acceleration term, which is of order ρu_S^2, are thus given as follows:

$$\frac{\rho \frac{d}{dt}(\dot{x}_S L_{\mathrm{eff}})}{\rho_\infty (u_S^2/2)} \sim \rho \frac{L^2}{u_S^2 T^2} \frac{T^2}{T_c^2} \sim \mathrm{St}^2 \frac{T^2}{T_c^2} \sim O(10^{-4}) \frac{T^2}{T_c^2}, \tag{36}$$

$$\frac{\rho u_S \dot{x}_S}{\rho (u_S^2/2)} \sim \frac{L}{u_S T} \frac{T}{T_c} \sim \mathrm{St} \frac{T}{T_c} \sim O(10^{-2}) \frac{T}{T_c}, \tag{37}$$

$$\frac{\rho u_S (2 S_S'^2/S_A^2) \dot{x}_S}{\rho (u_S^2/2)} \sim \frac{L}{u_S T} \frac{T}{T_c} \frac{S_S'^2}{S_A^2} \sim \mathrm{St} \frac{T}{T_c} \frac{S_S'^2}{S_A^2} \sim O(10^{-6}) \frac{T}{T_c}, \tag{38}$$

$$\frac{2\rho (\dot{V}_{AS}/S_A^2) S_S' \dot{x}_S}{\rho_\infty (u_S^2/2)} \sim \frac{v_W L_G L}{u_S S_S'} \frac{L}{u_S T} \frac{T}{T_c} \frac{S_S'^2}{S_A^2} \sim \mathrm{St}^2 \frac{T}{T_c} \frac{S_S'^2}{S_A^2} \sim O(10^{-6}) \frac{T}{T_c}, \tag{39}$$

$$\frac{\rho (S_S'^2/S_A^2) \dot{x}_S^2}{\rho (u_S^2/2)} \sim \frac{L^2}{u_S^2 T^2} \frac{T^2}{T_c^2} \frac{S_S'^2}{S_A^2} \sim \mathrm{St}^2 \frac{T^2}{T_c^2} \frac{S_S^2}{S_A^2} \sim O(10^{-6}) \frac{T^2}{T_c^2}. \tag{40}$$

The first two terms are the largest relative to the convection term, the rest being of at least four orders of magnitude smaller. The importance of these two largest terms depends critically on the ratio T/T_c. If $T/T_c \sim O(10^2)$, for example, then these two terms have the same importance as convection. On the other hand, if $T/T_c \sim O(10)$, as suggested by the computations of Alipour and Scherer (2004), and the modeling of Pelorson *et al.* (1994), then only the second term is relevant and represents a 10% correction to convection, when the separation point is in motion.

Thus, under the assumption that $T/T_c \sim O(10)$, all the separation point terms but the second are negligible, and Eq. (32) reduces to

$$\rho \frac{u_S^2}{2} - \rho \dot{x}_S u_S - \Delta p_{AD} + \bar{F}_f = 0. \tag{41}$$

Note that the surviving separation point motion term comes directly from the unsteady acceleration term [see Eq. (2)]. This result strongly supports the arguments of Sobey (1983) and McGowan (1993) that a moving separation point results in an inherently unsteady flow, as an approximately 10% correction to the convective acceleration.

F. Behavior of jet speed and glottal volume flow

The analysis of Secs. III B–III E has identified two temporal regimes, one in which unsteady effects such as local unsteady acceleration and wall motion are relevant, and one in which the only relevant unsteady mechanism is due to separation point motion. This article focuses on this second regime, in which Eq. (5) reduces to

$$\rho \frac{u_S^2}{2} - \rho \dot{x}_S u_S - G = 0, \tag{42}$$

where

$$G = \Delta p_{AD} - \bar{F}_f. \tag{43}$$

This equation is a quadratic in u_S, so

$$u_S = \dot{x}_S + \sqrt{\dot{x}_S^2 + \frac{2G}{\rho}}. \tag{44}$$

Given that the first term under the square root sign is $O(L^2/T_c^2)$ and the second term is $O(u_S^2)$, their ratio is proportional to $\mathrm{St}^2 T^2/T_c^2 \sim O(10^{-2})$ and the approximate expression for the air speed at the separation point is

$$u_S = \dot{x}_S + \sqrt{\frac{2G}{\rho}}. \tag{45}$$

From this equation the dominant effects determining flow through the glottis can be determined:

(1) The jet speed u_S is determined primarily by the instantaneous net force $G = \Delta p_{AD} - F_f$. The two terms in

G vary differently with time, but the time variations in u_S are in phase with the time variations in G.

(2) Friction retards the flow, particularly toward the end of the cycle, as the flow is "squeezed" through the growing boundary layers by the closing vocal fold walls. As a result, friction increases with time during glottal closure, reducing the net driving force G at the end of the cycle. This behavior is also indicated by the experiments of Deverge *et al.* (2003).

(3) The speed of separation point motion appears directly as a correction to the speed of glottal air flow. When the separation point moves downstream (is positive), the jet speed is enhanced, while upstream separation point motion retards jet speed. If the separation point moves upstream during closure, the jet speed will decrease faster than a Bernoulli prediction.

The glottal volume velocity $U_g = u_S S_S'$ is given by the following expression:

$$U_g \approx S_S'(t)\dot{x}_S + S_S'(t)\sqrt{2\frac{G}{\rho_\infty}}. \tag{46}$$

From this expression, it can be seen that glottal volume flow waveform is affected by temporal fluctuations in three mechanisms: (1) G, (2) S_S', through variations in both glottal geometry and the blockage ratio δ_S/h_S, and (3) movement of the separation point x_S, where the effect of separation point movement is seen to be a correction of $O(10^{-1})$. For example, the effect of the time variation of Δp_{AD} (due to vocal tract inertance) is that the volume velocity waveform is skewed such that the flow shutoff occurs more rapidly than the flow initiation. This effect is well known (see, e,g., Titze, 1994) to strongly affect the sound power and quality of the voice. From Eq. (46), it is clear that other mechanisms can further enhance the rapidity of the volume flow shutoff. First, an upstream movement of the separation point causes a deceleration of U_g, both because dx_S/dt is negative and because S_S' decreases with time in a diverging glottis. (A converse effect occurs if the separation point moves downstream.) Furthermore, the effective flow area S_S' is a strong function of the blockage ratio δ_S/h_S, which is expected to increase during the vibration cycle because h_S decreases much faster than δ_S during closure. This results in a faster decrease of the effective flow area S_S' than that of the true cross-sectional area S_S. Again, the validity of these statements does not necessarily apply to the final deceleration of the flow during the final few percent of the interval where the glottis is open, during which time the wall motion terms become important.

G. Discussion

The present analysis explains why *in vitro* experiments in which the vocal folds are mimicked by a time-varying diaphragm or shutter (see, e.g., Coker *et al.*, 1996; Mongeau *et al.*, 1997; Barney *et al.*, 1999; Zhang *et al.*, 2002) produce convincing replications of glottal volume flow waveforms even though the glottis in these experiments does not change shape (from a converging to a diverging shape) during a glottal cycle. Because the volume flow is determined largely by the transglottal pressure gradient and the glottal flow area $S_S'(t)$, the vibration pattern is almost irrelevant, as long as $S_S'(t)$ has similar behavior to the real flow. Because separation point motion is due more to changing glottal shape than other factors, the behavior of the separation point [and hence $S_S'(t)$] in these experimental approximations *will* differ from that found in real vocal fold vibration. In the experiments mentioned, the flow separation most likely (1) occurs near the "glottal" exit and (2) does not move very much.

The results presented here rely on the validity of the underlying assumptions laid out in Sec. I, which are now discussed. The first assumption is that the glottal jet is essentially quasisteady, resulting in a uniform pressure downstream of the separation point (i.e., that $p_S \approx p_D$). Experimental evidence (Hofmans *et al.*, 2003) strongly suggests that this assumption is not correct. In fact, it seems likely that, because (1) the volume occupied by the jet and (2) the momentum distribution inside that volume both exhibit strong time variations (including turbulent fluctuations), $p_S - p_D$ is a nonzero function of time. Thus, the appearance of a "quasisteady" regime, identified using the theoretical approach used here, rests on the traditional assumption that the jet development is itself quasisteady. Recent studies (Mongeau *et al.*, 1997; Vilain *et al.*, 2004), which apparently confirm this approximation for glottal flow as a whole, did so using transglottal measurements across a model glottis in which the supraglottal pressure tap was either vented to an open room (Mongeau *et al.*, 1997; Deverge *et al.*, 2003; Vilain *et al.*, 2004) or placed far downstream of the glottal exit. This experimental arrangement essentially *enforces* the *assumed* condition that the pressure downstream of separation is uniform or, equivalently, that the jet is quasisteady. The degree to which the glottal jet is unsteady, and how it might thereby contribute to glottal flow unsteadiness, is currently under study by the authors.

Another assumption used is that the vocal fold motion and the flow are symmetric about the centerline. In a symmetrical glottis, the flow will not exhibit asymmetrical behavior upstream of the separation point. In other words, flow in a symmetrical glottis can only be asymmetrical in the jet region. The experiments of Hofmans *et al.* (2003) show that the glottal jet may become asymmetrical, but not the flow upstream of the separation point. The assumption of uniform pressure in the jet region is equivalent to saying that no matter what form glottal jet motion takes, its dynamic importance is limited to an instantaneous and total diffusion of jet momentum. In other words, the dynamic effect of flow asymmetry is negligible, under the assumption that the jet is quasisteady. For cases of mild vocal fold asymmetry, at least while the motion of the folds is synchronized, the same is likely true. Once vocal fold motion becomes desynchronized, vocal fold and flow asymmetry become severe, so the validity of the current analysis is likely compromised.

It is clear, then, that most unsteady effects can be neglected only if the glottal jet is quasisteady and the glottis is symmetrical. This strongly motivates a thorough investigation of the behavior of the glottal jet and the effect of vocal

fold asymmetry on the unsteadiness of glottal flow.

IV. SUMMARY

This paper has presented a theoretical evaluation of the common assumption that glottal flow essentially involves a balance between transglottal pressure gradient and convective acceleration, i.e., that all other effects (separation point motion, wall motion, viscous effects such as friction and viscous blockage) are negligible. An equation of motion was derived in a form which explicitly includes the effects of vocal fold wall motion, viscous blockage, and motion of the separation point. Using the scales of motion describing phonation, the order of magnitude of the terms describing each effect was estimated and compared to that of the dominant convective acceleration term.

Two temporal regimes were identified. The first is dominated by unsteady effects, the local acceleration and acceleration by the vocal fold walls. The Strouhal number, which is typically $O(10^{-2})$, indicates the magnitude of this interval relative to the vibration period. The second temporal regime was the major focus of the paper: the interval in which the flow appears to be essentially quasisteady, with second-order corrections due to the motion of the separation point and unsteady diffusion of the boundary layers.

Expressions for the jet speed and volume flow rate were obtained from the simplified equation of motion. These show that the glottal volume flow waveform is determined primarily by the transglottal pressure gradient and the cross-sectional area at the separation point, but is also influenced by temporal fluctuations in three other factors: (1) the movement of the separation point, (2) the viscous blockage, and (3) friction, all of which are not likely to vary in phase with either the transglottal pressure or the glottal area. Thus these effects also contribute to the asymmetry of the volume flow waveform.

The importance of friction, relative to convection, was shown to be controlled by the quotient of the glottis aspect ratio and the square root of the Reynolds number, a result equivalent to the assertion of Deverge *et al.* (2003) that the blockage ratio is determined by the reciprocal of the square root of the product of Reynolds number and the aspect ratio. The Womersley number was also shown to reflect how strongly the blockage ratio affects the transverse diffusion time and the degree to which the true glottal area S_S and the effective glottal area S'_S are in phase. For typical normal phonation, an increase of the blockage ratio and a phase shift between S_S and S'_S are expected during glottal closure.

Finally, the above conclusions are only as valid as the assumptions upon which they are based. The assumptions that (1) glottal jet behavior is quasisteady and (2) that the pressure downstream of the separation point is uniform are currently under study.

ACKNOWLEDGMENT

The authors acknowledge partial support from NIH Grant No. 1R01DC054642-01.

NOMENCLATURE

f = frequency
$\bar{F}_f$ = friction force on air in glottis
G = net force on glottal flow (transglottal pressure difference minus friction force)
$h(x,t)$ = glottal half-width
h_A, h_S = glottal half-width at x_A and x_S, respectively
L = axial length of vocal folds
L_d = boundary layer development length scale
L_{eff} = effective axial length of glottis
L_G = vocal fold length
n_i = ith component of unit normal vector on vocal fold wall
p = air static pressure
Re = Reynolds number, Uh_S/ν, based on glottal width h_S
S_A, S_S = physical cross sectional areas at x_A and x_S, respectively
S'_S = aerodynamic cross sectional area at x_S, corrected for viscous blockage
S_{wall} = area of vocal fold walls
St = Strouhal number $U/(LT)$
t = time
T = period of vocal fold vibration
T_o, T_a, T_{us}, T_v, T_c = time scales describing: interval glottis is open, glottal flow acceleration/deceleration, interval time air displacement by vocal fold wall motion is important, viscous diffusion across boundary layer, and motion of separation point, respectively
u = axial component of velocity on centerline
u_i = component of air velocity vector in ith coordinate direction
U_g = glottal volume flow rate
v_W = velocity of vocal fold walls
V_{AB} = glottal volume between axial points x_A and x_B
$\dot{V}_{AB}$ = Time rate of change of V_{AB}
Wo = Womersley number $[h^2/(T\nu)]^{1/2}$
x_i = ith component of spatial coordinate $i=1$: axial, $i=2$, transverse
Δp_{AD} = pressure difference between x_A and x_D (transglottal pressure difference)
Δh_S = change in h_S during vocal fold vibration
δ_S = boundary layer thickness in glottis, at separation point
μ = air dynamic viscosity
ν = air kinematic viscosity, μ/ρ
ρ = air density

APPENDIX: MASS CONSERVATION EQUATION IN TIME-VARYING VOLUME

In this appendix the conservation of mass equation for the incompressible flow through a control volume V_{AB} bounded by a control surface S consisting of a time-varying inlet and outlet, as well as impermeable, moving walls, is

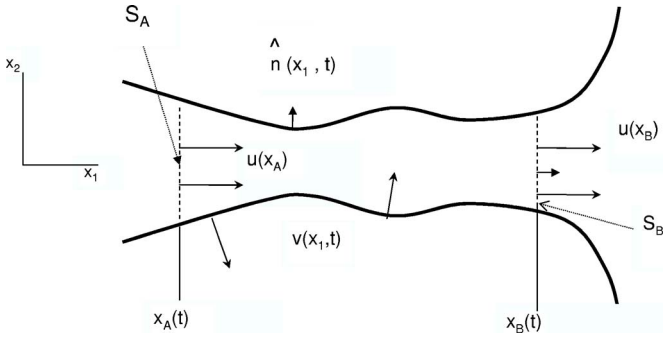


FIG. 2. Sketch of control volume for flow through a duct of time-varying shape, and whose entrance and exit locations (x_A and x_B, respectively) change with time. The unit outward normal vector is **n**, and the wall velocity is $\mathbf{v}(x_1,t)$.

presented. From this equation, the expression for the flow speed at any location x_1 inside V_{AB} is derived. This relation is used to derive Eq. (5).

The volume is shown in Fig. 2. The inlet (area S_A) is located at $x_1=x_A(t)$, the outlet (area S_B) at $x_2=x_B(t)$. The wall orientation is denoted by the unit outward normal vector, n, whose components in the x_1, x_2, and x_3 directions are n_1, n_2, and n_3, respectively. The conservation of mass equation for this case is given by (Panton, 1994)

$$\int_{S(t)} (v_i - u_i) n_i \, dS = -\frac{dV_{AB}}{dt}, \tag{A1}$$

where v_i is the ith component of the velocity of the wall. Because of the no-slip and no-flux boundary conditions, $v_i - u_i = 0$ on the solid walls, the only contributions from the surface integral come from the inlet and outlet, so that

$$(u_B - \dot{x}_B) S'_B - (u_A - \dot{x}_A) S'_A = -\frac{dV_{AB}}{dt}, \tag{A2}$$

where $u_A S'_A$ and $u_B S'_B$ are the inlet and outlet volume fluxes (volume velocities), respectively, averaged over their respective cross-sectional areas. In addition, S'_A and S'_B are the physical cross-sectional areas, corrected for viscous blockage. The other terms represent the change in V_{AB} by the motion of the inlet and outlet surfaces, at speeds dx_A/dt and dx_B/dt, respectively.

In the derivation of Eq. (6), an expression is needed for $u(x_1)$, i.e., u at a x_1 location inside V_{AS}, For a control volume bounded by $x_1=x$ on the right, and $x_1=x_B(t)$ on the left, then

$$u(x) = (u_B - \dot{x}_B)\frac{S_B}{S(x)} + \frac{1}{S(x)}\frac{dV_{AB}}{dt}. \tag{A3}$$

Alipour, F., and Scherer, R. (**2004**). "Flow separation in a computational oscillating vocal fold model," J. Acoust. Soc. Am. **116**, 1710–1719.

Alipour, F., Scherer, R., and Patel, V. (**1995**). "An experimental study of pulsatile flow in canine larynges," ASME J. Fluids Eng. **117**, 577–581.

Barney, A., Shadle, C. H., and Davis, P. O. A. L. (**1999**). "Fluid flow in a dynamic mechanical model of the vocal folds and tract. I. Measurements and theory," J. Acoust. Soc. Am. **105**, 444–455.

Bejan, A. (**1984**). *Convection Heat Transfer* (Wiley-Interseience, New York).

Berke, G., Moore, D., Monkewitz, P., Hanson, D., and Gerrat, B. (**1989**). "A preliminary study of particle velocity during phonation in an in vivo canine model," J. Voice **3**(4), 306–313.

Coker, C. H., Krane, M. H., Reis, B. Y., and Kubli, R. A. (**1996**). "Search for unexplored effects in speech production," *Proceeding of ICSLP '96*, Philadelphia, PA.

Deverge, M., Pelorson, X., Vilain, C., Lagrée, P.-Y., Chentouf, F., Willems, J., and Hirschberg, A. (**2003**). "Influence of collision on the flow through *in vitro* rigid models of the vocal folds," J. Acoust. Soc. Am. **114**, 3354–3362.

Flanagan, J. L., and Ishizaka, K. (**1976**). "Computer model to characterize the air volume displaced by the vibrating vocal cords," J. Acoust. Soc. Am. **63**, 1559–1565.

Hirschberg, A. (**1992**). "Some fluid dynamic aspects of speech," Bull. Comm. Parl. **2**, 1–30.

Hofmans, G., Groot, G., Rancci, M., Graziani, G., and Hirschberg, A. (**2003**). "Unsteady flow through in-vitro models of the glottis," J. Acoust. Soc. Am. **113**(3), 1658–1675.

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–l268.

Liljencrants, J. (**1996**). "*Experiments with analysis by synthesis of glottal airflow*," *Proceedings of ICSLP '96*, Philadelphia, PA.

Lous, N. J. C., Hofmans, G. C. J., Velduis, R. N. J., and Hirschberg, A. (**1998**). "A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design," Acta Acust. (Beijing) **84**(6), 1135–1150.

McGowan, R. S. (**1993**). "The quasi-steady approximation, boundary-layer separation, and vibrations in the vocal tract," J. Acoust. Soc. Am. **93**, 2295.

Mongeau, L., Franchek, N., Coker, C. H., and Kubli, R. A. (**1997**). "Characteristics of a pulsating jet through a small, modulated orifice, with application to voice production," J. Acoust. Soc. Am. **102**, 1121–1133.

Panton, R. (**1994**). *Incompressible Flow* (Wiley, New York).

Peloroson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (**1994**). "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.

Rosenhead, L. (Ed.) (**1963**). *Laminar Boundary Layers*, (Dover, New York).

Sobey, I. J. (**1983**). "The occurrence of separation in oscillatory flow," J. Fluid Mech. **134**, 247–257.

Story, B., and Titze, I. (**1995**). "Voice simulation with a body-cover model of the vocal folds," J. Acoust. Soc. Am. **97**, 1249–1260.

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall Englewood Cliffs, NJ).

Vilain, C. E., Perorson, X., Fraysse, C., Deverge, M., Hirschberg, A., and Willems, J. (**2004**). "Experimental validation of a quasi-steady theory for the flow through the glottis," J. Sound Vib. **276**, 475–490.

Zhang, Z., Mongeau, L., and Frankel, S. (**2002**). "Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes," J. Acoust. Soc. Am. **112**, 1652–1663.

How far, how long: On the temporal scope of prosodic boundary effects[a)]

Dani Byrd,[b)] Jelena Krivokapić, and Sungbok Lee[c)]
USC Department of Linguistics, University of Southern California, 3601 Water Way, GFS 301, Los Angeles, California 90089-1693

(Received 16 February 2005; revised 30 May 2006; accepted 1 June 2006)

Acoustic lengthening at prosodic boundaries is well explored, and the articulatory bases for this lengthening are becoming better understood. However, the temporal scope of prosodic boundary effects has not been examined in the articulatory domain. The few acoustic studies examining the distribution of lengthening indicate that boundary effects extend from one to three syllables before the boundary, and that effects diminish as distance from the boundary increases. This diminishment is consistent with the π-gesture model of prosodic influence [Byrd and Saltzman, J. Phonetics **31**, 149–180 (2003)]. The present experiment tests the preboundary and postboundary scope of articulatory lengthening at an intonational phrase boundary. Movement-tracking data are used to evaluate durations of consonant closing and opening movements, acceleration durations, and consonant spatial magnitude. Results indicate that prosodic boundary effects exist locally near the phrase boundary in both directions, diminishing in magnitude more remotely for those subjects who exhibit extended effects. Small postboundary effects that are compensatory in direction are also observed. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2217135]

PACS number(s): 43.70.Bk [AL] Pages: 1589–1599

I. INTRODUCTION

A. Articulatory studies on phrase boundary effects

Prosodic structure shapes the production of individual phonological units at phrase boundaries and under accent. In the vicinity of prosodic boundaries, segments exhibit acoustic final lengthening (e.g. Oller, 1973; Klatt, 1976; Wightman *et al.*, 1992) and initial lengthening (Oller, 1973). Articulatory studies show that at phrase edges "gestures get larger, longer, and further apart" (Byrd and Saltzman, 2003, p. 159; Byrd *et al.*, 2000). Articulations are spatially more extreme and temporally longer (Edwards *et al.*, 1991; Beckman and Edwards, 1992; Fougeron and Keating, 1997; Byrd and Saltzman, 1998; Fougeron, 2001; Cho, 2006; Cho and Jun, 2000; Cho and Keating, 2001; Tabain, 2003; Keating *et al.*, 2004; Tabain and Perrier, 2005).

Phrase final and initial articulatory lengthening have been observed to increase cumulatively for larger prosodic boundaries (phrase finally: Byrd and Saltzman, 1998; Byrd, 2000; Cho, 2006; Tabain, 2003; Tabain and Perrier, 2005; and phrase initially: Byrd and Saltzman, 1998; Cho and Keating, 2001; Fougeron, 2001; Cho, 2006; Tabain, 2003; Keating *et al.*, 2004), though there is individual variation. Articulatory studies have also shown there to be less temporal overlap between articulations separated by a boundary or adjacent to a boundary (McClean, 1973; Byrd *et al.*, 2000; Byrd and Saltzman, 1998; Byrd, 2000; see also Hacopian, 2003). This depends on boundary strength; gestures are less overlapped across stronger boundaries (Byrd, 2000; Cho, 2004).

In the spatial domain, Fougeron and Keating (1997) and others show the effects of increased linguapalatal contact to be cumulative—higher (stronger) prosodic boundaries induce greater linguapalatal contact, both phrase initially (Fougeron and Keating, 1997; Cho and Keating, 2001; Fougeron, 2001; Keating *et al.*, 1999; Tabain, 2003; Keating *et al.*, 2004) and phrase finally (Fougeron and Keating, 1997 [although their results are less consistent across speakers than their results for phrase-initial effects], Keating *et al.*, 1999; see also Hacopian, 2003; Tabain, 2003). Both phrase-final and phrase-initial magnitude results tend to vary across speakers and across segments examined (see, e.g., Byrd *et al.*, 2005), and different studies find spatial effects at different prosodic domains.

B. Acoustic and articulatory studies on the temporal scope of phrase boundary effects

While the articulatory and acoustic effects of boundaries have been examined in a number of studies, the *temporal scope* of boundary effects—i.e., the interval preceding and following a boundary over which they occur—has been far less studied. A few studies have addressed this question in the domain of acoustics. Oller (1973) examined onset and coda lengthening and found that generally both lengthen in final syllables. Berkovits (1993a, 1993b) showed that in phrase-final bisyllabic words with stress on the second (final) syllable, lengthening of segments within the final syllable is largest in the final segment and decreases in the preceding segments. Furthermore, phrase-final lengthening extends to

[a)]Portions of this work were presented in "On the temporal scope of boundary effects in articulation" Acoustical Society of America meeting, San Diego, California, November 2004.

[b)]Author to whom correspondence should be addressed; electronic mail: dbyrd@usc.edu

[c)]Also at: University of Southern California, Viterbi School of Engineering, Los Angeles, CA 90089.

the initial syllable (of bisyllabic words) but is lesser in magnitude than in the final syllable. Berkovits (1994) examined lengthening of phrase-final bisyllabic words with initial and final stress and found that words with initial stress show lengthening on both the initial and the final syllable and again, that the lengthening in the segments of the final syllable increases progressively closer towards the boundary. Words with final stress in this experiment show lengthening only on the final syllable, and again, segmental lengthening increases progressively towards the boundary. Overall, the initial syllable of two-syllable words lengthens less than the final syllable, and lengthening is mainly on the final syllable (Berkovits, 1993b, 1994).

Shattuck-Hufnagel and Turk (1998) examined lengthening effects, and they showed preboundary lengthening beyond the final syllable, back to the stressed syllable of a word. Lengthening is strongest in the final syllable. The postboundary effects show more lengthening on the onset than the rime vowel, which does not appear to lengthen. Turk (1999) reported that two-syllable words show phrase final lengthening back to the rime of the stressed syllable. Significant lengthening mainly occurs on the rimes, not on the onsets. (In the case of words with stress on the second syllable, one of two subjects showed a significant *shortening* effect on the rime of the first syllable.) Lengthening in the final syllable increases from nucleus to coda.

Cambier-Langeveld (1997), in her acoustic study on Dutch, examined the preboundary extent of the effect of prosodic boundaries. She found that the lengthening is largest in the final segment, and decreases in the preceding ones. Generally, the lengthening domain is the final syllable, but significant lengthening on the nucleus of the penultimate syllable occurred in one token word (*mode*), whose last syllable consists of an onset and a *reduced* vowel. (A trend for lengthening of the penultimate nucleus also occurred in the word *tandem*, in which the vowel of the final syllable is also reduced.) As Cambier-Langeveld noted, the effect of syllable weight on the domain of final lengthening could be explained by the fact that these syllables are phonetically shorter than the syllables with an unreduced vowel. Finally, Cambier-Langeveld pointed out that even though the amount of lengthening increases with higher prosodic boundaries, the domain of lengthening does not become larger. This result is similar to Berkovits's findings, who reported that, in words with initial stress, the first syllable accounts for 25% of the lengthening, the final syllable for 75%. In words with final stress, 5% of the average lengthening was due to the initial syllable and 95% was due to the lengthening of the final syllable (Berkovits 1993a, 1994). The figures in Turk (1999) point to a similar distribution of lengthening.

In the articulatory domain, there are only a few studies that tangentially examine the temporal extent or scope of phrasal lengthening. Byrd and Saltzman (1998) compared opening movements for C_1 and closing movements for C_2 in $[C_1V_1\#C_2V_2]$ sequences and show that the postboundary temporal effect is stronger. Byrd (2000) reported that, in $[C_1V_1\#C_2V_2]$ sequences, V_1 exhibits longer articulatory duration than V_2. A further indication of progressive final lengthening is evident in the results of Edwards *et al.* (1991) and Beckman and Edwards (1992). They examined jaw movement in a VC sequence phrase finally and found that there is an increase in duration of the jaw opening and jaw closing movements, and that the effects are more prominent in the final jaw closing. Fougeron and Keating (1997) examined palatal contact in $[n_1o_1\#n_2o_2]$ sequences. They found for the consonants that there is a boundary effect not just domain initially (n_2), but also in some cases domain finally (n_1); however, the effect is much stronger closer to the boundary (that is, for n_2). Similarly for the vowel, the boundary effect was much stronger for the vowel closer to the boundary (o_1) than for the vowel farther away from the boundary (o_2). Their results can be taken to indicate a progressively decreasing effect of the phrase boundary (see also, Gendrot, 2005).

Together, the acoustic and articulatory studies point to decreased lengthening with distance from a phrasal boundary. Absolute phase final segments lengthen more than preceding ones. Furthermore, even though the degree of lengthening is a function of boundary type, the temporal locale of lengthening seems often to be largely, though not entirely, limited to the boundary-adjacent segments. Articulatory studies addressing the temporal scope of the boundary effects are sparse.

C. The π-gesture framework

Findings as to the temporal scope of phrasal effects on articulations have theoretical implications for the prosodic gesture (π-gesture) framework of Byrd and Saltzman (Byrd and Saltzman, 2003; Byrd *et al.*, 2000; Byrd, 2000). The π-gesture model extends the notion of phonological units being defined as intrinsically temporal (e.g., Browman and Goldstein, 1992), by viewing prosodic events such as phrase boundaries as having a temporal interval of activation, similar to constriction gestures. Many of the phrasal effects of longer duration, greater magnitudes, and decreased intergestural overlap have been successfully captured by simulations of boundaries, implemented as π-gestures, interacting with articulatory gestures within this model (Byrd and Saltzman, 2003). In this simulation work, a π-gesture on the prosodic tier influences all constriction gestures that are concurrently active with it. This overlap among suprasegmental and constriction gestures is shown in Fig. 1.

During the activation of a phrasal π-gesture, the central clock controlling the rate of activation-unfolding for active gestures is slowed (Byrd and Saltzman, 2003). The amount of slowing down depends on the strength of activation of the π-gesture, which is a reflection of the strength of the prosodic boundary. Because the π-gesture has a temporally discrete and smoothly varying interval of activation, the π-gesture model predicts that boundary effects should be local, that is, not discontinuous (Byrd and Saltzman, 2003). Further, because the π-gesture waxes and wanes in activation, the lengthening associated with a π-gesture is predicted to likewise increase in degree as the π-gesture's peak activation is reached and then decrease (Byrd and Saltzman, 2003).[1]

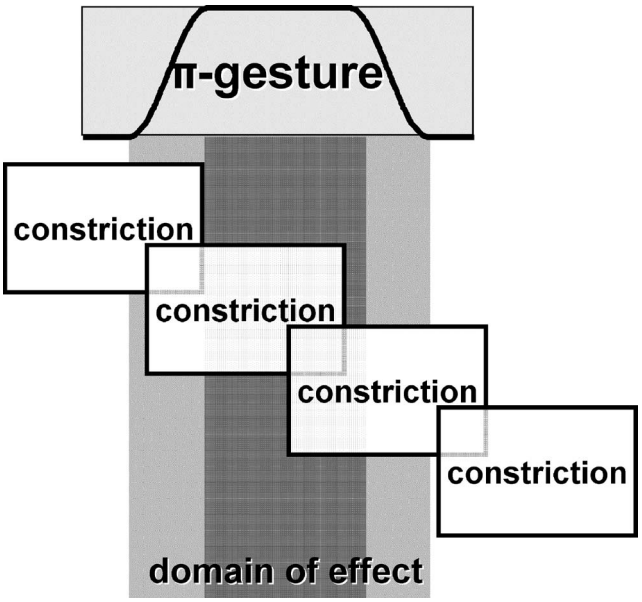


FIG. 1. A schema of the π-gesture model representing the overlap of a π-gesture at a phrasal juncture with several closing movement gestures. Shading darkness represents strength of the π-gesture effect as the π-gesture's activation increases to maximum (darkest shading). The constriction gestures are representative of *any* consonant or vowel. In our experiment, the constrictions potentially represent the consonants N D D N.

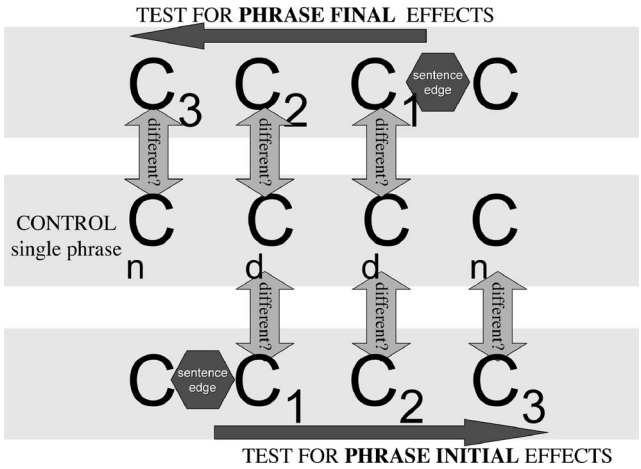


FIG. 2. Schematized experimental design.

D. Goals of the present study

The primary goal of this study is an empirical one—that is, to investigate the temporal scope of prosodic lengthening in the articulatory domain, since such articulatory data are not currently available in the literature. This articulatory kinematic experiment explicitly investigates the preboundary and postboundary extent of phrasal effects in the articulatory domain. Based on the results from the acoustic studies, we expect that the strongest articulatory lengthening will appear near the phrase edge but that lengthening could also be possible more remotely. We also seek to add to the description of the general nature of prosodically-driven temporal perturbation to articulation. We report spatial magnitude effects as well, primarily for the purpose of determining if they mirror the temporal effects observed for the vertical tongue tip movement, i.e., the tongue tip *y* trajectories (and, of course, for the purpose of adding to the limited data in the kinematic literature on phrasal spatial effects). We do not evaluate these spatial results in a theoretical context here.

We also pursue the additional secondary goal of evaluating whether the pattern of results is consistent with the π-gesture model of Byrd and Saltzman (2003). Because a π-gesture has a temporally discrete interval of activation with a smoothly waxing and waning function, it would be expected that boundary effects should be local to the boundary and not discontinuous (Byrd and Saltzman, 2003). Further, because the π-gesture waxes and wanes in activation, the degree of lengthening associated with a π-gesture is likewise expected to increase (as the π-gesture's peak activation is reached) and then decrease (Byrd and Saltzman, 2003). Lastly, prosodically driven lengthening is expected to be generally symmetrical [this is not a requirement of the model, merely the null hypothesis adopted in Byrd and Saltzman (2003)].

II. METHOD

A. Stimuli and subjects

Three stimuli, given in Table I, were used to test phrase boundary effects on preboundary and postboundary sequences of consonants, with the same phonological string appearing in each. The target sequence in each utterance was […nVdVdVnV…]. The articulation of the tongue tip consonants will be investigated using articulatory movement tracking of a transducer placed on the tongue tip. For ease of presentation, we will denote the target sequence as [N D D N].

In order to avoid any confounding effect of accent at the phrase boundary, subjects were asked to emphasize (place in

TABLE I. Stimulus 1 is testing for preboundary effects; the consonants to be measured are N D D. Stimulus 2 is testing for postboundary effects; the consonants to be measured are D D N. Stimulus 3 is the control utterance and contains a phrase-medial sequence of consonants. (Subjects were instructed to emphasize the underlined word to control somewhat for focal accent placement.)

Effect	Consonants	Sentence
Preboundary effects	N D D	Birdhunting, we were shocked to see a **n**ew **d**o**d**o. Knock on wood that there are more.
Postboundary effects	D D N	At the zoo, we were shocked to see a Gnu. **D**o**d**o **kn**ocking about, however, would have been more surprising.
Control	N D D N	Birdhunting, we were shocked to see a **n**ew **d**o**d**o **kn**ocking on wooden posts.

focus) the underlined words ("shocked" and "more"), thus ensuring that the words containing the experimental target consonants did not receive a tonic accent. Sentence 1 probes for preboundary effects using the consonants (N D D). These consonants are compared to the parallel consonants in the control sentence (Sentence 3) in which the sequence is entirely phrase medial. Sentence 2 probes for postboundary effects using the consonants (D D N). They are likewise compared to the parallel phrase-medial consonants in the control utterance. The experimental design is schematized in Fig. 2. It should be noted that the preboundary consonant is not immediately before the boundary but is rather the onset consonant of the preboundary syllable; unlike the postboundary onset consonant, which is absolutely phrase-initial.

Four subjects, all native speakers of American English with no known speech or hearing disorders, participated. Subjects will be referred to as Subject A, Subject D (the first author), Subject E, and Subject J. The stimuli were pseudorandomized in blocks of three stimuli with no adjacent identical tokens. Subjects read each utterance 12 times and were instructed to read in a casual conversational style. (For Speaker D, 3 of the 36 tokens [1 control and 2 postboundary] were lost due to data collection error.)

The audio recordings were evaluated using ToBI guidelines (Beckman and Elam, 1997). They were realized with the expected intonation: The utterance testing the postboundary effect was realized with an intonation phrase boundary (marked by a boundary tone and a break index 4) after the word "Gnu." The utterance testing the preboundary effect was realized with an intonation phrase boundary (marked by a boundary tone and a break index 4) after the word "dodo." The control utterance was realized as one or as two intonation phrases ("birdhunting" was sometimes realized as a separate intonation phrase)—but, crucially, the relevant consonants under examination were all within one intonation phrase. Thus, we confirm that the experiment will be able to test the effects of the presence of an intonational phrase boundary on the consonant sequence using the no-boundary sentence as a control.

B. Data collection

The Carstens Articulograph (AG200) was used to track a sensor adhered to the tongue tip. (For a subset of the subjects, a tongue dorsum sensor was adhered as well, though it was irrelevant for the analysis design of this study.) Reference sensors were tracked on the maxilla and bridge of the nose and a sample of the occlusal plane of each subject was acquired. Articulatory data were sampled at 200 Hz, and acoustic data at 16 kHz. After data collection, data were corrected for head movement and rotated to the occlusal plane. The tongue tip y (vertical) signal was differentiated, and signals were smoothed before and after differentiation with a ninth-order Butterworth filter of cutoff frequency 15 Hz. [For all speakers' tongue tip constrictions, the x (horizontal) component of the movement was negligible; the consonants were dominantly formed in the vertical direction.]

C. Data analysis

For each of the consonants to be analyzed (N D D for the preboundary effects, D D N for the postboundary effects, and N D D N for the control condition), three timepoints were defined: The onset of the closing movement, the closing movement extremum (also initiation of consonant opening movement), and the end of the opening movement. These timepoints were defined by zero-crossings in the tongue tip velocity trajectory (For the last consonant ([n]) in all conditions, the endpoint of opening movement could not be reliably identified and is not analyzed.) By definition, the end of opening movement for consonant one is also the onset of closing movement for consonant two, etc.; that is, a total of eight timepoints were identified. There were cases where there were two zero-crossings for the tongue tip extremum for a consonant—the first consonant ([n]) across all sentences and the second consonant [d] in the [N # D D N] sentence—in these cases, the spatially highest peak was selected to designate the extremum point. Further, at the boundary, there were cases for subjects A, J, and E in which there were multiple zero-crossings at the end of the preboundary consonant opening, defining both the end of the opening movement and simultaneously the beginning of the closing movement. The zero-crossing selected for this point was that which was temporally closest to the highest peak velocity of the upcoming postboundary consonant.

Additionally, the spatial position of the tongue tip y trajectory was measured at each of these timepoints. Finally, for each closing and opening movement, the time of peak velocity was also measured, yielding an additional eight timepoints. In the case of preboundary [d] opening and postboundary [n] closing, multiple velocity peaks sometimes occurred, and in such cases, the fastest peak was chosen. Thirteen atypical tokens were excluded from the analysis.

For each consonant's closing movement and opening movement, the following dependent variables were derived from the measured timepoints of the tongue tip vertical (y) transducer:

1 Duration:
- For closing movement: Time from onset of closing movement to extremum.
- For opening movement: Time from extremum to end of opening movement.

2 Time-to-peak-velocity:
- For closing movement: Time from onset of closing movement to peak closing velocity.
- For opening movement: Time from onset of opening movement to peak opening velocity.

3 Extremum position (one tongue tip y extremum for each [n], [d], [d], and [n]).

4 Displacement:
- For closing movement: Euclidean distance between position at closing movement onset and at extremum.
- For opening movement: Euclidean distance between position at extremum and at end of opening movement.

These kinematic landmarks and derived variables are

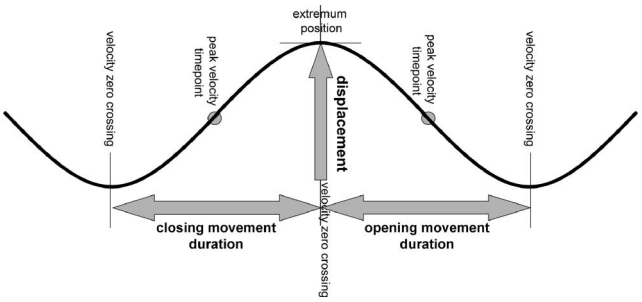


FIG. 3. Schema of the measured points in the tongue tip vertical (y) trajectory.

schematized in Figure 3 for a single consonant (note that displacement can potentially be affected by either or both onset and offset position). Time-to-peak-velocity has proven to be a good indicator of gestural stiffness within a mass-spring gestural model as modeled within Task Dynamics (Saltzman and Munhall, 1989) and similar models of movement control, i.e., the parameter shaping the internal temporal properties of an articulatory gesture. Stiffness is the primary control parameter that has been implicated in phrasal lengthening (e.g., Edwards *et al.*, 1991; Beckman and Edwards, 1992; Byrd *et al.*, 2000) and a slowed gating-in of target stiffness is one of the consequences of clock slowing within the π-gesture model (Byrd and Saltzman, 2003).

Two-tailed t-tests were conducted testing the preboundary and postboundary effect on duration, time-to-peak-velocity, extremum position, and displacement. Criterial significance was set at $p<0.05$. All and only statistically significant results are reported.

III. RESULTS

In presenting the details of temporal scope of effect, it will be convenient to numerically designate the consonant position relative to the utterance edge. To do this, we will use the designation C1 to refer to the consonant that is either the final (for the preboundary condition) or initial (for the postboundary condition) consonant in the test utterance, C2 to refer to the next-most preceding/following consonant, etc. Recall, however, that the while the postboundary C1 is immediately following the boundary, the preboundary C1 is the onset consonant of the preboundary syllable and thus not immediately at the final phrase edge.

Regarding acoustic pauses at the boundary, Speakers A, E, and J's pauses (i.e., silences) in the preboundary test utterance range from 350 to 650 ms; while for Speaker D's, only three were longer than 200 ms (the longest at 350 ms). Speakers A, E, and J's pauses (i.e., silences) in the postboundary test utterance range from 200 to 970 ms; while for Speaker D's, only four were longer than 200 ms (the longest at 350 ms), as with her preboundary condition sentence. In the preboundary condition, Speakers A, E, and J place the end of the opening movement for preboundary C1 in the second half of the acoustic pause; whereas for speaker D, this movement ends in all but one case during the vowel. Likewise, in the postboundary condition, Speakers A, E, and J place the beginning of the closing movement for postboundary C1 in the second half of the acoustic pause; whereas for speaker D, this movement starts during the preboundary vowel).

Another point to consider is that while this study explicitly controls for tonic accent, the possibility of variation in

TABLE II. Preboundary temporal scope of effect on C1: t-test results and means (SD) (ms). The reported n is the total number of tokens in the test plus the control conditions.

Preboundary temporal scope of effect on C1: C3 C2 C1]				
	C1 closing movement duration	C1 closing movement time-to-peak-velocity	C1 opening movement duration	C1 opening movement time-to-peak-velocity
Subject A $n=24$	n.s. Control: 76 (5) Test: 79 (5)	n.s. Control: 45 (4) Test: 48 (4)	$t=15.351$, $p<0.0001$ Control: 86 (13) Test: 564 (107)	$t=3.959$, $p=0.0007$ Control: 31 (3) Test: 36 (3)
Subject D $n=21$	$t=4.120$, $p=0.0006$ Control: 98 (7) Test: 113 (9)	$t=2.462$, $p=0.0235$ Control: 56 (7) Test: 65 (8)	$t=2.445$, $p=0.0244$; n=20 Control: 82 (6) Test: 143 (83)	n.s. Control: 45 (8) Test: 49 (6)
Subject E $n=23$	n.s. Control: 88 (7) Test: 86 (2)	n.s. Control: 55 (9) Test: 52 (3)	$t=19.268$, $p<0.0001$ Control: 96 (10) Test: 560 (79)	n.s. Control: 40 (5) Test: 41 (8)
Subject J $n=18$	n.s. Control: 69 (2) Test: 69 (3)	n.s. Control: 36 (2) Test: 37 (3)	$t=8.634$, $p<0.0001$ Control: 62 (8) Test: 396 (116)	$t=6.288$, $p<0.0001$ Control: 32 (3) Test: 41 (3)

TABLE III. Postboundary temporal scope of effects on C1 *t*-test results and means (SD) (ms). The reported *n* is the total number of tokens in the test plus the control conditions.

Post-boundary temporal scope of effect on C1: [C1 C2 C3				
	C1 closing movement duration	C1 closing movement time-to-peak-velocity	C1 opening movement duration	C1 opening movement time-to-peak-velocity
Subject A $n=24$	$t=3.584$, $p=0.0017$ Control: 70 (7) Test: 151 (78)	$t=3.071$, $p=0.0056$ Control: 37 (5) Test: 92 (61)	n.s. Control: 149 (21) Test: 131 (35)	n.s. Control: 97 (19) Test: 96 (31)
Subject D $n=18$	$t=5.857$, $p<0.0001$ Control: 92 (7) Test: 225 (76)	$t=4.940$, $p=0.0001$ Control: 49 (6) Test: 125 (51)	$t=3.720$, $p=0.0019$ Control: 107 (5) Test: 137 (26)	$t=3.927$ $p=0.0012$ Control: 63 (6) Test: 94 (25)
Subject E $n=22$	$t=5.067$, $p<0.0001$ Control: 65 (11) Test: 98 (18)	$t=5.369$, $p<0.0001$ Control: 29 (4) Test: 51 (13)	n.s. Control: 122 (6) Test: 118 (5)	n.s. Control: 73 (7) Test: 75 (3)
Subject J $n=20$	$t=2.448$, $p=0.0248$ Control: 91 (18) Test: 174 (100)	$t=2.964$, $p=0.0083$ Control: 31 (8) Test: 104 (73)	$t=4.063$, $p=0.0007$ Control: 120 (10) Test: 167 (34)	$t=4.194$, $p=0.0005$ Control: 85 (8) Test: 135 (35)

pre- and post-tonic accent exists. In the postboundary condition, all subjects place a very small accent on "dodo knocking" in *both* the control and the postboundary test sentences. Regarding the postboundary test sentence, for two speakers (A and J), there were differences in the accent size of F0 perturbation in the postboundary and control conditions, although, as noted above, the control condition also had a post-tonic accent. For the other two speakers, these differences do not occur. Therefore, it is reasonable to conclude that the consistency, observed below across subject results, reflects the prosodic boundary effects, not any accentual differences.

We will start with a discussion of the preboundary effects, and follow that with a discussion of the postboundary effects, with temporal effects considered first and spatial effects following.

A. Preboundary temporal scope of effect: C3 C2 C1] (N D D])

T-test results for the preboundary scope of temporal effects on C1 are shown in Table II; there were no significant results for preboundary C2 and C3 (these means are included in the Appendix). All subjects show a significant and large effect of a phrase boundary on the duration of opening movement of the preboundary C1 consonant, which was longer in the test than control utterance. Two subjects (A and J) show a significant effect in their time-to-peak-velocity for C1 opening movement in that reaching peak velocity takes longer in the test than in the control condition. One subject (D) shows a significant effect on early movement intervals, in that C1 closing movement and C1 closing movement time-to-peak-velocity are longer in the test than in the control condition.

B. Postboundary temporal scope of effect: [C1 C2 C3 ([D D N)

In the postboundary direction, all subjects have longer closing durations and time-to-peak-velocities for C1 in the test condition as compared to the control condition. *T*-test results are given in Table III. For two subjects (D and J), C1 opening movement also shows longer durations and time-to-peak-velocity.

However, an interesting and unexpected result was observed for C2 and C3, and this result was obtained across subjects for both the closing and opening movements of the postboundary C2 and C3. (Recall that while C3 opening movement time-to-peak-velocity was measured, C3 opening movement duration was not). Subjects show *shorter* durations for these consonants in the test condition compared to the control no-boundary condition. These effects were quite uniformly significant (see Table IV).

In sum, there is a significant lengthening effect on C1 closing and (for some subjects) opening movement, and a significant shortening effect for C2 and C3. It is important to note that while this shortening is compensatory in direction (relative to the preceding lengthening), the shortening was not nearly as large in magnitude as the phrase-initial lengthening.

C. Scope of spatial effects

The preboundary spatial effects are inconsistent across subjects and across consonants. Means are given in Table V and the Appendix and significant results detailed below. While all speakers have temporally longer C1 opening movement and two had longer time-to-peak-velocities, there are no consistent preboundary effects in the spatial domain. While three subjects have a significant effect on preboundary C1 opening displacement, some have larger movements

TABLE IV. Postboundary temporal effects on C2 and C3 *t*-test results and means (SD) (ms). NB: These are all shortening effects in direction for the test condition. The reported *n* is the total number of tokens in the test plus the control conditions.

	Postboundary temporal scope of effect on C2: [C1 C2 C3			
	C2 closing movement duration	C2 closing movement time-to-peak-velocity	C2 opening movement duration	C2 opening movement time-to-peak-velocity
Subject A $n=24$	$t=4.229$, $p=0.0003$ Control: 76 (5) Test: 68 (4)	$t=7.317$, $p<0.0001$ Control: 45 (4) Test: 35 (2)	$t=4.095$, $p=0.0005$ Control: 86 (13) Test: 64 (13)	n.s. Control: 31 (3) Test: 31 (6)
Subject D $n=18$	$t=4.283$, $p=0.0006$ Control: 98 (7) Test: 84 (8)	$t=3.908$, $p=0.0013$ Control: 56 (7) Test: 44 (6)	$t=6.494$, $p<0.0001$ Control: 82 (6) Test: 67 (3)	$t=3.599$, $p=0.0024$ Control: 45 (8) Test: 34 (5)
Subject E $n=22$	$t=3.551$, $p=0.002$ Control: 88 (7) Test: 80 (4)	$t=3.691$, $p=0.0014$ Control: 55 (9) Test: 45 (5)	$t=8.011$, $p<0.0001$ Control: 96 (10) Test: 65 (9)	$t=4.58$, $p=0.0002$ Control: 40 (5) Test: 32 (3)
Subject J $n=20$	$t=2.240$, $p=0.0379$ Control: 69 (2) Test: 65 (5)	$t=3.176$, $p=0.0052$ Control: 36 (2) Test: 33 (3)	$t=6.484$, $p<0.0001$ Control: 62 (8) Test: 40 (6)	$t=7.131$, $p<0.0001$ Control: 32 (3) Test: 21 (4)
	Postboundary temporal scope of effect on C3: [C1 C2 C3			
	C3 closing movement duration	C3 closing movement time-to-peak-velocity		C3 opening movement time-to-peak-velocity
Subject A $n=24$	$t=3.501$, $p=0.002$ Control: 84 (7) Test: 74 (7)	$t=4.357$, $p=0.0003$ Control: 47 (6) Test: 37 (4)		n.s. Control: 80 (7) Test: 76 (14)
Subject D $n=18$	$t=4.233$, $p=0.0006$ Control: 90 (7) Test: 70 (13)	n.s. Control: 40 (6) Test: 39 (2)		$t=2.473$, $p=0.025$ Control: 54 (5) Test: 49 (4)
Subject E $n=22$	$t=6.279$, $p<0.0001$ Control: 91 (6) Test: 77 (4)	$t=4.220$, $p=0.0004$ Control: 50 (9) Test: 37 (3)		$t=3.606$, $p=0.0018$ Control: 73 (5) Test: 65 (6)
Subject J $n=20$	$t=5.348$, $p<0.0001$ Control: 115 (18) Test: 81 (9)	$t=2.107$, $p=0.0494$ Control: 38 (7) Test: 33 (3)		n.s. Control: 62 (16) Test: 54 (3)

(Subject D, $t=4.8$, $p=0.0001$, $n=21$) and some smaller (Subject A $t=5.154$, $p<0.0001$, $n=24$; Subject E, $t=5.035$, $p<0.0001$, $n=23$). With regard to C1 closing, though one subject (D) has a longer preboundary C1 closing duration, no spatial effect is observed for her; only one subject (J) has an effect on C1 closing movement displacement ($t=2.992$, $p=0.0086$, $n=18$), such that the boundary condition is smaller than control. Subject J also shows smaller displacements in the boundary condition for C2 opening ($t=3.432$, $p=0.0034$, $n=18$) and closing movements ($t=2.896$, $p=0.0105$, $n=18$). Two subjects (A and J) have an effect on C3 opening (A: $t=2.137$, $p=0.044$, $n=24$; J: $t=2.403$, $p=0.0287$, $n=18$) in opposite directions, J again with smaller displacements in the boundary condition. Finally, Subject E shows a significant effect on C3 closing movement displacement ($t=2.120$, $p=0.0461$, $n=23$).

For extremum position, two subjects (E and J) show a significant effect for C2 (E $t=3.965$, $p=0.0007$, $n=23$; J $t=2.367$, $p=0.0309$, $n=18$) and one subject (E) for C3 ($t=3.556$, $p=0.0019$, $n=23$). In all instances, the test condition showed less extreme positions than the control condition.

In contrast, postboundary spatial effects are more consistent and pattern consistently with the observed temporal effects. Results are shown in Table V. All subjects show larger displacements for C1 closing movement in the boundary condition. Note that this was the position for which all subjects show postboundary lengthening. For C1 opening movement, one subject has a slightly larger displacement in the test condition; and for C2 displacement, Subject E shows more displacement and Subject J shows less displacement in the test condition. A consistent pattern of smaller displacements was obtained for C2 opening movement and C3 closing movement for all subjects' boundary condition. Note that these were the two positions that also show consistent compensatory temporal shortening.

We note that postboundary C1 displacement differences are driven almost entirely by differences in onset position of the movement; namely for C1, the tongue tip started at a

TABLE V. Postboundary spatial effects *t*-test results and means (SD) (mm). The reported *n* is the total number of tokens in the test plus the control conditions.

	Postboundary spatial scope of effect: [C1 C2 C3				
	C1 closing movement displacement	C1 opening movement displacement	C2 closing movement displacement	C2 opening movement displacement	C3 closing movement displacement
Subject A $n=24$	$t=5.503$, $p<0.0001$ Control: 4.0 (1) Test: 6.5 (1)	n.s. Control: 12.8 (1.1) Test: 11.5 (2)	n.s. Control: 10.4 (1.1) Test: 9.1 (2)	$t=5.662$, $p<0.0001$ Control: 7.7 (1.4) Test: 4.3 (1.6)	$t=5.491$, $p<0.0001$ Control: 9.8 (1.5) Test: 6.3 (1.6)
Subject D $n=18$	$t=9.38$, $p<0.0001$ Control: 3.1 (0.5) Test: 6.9 (1.2)	n.s. Control: 9.5 (1.2) Test: 10.3 (1.1)	n.s. Control: 8.6 (1) Test: 7.9 (1)	$t=3.455$, $p=0.0033$ Control: 5.5 (1.5) Test: 3.3 (0.7)	$t=2.973$, $p=0.009$ Control: 5.5 (1.4) Test: 3.7 (0.8)
Subject E $n=22$	$t=10.593$, $p<0.0001$ Control: 3.9 (2.1) Test: 11.3 (0.9)	$t=2.433$, $p=0.0245$ Control: 12.6 (1.3) Test: 13.6 (0.6)	$t=3.568$, $p=0.0019$ Control: 10.6 (1.2) Test: 12.1 (0.7)	$t=7.951$, $p<0.0001$ Control: 9.5 (0.7) Test: 5.9 (1.3)	$t=9.943$, $p<0.0001$ Control: 12.1 (0.5) Test: 7.3 (1.5)
Subject J $n=20$	$t=6.972$, $p<0.0001$ Control: 1.6 (0.8) Test: 7.2 (2.3)	n.s. Control: 13.2 (1) Test: 13.1 (1.6)	$t=2.154$, $p=0.045$ Control: 9.5 (0.6) Test: 8.7 (0.8)	$t=6.707$, $p<0.0001$ Control: 6.3 (1.7) Test: 2.1 (1.1)	$t=5.732$, $p<0.0001$ Control: 9.7 (1.9) Test: 5.6 (1.3)

lower position. (That is, for all subjects, there is a significant effect for C1 closing movement in that the test condition had lower onset positions.) For C2 and C3 displacement changes, however, both onset and extremum positions change differentially—contributing to individuals' displacement effects. Subject D had a postboundary effect on C2 extremum position ($t=2.918$, $p=0.0101$, $n=18$) and Subject E had an effect on C3 extremum position ($t=3.387$, $p=0.0029$, $n=22$), with the direction of the effect opposite for each subject. As for onset position, Subject E had an effect on C2 onset position, and for C3 all subjects except D showed such an effect—but the direction of the effect varied.

D. Summary of results

1. Summary of preboundary effect

The most salient preboundary results are the temporal effects. All subjects have longer C1 opening movement durations, and two subjects have a longer time-to-peak-velocity as well in this position. One subject also has longer duration and time-to-peak-velocity for the C1 closing movement. Recall that this closing movement was for the consonantal onset of the preboundary syllable, i.e., that a vowel removes this consonant from immediate adjacency to the phrase edge.

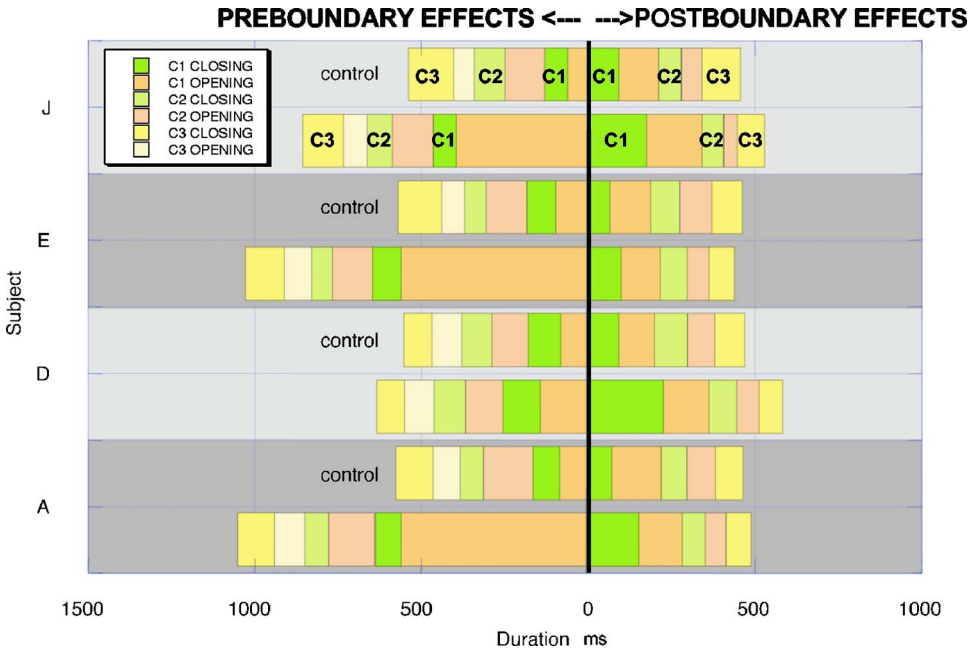


FIG. 4. (Color online) Consonant closing movement and opening movement duration results. The top bar for each speaker's pair is control and the bottom bar is test. (The bold line at center is merely a graphical convenience to separate results in the two conditions.)

The spatial effects are inconsistent across subjects and conditions. For the C1 opening movement, there are no consistent spatial effects to accompany the temporal effects—some subjects have larger C1 opening movements and some smaller. For the speaker who had a longer C1 closing movement duration, no spatial effect was observed.

2. Summary of postboundary effect

The temporal effects show a clear postboundary pattern. For C1, there was a lengthening effect. The C1 closing movement has a longer duration and longer time-to-peak-velocity for all subjects. The C1 opening movement shows longer opening movement duration and longer time-to-peak-velocity for two subjects.

However, the closing movements and the opening movement examined for C2 and C3 show a *shortening* after C1 lengthening. All subjects show a shortening in duration and time-to-peak-velocity for the C2 closing movement. For the C2 opening movement, all subjects show shortening in duration, and three subjects show shortening of time-to-peak-velocity. For the C3 closing movement, all subjects show shorter duration, and three subjects show shorter time-to-peak-velocity. For the C3 opening movement, time-to-peak-velocity is shorter for two subjects. Despite its compensatory direction, the degree of shortening is quite small; much smaller than previous lengthening.

Overall, the postboundary spatial effects consistently correspond to the temporal effects. That is, for C1 closing movement, all subjects show a larger displacement in the boundary condition. For C2 opening movement and for C3 closing movement, all subjects show smaller displacement in the boundary condition.

A summary of the temporal results that indicates the scope of boundary related temporal lengthening is shown in Fig. 4. (The bold line down the center of Fig. 4 is presented solely for the graphical ease of placing the preboundary and postboundary results on a single graph, i.e., it has no theoretical or methodological significance.)

IV. DISCUSSION

The main findings of the experiment are: (a) Preboundary temporal effects, (b) postboundary temporal and spatial effects, (c) followed further into the phrase by postboundary compensatory shortening effects (accompanied by smaller displacements).

The prosodic π-gesture model (Byrd and Saltzman, 2003) predicts that activation slowing of articulatory gestures will increase as the π-gesture's activation increases and wane as the π-gesture's activation decreases. Our results show a strong local lengthening of the opening movement of the C1 preboundary consonant (in the CV syllable preceding the boundary) and of the closing movement of the C1 postboundary consonant. One subject (D) also shows lengthening of the closing movement of the preboundary C1 consonant, and, importantly, this effect is smaller in magnitude than the effect on the phrase-final C1 opening (a change of 15 ms for closing movement versus 61 ms for opening movement). Postboundary, two subjects also have effects on the opening movement of C1—again, the effect is smaller than that shown for the immediately phrase-initial C1 closing (Subject D: Closing movement, the mean difference is 133 ms versus opening movement, 30 ms; Subject J closing movement, the mean difference is 91 ms versus opening movement, 48 ms). Thus, for the subjects for whom the boundary effect extends further than the boundary-adjacent movement—further leftward than C1 release preboundary, and further rightward than C1 closing postboundary—the effect of the boundary decreases with distance from the phrase edge (i.e., for Subject D, preboundary, and for Subjects D and J, postboundary). These decreasing effects are compatible with the π-gesture model in which the activation of the prosodic gesture extends over an interval, waxing and waning in strength.

The phrase-final and initial lengthening effects are roughly symmetrical, though the time-to-peak-velocity is affected only for two subjects preboundary; but for all four subjects, postboundary—presumably because the postboundary consonant is immediately at the phrase edge, unlike the preboundary syllable-onset consonant. Spatial effects are consistent for the postboundary articulations and pattern with the temporal effects, but this is not the case in the preboundary direction.

Particularly interesting results are the compensatory effects shown postboundary for C2 and C3, which to our knowledge have not been reported previously. It should be noted that they are much smaller than the lengthening effects, so are not "literally" compensatory. It is our view that the compensatory shortening is not a direct effect of the π-gesture; in other words, it is not a planned prosodic shortening that is occurring. Rather, it appears that, following the prosodic lengthening, there is an attraction of the constriction gestures back in the direction of the prosodically unmodified or unperturbed timing. If the prosodic lengthening were not compensated for at all, the effect of a phrase boundary would be to prolong the time of the utterance for precisely the duration of the local phrasal lengthening; this is unlikely to be the case. Studies on focal accent have suggested that while local gesture duration does affect global utterance duration, it does not prolong the utterance duration for the whole amount of the prosodic lengthening (Weismer and Ingrisano, 1979; consider also Lindblom and Rapp, 1973; Saltzman *et al.*, 2000). The compensatory shortening shown after a phrasal lengthening in our results indicates that the timing that would have been expected without prosodic warping is partially restored.

We conclude from the results of this articulatory kinematic experiment that the temporal prosodic effects on articulation are near the juncture, though not confined to the articulations immediately preceding or following the juncture; and they wax and wane in magnitude, both phrase finally and phrase initially. We further suggest that local gesture-to-gesture speech timing is sensitive to prosodic events, even downstream of those prosodic modifications.

APPENDIX

TABLE A1. Means (SD) (ms) for preboundary C2 and C3 durations (no significant differences exist).

n includes test+control	C2 closing movement duration	C2 closing movement time-to-peak-velocity	C2 opening movement duration	C2 opening movement time-to-peak-velocity	C3 closing movement duration	C3 closing movement time-to-peak-velocity	C3 opening movement duration	C3 opening movement time-to-peak-velocity
Subject A *n*=24	Control: 70 (7) Test: 72 (8)	Control: 37 (5) Test: 38 (5)	Control: 149 (21) Test: 136 (18)	Control: 97 (19) Test: 90 (15)	Control: 111 (30) Test: 110 (34)	Control: 40 (5) Test: 41 (5)	Control: 82 (25) Test: 92 (24)	Control: 80 (23) Test: 88 (19)
Subject D *n*=21	Control: 92 (8) Test: 95 (9)	Control: 49 (6) Test: 45 (7)	Control: 107 (5) Test:112 (10)	Control: 63 (6) Test: 63 (6)	Control: 82 (6) Test: 83 (7)	Control: 41 (4) Test: 43 (6)	Control: 90 (9) Test: 87 (8)	Control: 90 (9) Test: 87 (8)
Subject E *n*=23	Control: 65 (11) Test: 62 (3)	Control: 29 (4) Test: 30 (3)	Control: 122 (6) Test: 120 (4)	Control: 73 (7) Test: 69 (5)	Control: 129 (19) Test: 117 (32)	Control: 47 (3) Test: 50 (13)	Control: 69 (14) Test: 81 (24)	Control: 66 (12) Test: 73 (14)
Subject J *n*=18	Control: 91 (18) Test: 76 (18)	Control: 31 (9) Test: 47 (24)	Control: 120 (10) Test: 120 (6)	Control: 85 (8) Test: 84 (5)	Control: 133 (33) Test: 122 (34)	Control: 37 (8) Test: 38 (4)	Control: 64 (26) Test: 71 (28)	Control: 62 (21) Test: 67 (22)

TABLE A2. Means (SD) for preboundary displacement (mm) (* indicates a significant difference as detailed in text).

n includes test+ control	C3 closing movement displacement	C3 opening movement displacement	C2 closing movement displacement	C2 opening movement displacement	C1 closing movement displacement	C1 opening movement displacement
Subject A *n*=24	Control: 3.8 (0.7) Test: 4.1 (0.8)	*Control: 2 (0.8) Test: 2.8 (1.0)	Control: 4 (1.0) Test: 4.4 (0.9)	Control: 12.8 (1.1) Test: 12.8 (0.9)	Control: 10.4 (1.1) Test: 10.5 (1.3)	*Control: 7.7 (1.4) Test: 4.4 (1.7)
Subject D *n*=21	Control: 4.9 (0.7) Test: 4.9 (0.8)	Control: 3 (0.5) Test: 3 (0.8)	Control: 3.1 (0.5) Test: 3.3 (0.5)	Control: 9.5 (1.2) Test: 9.9 (0.8)	Control: 8.6 (1.0) Test: 9.4 (1.4)	*Control: 5.5 (1.5) Test: 9.0 (1.8)
Subject E *n*=23	Control: 9.7 (0.7) Test: 8.4 (2.0)	Control: 4.3 (1.7) Test: 3.9 (1.3)	Control: 3.9 (2.1) Test: 3.6 (1.1)	Control: 12.6 (1.3) Test: 12 (1.1)	Control: 10.6 (1.2) Test: 10.7 (0.7)	*Control: 9.5 (0.7) Test: 8.0 (0.7)
Subject J *n*=18	Control: 3.2 (0.9) Test: 3.2 (0.6)	*Control: 1.2 (0.5) Test: 0.7 (0.4)	*Control: 1.6 (0.8) Test: 8 (0.3)	*Control: 13.2 (1.0) Test: 11.7 (0.8)	*Control: 9.4 (0.6) Test: 8.1 (1.2)	Control: 6.3 (1.7) Test: 4.6 (2.7)

TABLE A3. Means (SD) for extremum position (mm) (* indicates a significant difference). Negative numbers refer to positions below the occlusal plane.

	Preboundary			Postboundary		
n includes test+control	C3 extremum	C2 extremum	C1 extremum	C1 extremum	C2 extremum	C3 extremum
Subject A *n*=24	Control: −2.3 (0.4) Test: −2.1 (0.8)	Control: −4.0 (5.0) Test: −0.5 (0.5)	Control: −2.7 (0.4) Test: −2.8 (0.9)	Control: −0.4 (0.5) Test: −0.6 (0.7)	Control: −2.7 (0.4) Test: −3.0 (0.7)	Control: −0.6 (0.7) Test: −1.0 (0.5)
Subject D *n*=21 (prebound); *n*=18 (postbound)	Control: 5.5 (0.8) Test: 5.3 (1.0)	Control: 5.6 (0.7) Test: 5.6 (0.9)	Control: 4.7 (0.4) Test: 5.1 (1.0)	Control: 5.6 (0.7) Test: 6.3 (0.9)	*Control: 4.7 (0.1) Test: 3.9 (0.3)	Control: 4.7 (0.6) Test: 4.3 (0.6)
Subject E *n*=23 (prebound); *n*=22 (postbound)	*Control: −5.089 (0.7) Test: −6.414 (1.0)	*Control: −5.5 (0.7) Test: −6.8 (0.8)	Control: −7.6 (0.6) Test: −8.0 (0.6)	Control: −5.5 (0.7) Test: −5.6 (0.5)	Control: −7.6 (0.6) Test: −7.1 (0.4)	*Control: −4.9 (0.6) Test: −5.8 (0.5)
Subject J *n*=18 (prebound); *n*=20 (postbound)	Control: −4.2 (0.7) Test: −4.4 (0.3)	*Control: −3.8 (0.5) Test: −4.3 (0.2)	Control: −7.6 (0.3) Test: −7.9 (1.0)	Control: −3.8 (0.5) Test: −3.7 (1.1)	Control: −7.6 (0.3) Test: −8.1 (1.1)	Control: −4.2 (0.7) Test: −4.6 (0.5)

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the NIH (Grant No. DC03172) and the assistance of Dr. James Mah, Elliot Saltzman and Shri Narayanan.

[1]It is interesting to speculate as to the accommodation of extremely long pauses, where in some sense the speech stream stops and restarts again, within the π-gesture framework. There are two (simultaneously viable) ways to view pauses within the π-gesture framework. First, one might view a pause as resulting from an extremely strong π-gesture, such that there is no qualitative difference between strong boundaries with and without a pause; a strong enough π-gesture simply slows articulation to an "effective" stop. It may be sensible in this regard to think about boundaries with pauses as having a flattened or plateaued π-gesture activation trajectory shape (Byrd and Saltzman 2003). A second possibility is to have overlapping phrase-final and phrase-initial π-gestures, *but* the model critically assumes these to have the same quality and nature of effect (Byrd and Saltzman 2003). In either case, a pause could be understood as a period during which planning of a gestural score for the upcoming utterance is occurring (see, Goldstein *et al.* 2006). (We thank Elliot Saltzman for helpful input on this matter.)

Beckman, M. E., and Edwards, J. (**1992**). "Intonational categories and the articulatory control of duration," in *Speech Perception, Production and Linguistics Structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Ohmsha, Tokyo, Japan), pp. 359–375.

Beckman, M. E., and Elam, G. A. (**1997**). "Guidelines for ToBI labelling," version 3.0, unpublished available online ⟨http://www.ling.ohio-state.edu/~tobi/ame_tobi/labelling_guide_v3.pdf⟩.

Berkovits, R. (**1993a**). "Progressive utterance-final lengthening in syllables with final fricatives," Lang Speech **36**, 89–98.

Berkovits, R. (**1993b**). "Utterance-final lengthening and the duration of final-stop closures," J. Phonetics **21**, 479–489.

Berkovits, R. (**1994**). "Durational effects in final lengthening, gapping, and contrastive stress," Lang Speech **37**, 237–250.

Browman, C. P., and Goldstein, L. (**1992**). "Articulatory phonology: An overview," Phonetica **49**, 155–180.

Byrd, D. (**2000**). "Articulatory vowel lengthening and coordination at phrasal junctures," Phonetica **57**, 3–16.

Byrd, D., Kaun, A., Narayanan, S., and Saltzman, E. (**2000**). "Phrasal signatures in articulation," in *Papers in Laboratory Phonology V. Acquisition and the Lexicon*, edited by M. B. Broe and J. B. Pierrehumbert (Cambridge University Press) pp. 70–87.

Byrd, D., Lee, S., Riggs, D., and Adams, J. (**2005**). "Interacting effects of syllable and phrase position on consonant articulation," J. Acoust. Soc. Am. **118**, 3860–3873.

Byrd, D., and Saltzman, E. (**1998**). "Intragestural dynamics of multiple phrasal boundaries," J. Phonetics **26**, 173–199.

Byrd, D., and Saltzman, E. (**2003**). "The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening," J. Phonetics **31**, 149–180.

Cambier-Langeveld, T. (**1997**). "The domain of final lengthening in the production of Dutch," in *Linguistics in the Netherlands 1997*, edited by J. Coerts and H. de Hoop (John Benjamins, Amsterdam), pp. 13–24.

Cho, T. (**2004**). "Prosodically conditioned strengthening and vowel-to-vowel coarticuation in English," J. Phonetics **32**, 141–176.

Cho, T. (**2006**). "Manifestation of prosodic structure in articulation: Evidence from lip movement kinematics in English," in *Laboratory Phonology 8: Varieties of Phonological Competence*, edited by L. Goldstein (Walter De Gruyter Inc., New York).

Cho, T., and Keating, P. (**2001**). "Articulatory and acoustic studies on domain-initial strengthening in Korean," J. Phonetics **29**, 155–190.

Cho, T., and Jun, S. A. (**2000**). "Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean," Chicago Linguistics Society **36**, 31–44.

Edwards, J., Beckman, M. E., and Fletcher, J. (**1991**). "The articulatory kinematics of final lengthening," J. Acoust. Soc. Am. **89**, 369–382.

Fougeron, C. (**2001**). "Articulatory properties of initial segments in several prosodic constituents in French," J. Phonetics **29**, 109–135.

Fougeron, C., and Keating, P. (**1997**). "Articulatory strengthening at edges of prosodic domains," J. Acoust. Soc. Am. **101**, 3728–3740.

Gendrot, C. (**2005**). "Acoustic, kinematic, and aerodynamic aspects of word-initial and word-final vowels in preboundary context in French." In ZAS: Zeitschrift für Sprachwissenschaft **40**, 45–61.

Goldstein, L, Byrd, D., and Saltzman, E. (**2006**). "The role of vocal tract gestural action units in understanding the evolution of phonology," in *From Action to Language: The Mirror Neuron System*, edited by M. Arbib (Cambridge University Press, Cambridge, UK).

Hacopian, N. (**2003**). "A three-way VOT contrast in final position: Data from Armenian," J. Int. Phonetic Assoc. **33**, 51–80.

Keating, P., Cho, T., Fougeron, C., and Hsu, C. (**2004**). "Domain-initial articulatory strengthening in four languages," in *Phonetic Interpretation (Papers in Laboratory Phonology VI)*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press), pp. 143–161.

Keating, P., Wright, R., and Zhang, J. (**1999**). "Word-level asymmetries in consonant articulation," UCLA Working Papers in Phonetics **97**, 157–173.

Klatt, D. (**1976**). "Linguistics uses of segmental duration in English: Acoustic and perceptual evidence," J. Acoust. Soc. Am. **59**, 1208–1221.

Lindblom, B., and Rapp, K. (1973). "Some temporal regularities of spoken Swedish," PILUS (Papers from the Institute of Linguistics, University of Stockholm) **21**, 1–59.

McClean, M. (**1973**). "Forward coarticulation of velar movement at marked junctural boundaries," J. Speech Hear. Res. **16**, 286–296.

Oller, K. D. (**1973**). "The effect of position in utterance on speech segment duration in English," J. Acoust. Soc. Am. **54**, 1235–1247.

Saltzman, E., Löfqvist, A., and Mitra, S. (**2000**). "'Glue' and 'clocks': Intergestural cohesion and global timing," in *Papers in Laboratory Phonology V*, edited by M. B. Broe and J. B. Pierrehumbert (Cambridge University Press, Cambridge), pp. 88–101.

Saltzman, E. L., and Munhall, K. G. (**1989**). "A dynamical approach to gestural patterning in speech production," Ecological Psychol. **1**, 333–382.

Shattuck-Hufnagel, S., and Turk, A. (**1998**). "The domain of phrase-final lengthening in English," *The Sound of the Future: A Global View of Acoustics in the 21st Century, Proceedings 16th International Congress on Acoustics and 135th Meeting Acoustical Society of America*, 1235–1236.

Tabain, M. (**2003**). "Effects of prosodic boundary on /aC/ sequences: Articulatory results," J. Acoust. Soc. Am. **113**, 2834–2849.

Tabain, M., and Perrier, P. (**2005**). "Articulation and acoustics of /i/ in preboundary position in French," J. Phonetics **33**, 77–100.

Turk, A. E. (**1999**). "Structural influences on boundary-related lengthening in English," *Proceedings of the XIVth International Congress of Phonetic Sciences*, San Francisco, edited by J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey (The Regents of the University of California) **1**, 237–240.

Weismer, G., and Ingrisano, D. (**1979**). "Phrase-level timing patterns in English: Effects of emphatic stress location and speaking rate," J. Speech Hear. Res. **22**, 516–533.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (**1992**). "Segmental durations in the vicinity of prosodic phrase boundaries," J. Acoust. Soc. Am. **91**, 1707–1717.

Voiceless affricate/fricative distinction by frication duration and amplitude rise slope

Shigeki Mitani, Toshihiro Kitama, and Yu Sato[a)]
Department of Physiology, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Chuo, Yamanashi 409-3898, Japan

(Received 12 July 2005; revised 23 May 2006; accepted 2 June 2006)

Previous psychophysical studies have shown that the perceptual distinction between voiceless fricatives and affricates in consonant-vowel syllables depends primarily on frication duration, whereas amplitude rise slope was suggested as the cue in automatic classification experiments. The effects of both cues on the manner of articulation between /ʃ/ and /tʃ/ were investigated. Subjects performed a forced-choice task (/ʃ/ or /tʃ/) in response to edited waveforms of Japanese fricatives /ʃi/, /ʃu/, and /ʃɑ/. We found that frication duration, onset slope, and the interaction between duration and onset slope influenced the perceptual distinction. That is, the percent of /ʃ/ responses increased with an increase in frication duration (experiments 1–3). The percent of /ʃ/ responses also increased with a decrease in slope steepness (experiment 3), and the relative importance between slope portions was not even but weighted at onset (experiments 1 and 2). There was an interaction between the two cues of frication duration and steepness. The relative importance of the slope cue was maximum at a frication duration of 150 ms (experiment 3). It is concluded that the frication duration and amplitude rise slope at frication onset are acoustic cues that discriminate between /ʃ/ and /tʃ/, and that the two cues interact with each other. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2221390]

PACS number(s): 43.71.Es [JES] Pages: 1600–1607

I. INTRODUCTION

Several acoustic cues for manner of articulation between voiceless fricatives and affricates in consonant-vowel (CV) syllables have been demonstrated over the past decades. Some psychophysical experiments suggested that frication duration is one of acoustic cues for distinguishing voiceless fricatives from affricates. That is, Grim (1966) showed that shortening the frication duration by truncation of the initial part resulted in a change of manner of articulation from fricatives to plosives. A minimum frication duration of 60–80 ms was required for more than 75% correct identification of fricatives. Furui (1986) reported similar findings by elimination of the initial part of the frication of Japanese CV syllables. Repp *et al.* (1978) also shortened the frication duration by truncation of the middle part of frication and showed similar change of the manner of articulation between /ʃ/ and /tʃ/.

The rise time from the frication onset to the time at the maximum amplitude of the frication envelope also has been considered as an important cue for the manner of articulation. Cutting and Rosner (1974) investigated the distinction between fricatives and affricates of synthetic syllables with different rise times, and showed that the percent /ʃ/ responses increased with an increase in rise time. Conversely, the percent /tʃ/ responses decreased as rise time increased. Similar findings were reported by manipulation of the rise time of naturally spoken CV syllables (Howell and Rosen, 1983). Kluender and Walsh (1992), however, pointed out that previous studies concerning the rise time were confounded by covariation of frication duration (Cutting and Rosner, 1974; Howell and Rosen, 1983). Thus, Kluender and Walsh conducted two experiments. In the first experiment with a variable rise time and fixed frication duration at 140 or 160 ms, the rise time had little effect on the distinction between fricatives and affricates. On the other hand, in the second experiment with variable frication duration and fixed rise time at 30 or 80 ms, the percent /ʃ/ responses increased with increased frication durations. Kluender and Walsh (1992) suggested that the changes in rise time were not sufficient to signal the distinction between fricatives and affricates.

Weigelt *et al.* (1990) have demonstrated that the amplitude rise slope, which is the rate of increase in the root-mean-square (rms) amplitude of the acoustic wave, is a useful acoustic cue for automatic distinction between fricatives and affricates of CV syllables. Their algorithm using only amplitude rise slope but not frication duration achieved a 96.8% correct in the overall rate of the distinction.

There are no psychophysical studies investigating the effects of the amplitude rise slope on the distinction between fricatives and affricates. This study was designed to investigate the role of the amplitude rise slope on discrimination between the fricative /ʃ/ and the affricate /tʃ/.

II. EXPERIMENT 1

A. Method

1. Stimuli

Japanese syllables /ʃi/ /ʃu/, and /ʃɑ/ from a phonetically trained male speaker were recorded in a sound-proof room using a microphone with preamplifier (Bruel & Kjaer 4190/

[a)]Author to whom correspondence should be addressed. Electronic email: yu_sato@yamanashi.ac.jp

2690). The microphone output was fed into a band-pass filter (100–11 000 Hz: NF Electronic Instruments, p-86 and p-87) and then through a 12-bit analog-to-digital converter (National Instruments, PCI-MIO-16E-4), recorded in the hard disk of a computer at a 22-kHz sampling rate.

We defined the onset and offset of the frication. The onset was the first noticeable turbulence in the sound waveform (Behrens and Blumstein, 1988) and the offset was the intensity minimum immediately preceding the onset of vowel periodicity (Jongman *et al.*, 2000). The /ʃi/ syllable was 575 ms in total duration with a frication noise portion of 284 ms; the /ʃu/ syllable, 534 ms with a noise portion of 292 ms; and the /ʃɑ/ syllable, 410 ms with a noise portion of 256 ms. Waveform editing was carried out with the use of MATLAB (Mathworks) as follows. The initial 34, 42, and 6 ms of very low energy frication noise were deleted from /ʃi/, /ʃu/, and /ʃɑ/ syllables, respectively, resulting in 250-ms frication noises [Fig. 1(a)]. We estimated rms amplitudes of the frication portion every 5 ms, constructing an rms sound-envelope [Fig. 1(b)]. The maximum amplitudes of the rms sound-envelopes (MArms) of /ʃi/, /ʃu/, and /ʃɑ/ syllables were 0.125, 0.152, and 0.072 Pa, respectively, which were normalized to 0.125 Pa. The amplitude rise phases from frication onset to MArms of /ʃi/, /ʃu/, and /ʃɑ/ syllables were 230, 220, and 230 ms, respectively, whereas the amplitude fall phases were 20, 30, and 20 ms, respectively.

The following procedures of waveform editing were done across all /ʃi/, /ʃu/, and /ʃɑ/ syllables. Sound wave amplitudes of the frication noise during the rise phase [Fig. 1(a)] were divided by the rms envelope amplitudes [Fig. 1(b)] and multiplied by MArms, resulting in equalized rise-phase frication and nonmanipulated fall-phase frication [sound wave in Fig. 1(c) and its rms envelope in Fig. 1(d)].

The signals represented by Fig. 1(c) were then multiplied by a rise function, 250 ms in duration [Fig. 1(e)] and given by

$$f(t) = 0 \qquad \text{for } 0 \le t < t_0,$$
$$f(t) = ((t - t_0)/\mathrm{RT})^p \qquad \text{for } t_0 \le t < \mathrm{RT} + t_0,$$
$$f(t) = 1 \qquad \text{for } \mathrm{RT} + t_0 \le t < 250 \text{ ms},$$

where t is the time from frication onset, t_0 is the onset time of rise slope, p is the exponent, taking the value of 0.5, 1.0, or 2.0 (as shown in Fig. 4(a)), and RT is the rise time duration from t_0 to the time of the maximum amplitude (RT is fixed in this study at 60 ms). The fricative duration (FD) was systematically shifted from 110 to 190 ms in steps of 20 ms (t_0 was correspondingly varied from 140 to 60 ms). Finally, the generated stimulus sound wave with $p = 0.5$ ($S^{0.5}$) together with its rms envelope are shown in Figs. 1(f) and 1(g), respectively.

We checked the change of spectral envelope by amplitude manipulation. The spectral analysis in the frication portion was performed with the use of a Hamming window and linear predictive coding (LPC, 24 poles). The LPC spectral contrasts of the original fricative noises of /ʃ/ (/ʃi/, /ʃu/, and /ʃɑ/) changed little after processing [Fig. 1(h)]. Thus, in this study, we succeeded in preserving the spectral cues despite manipulation of temporal cues.

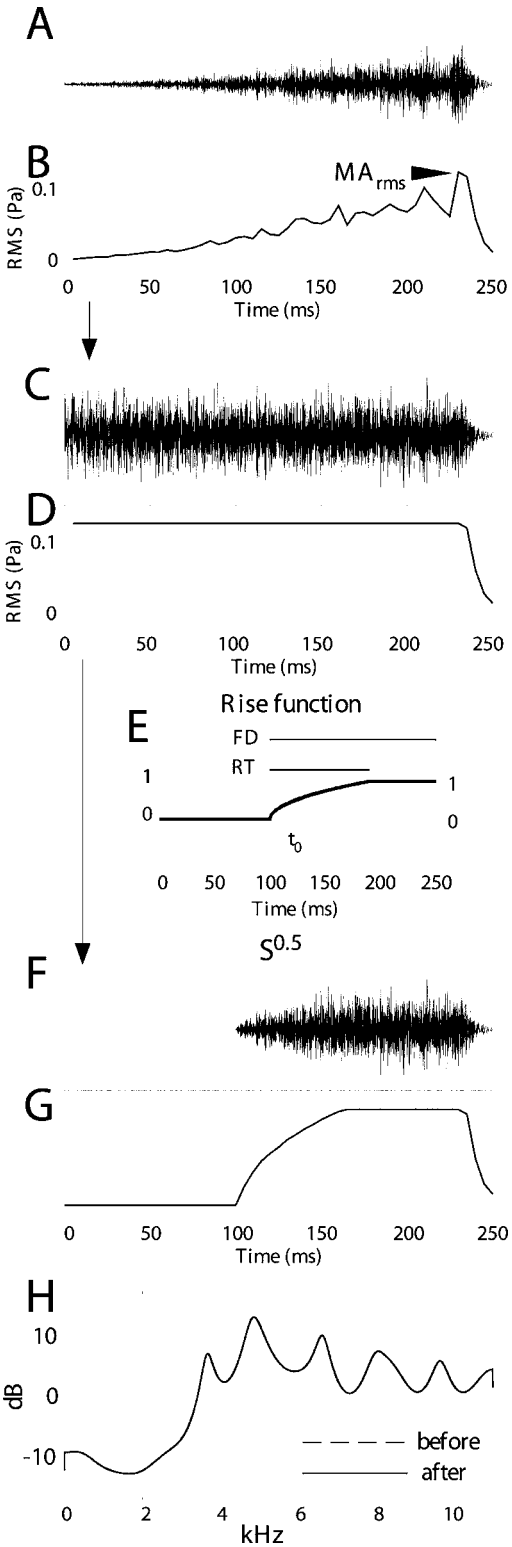


FIG. 1. Procedures to edit sound wave. (a) and (b) Original sound wave and rms envelope of frication portion of /ʃɑ/, respectively. (c) and (d) Equalized sound wave and rms envelope. (e) Amplitude rise function: exponent=0.5. (f) and (g) stimulus sound wave and rms envelope of $S^{0.5}$. (h) LPC envelopes before and after manipulation. Mean of spectral amplitudes in dB SPL was normalized as 0.

This study used only one token for each of the three fricative-vowel combinations for all of the perception tests. We investigated LPC envelopes in ten speakers (25–37 years of age, five males and five females). As shown in Fig. 2(a),

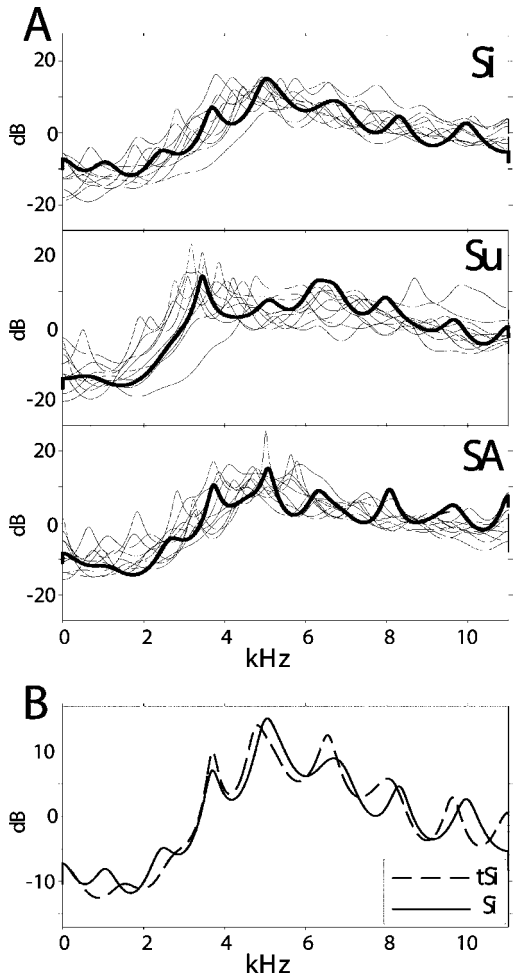


FIG. 2. Spectral analyses of frication noises. (a) LPC envelopes of tokens of a male speaker employed as stimulus envelopes (thick lines) and those of ten speakers not employed (thin lines). (b) LPC envelopes of original /ʃ/ and t/ʃ/ of a male speaker. Spectral envelope of /ʃ/ (solid line) is the same as that of /ʃ/ (thick solid line) in (a). Mean of spectral amplitudes in dB SPL was normalized as 0.

LPC envelopes of /ʃ/ (/ʃi/, /ʃu/, and /ʃɑ/) adopted in this study (thick lines) are similar to those in the ten speakers (thin lines). Thus, we considered that the tokens adopted in this study were typical concerning spectral parameters.

We also checked spectral differences among original /ʃ/ (/ʃi/, /ʃu/, and /ʃɑ/) and /tʃ/ (/tʃi/, /tʃu/, and /tʃɑ/), respectively, in the ten speakers. The LPC spectral contrasts of original /ʃ/ were similar to those of original /tʃ/ [Fig. 2(b)]. Correctively, manipulation procedures employed in this study provided reasonable examples of speech concerned with spectral cues.

Typical values for fricatives and affricates in Japanese are shown in Table I. It shows mean values (and SD) of frication duration, rise time, and maximum amplitude across three fricative-vowel and three affricate-vowel combinations from the ten speakers. Neither vowel context nor speaker's gender affected the acoustic parameters of frication duration, rise time, and maximum amplitude significantly (a two way ANOVA, $p > 0.05$). Figure 3 shows ranges of frication duration (a), rise time (b), and maximum amplitude (c) of /tʃi/, /tʃu/, /tʃɑ/, /ʃi/, /ʃu/, and /ʃɑ/ (total 50 tokens, respectively) from the 10 speakers. All 50 tokens were clearly identified as /ʃi/, /ʃu/, /ʃɑ/, /tʃi/, /tʃu/, and /tʃɑ/, respectively, by the authors and authors' colleagues. Nevertheless, large overlaps were found among consonants with a different manner of articulation, suggesting that those acoustic parameters (frication duration, rise time, and maximum amplitude) are not enough for distinguishing fricatives from affricates. The range of frication duration (110–190 ms), rise time (60 ms), and maximum amplitude (0.125 Pa) adopted as stimulus parameters in this study are shown by vertical interrupted lines.

TABLE I. Mean and standard deviation (SD) of frication duration (FD), rise time (RT), and maximum amplitude (MA) across three affricate-vowel and three fricative-vowel combinations from speakers of five males (M) and five females (F).

		FD (ms)		RT (ms)		MA (Pa)	
	Gender	Mean	SD	Mean	SD	Mean	SD
t/ʃi/	M	98.16	11.12	68.80	15.19	0.142	0.066
	F	97.36	54.28	61.88	37.46	0.093	0.066
t/ʃu/	M	91.61	12.67	58.08	19.29	0.181	0.108
	F	103.40	61.49	61.64	38.51	0.106	0.090
t/ʃɑ/	M	64.36	19.97	41.08	12.38	0.098	0.050
	F	61.08	28.09	35.00	15.24	0.087	0.068
/ʃi/	M	218.72	34.44	168.84	41.08	0.118	0.066
	F	196.92	63.43	133.36	24.82	0.099	0.081
/ʃu/	M	224.52	40.91	176.00	48.19	0.142	0.054
	F	196.28	76.62	141.16	50.81	0.082	0.059
/ʃɑ/	M	163.64	35.62	121.44	25.67	0.083	0.032
	F	15.76	46.94	120.96	39.18	0.068	0.044

As shown in Fig. 1(g), the signal has been reduced to 0 Pa for a varying number of milliseconds at the beginning of the processed fricative. If we start stimuli at 0 ms in Fig. 1(g) under ambient noise, the length of the silence at the beginning is an unintentional additional cue that could well mimic the closed portion of an affricate. To avoid the unintentional silence cue, we provided an additional 1000 ms of 0 Pa before the 0 ms. Thus, the silent period is enough long (>1000 ms) so that listeners do not interpret the silence as an acoustic cue.

The processing method was thus designed to change temporal parameters of the fricative noise at onset but not offset, supposing that the onset is more important perceptually. In fact, the first step of the processing was to truncate the initial portion of the tokens in order to equalize the lengths of the three tokens. One may wonder how this is justified. However, remember that the original tokens were not used as stimuli but used just for providing reasonable spectral cues as speech to the stimuli. The truncation of the initial portions with low signal-to-noise ratio did not affect the spectral cues of the stimuli.

The above-mentioned manipulation resulted in three different accelerated patterns of rms rise slopes: (1) $S^{0.5}$ has steep rise slope at onset and a gradually decreasing steepness; (2) S^1 has linear rise of RMS slope; and (3) S^2 has a gentle rise of rms slope at frication onset and a gradually increasing steepness [Fig. 4(a)].

The stimuli were fed into a 12-bit digital/analog converter (National Instruments, PCI-MIO-16E-4) and then low-pass filtered (NF Electric Instruments p-86) with a cutoff

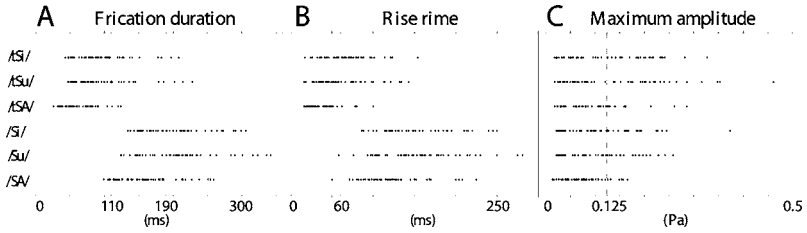


FIG. 3. Distribution of frication duration (a), rise time (b), and maximum amplitude (c) of /ʃi/, /ʃu/, /ʃɑ/, t/ʃi/, t/ʃu/, and t/ʃɑ/ of 50 tokens from ten speakers. Vertical interrupted lines show (a) range or (b) and (c) fixed value of stimulus parameters adopted in this study.

frequency of 11 kHz. The output from the filter was attenuated, and through a low-output-impedance power amplifier (Denon, PMA2000), the stimuli were presented from open-ear headphones (AKG K1000) binaurally in a sound proof room.

2. Participants and forced-choice tasks

Thirteen undergraduate students (18–30 years of age, nine males and four females) participated in this experiment. All students were native Japanese speakers with normal hearing (better than 15 dB HL for octave-separated frequencies between 125 and 8000 Hz in both ears). The subjects were paid for their participation. The experiments were undertaken with the understanding and written consent of each participant. The listeners were instructed to perform a forced choice task to label each stimulus token as either /ʃ/ or /tʃ/ by pressing one of the two labeled response buttons. The percent /ʃ/ responses was measured by giving 5 FDs (110, 130, 150, 170, and 190 ms) $\times 3$ slope patterns $\times 10$ repetitions $\times 3$ vowels $=450$ trials (random order) to each subject.

The forced-choice experiment design biases the subjects to interpret the stimuli as speech. The subjects were debriefed after the experiments, and they considered them to be reasonable examples of the particular segments they were forced to choose between.

B. Results

The percent of /ʃi/, /ʃu/, and /ʃɑ/ responses (mean and SE of 13 subjects) were plotted against FD [Figs. 4(b)–4(d)]. Because the results were similar in the three vowel conditions as shown in Figs. 4(b)–4(d), the percent /ʃi/, /ʃu/, and /ʃɑ/ responses were averaged as the percent /ʃ/ responses [Fig. 4(e)]. It was noted that (1) the percent /ʃ/ responses significantly increased with an increase in FD [a two-way ANOVA, $F(4,48)=90.60$, $P<0.01$], (2) the rise slope patterns also significantly affected the percent /ʃ/ response [a two-way ANOVA, $F(2,24)=37.63$, $P<0.01$], (3) there was a significant FD $\times$ slope interaction [$F(8,96)=11.97, P<0.01$], and (4) the percent /ʃ/ responses to S^1 or S^2 stimuli was always higher than those with $S^{0.5}$ stimuli. The differences in percent /ʃ/ responses between stimuli S^1 and $S^{0.5}$($DS^{1-0.5}$) and between stimuli S^2 and $S^{0.5}$($DS^{2-0.5}$) were significant (Scheffe *posthoc* test: $P<0.05$) but not between stimuli S^1 and S^2($P>0.05$) [4(e)].

As mentioned above, a significant interaction between FD and slope parameters was found. This means listeners identified two slope differences according to a shift of frication duration. $DS^{1-0.5}$ was thus plotted against FD [Fig. 4(f), broken line], making the $DS^{1-0.5}$-FD function. The function had rising and then falling slopes with increased FD, peaking at 150 ms of FD. The maximum difference (the maximum $DS^{1-0.5}$) was significantly different from the differences at other FD values (paired t test: $P<0.01$ at FD=110, 170, 190 ms; $P<0.05$ at FD=130 ms), suggesting that relative importance in slope cue changes with duration cue. $DS^{2-0.5}$ was also plotted against FD [Fig. 4(f), solid line], making the $DS^{1-0.5}$-FD function. The $DS^{2-0.5}$-FD function was similar to the $DS^{1-0.5}$-FD function, peaking at 150 ms of FD. The maximum difference was significantly different from the differences at other FD values (paired t test: $P<0.01$).

C. Discussion

In experiment 1, we found three major results. First, the subjects tended to recognize the manipulated consonant in the CV syllables as fricatives at longer frication durations. Inversely, they tended to recognize them as affricates at shorter frication durations. These findings are in accord with the previous studies that frication duration is one of cues for manner of articulation in CV syllables (Grim, 1966; Repp *et al.*, 1978; Furui, 1986; Kluender and Walsh, 1992; Hedrick,

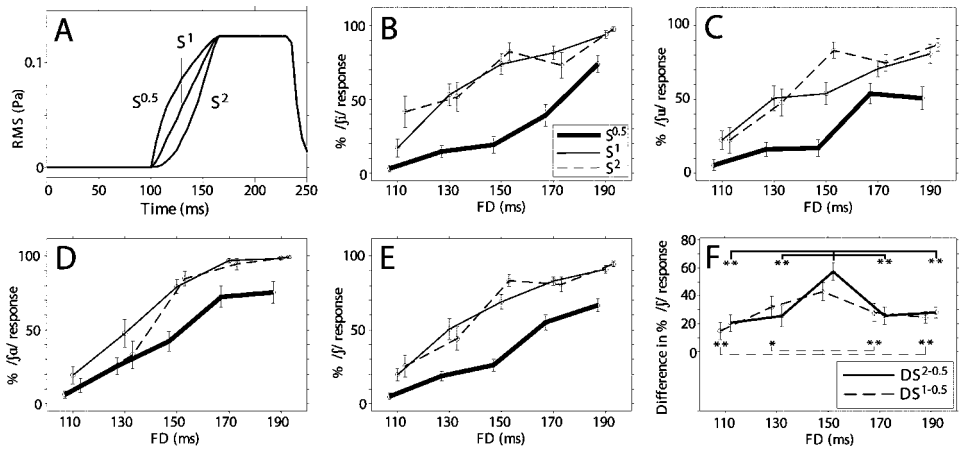


FIG. 4. Experiment 1. (a) The rms envelopes of stimuli $S^{0.5}$, S^1, and S^2. (b)–(d) Mean and SE of percent /ʃi/, /ʃu/, and /ʃɑ/ responses to stimuli $S^{0.5}$, S^1, and S^2. (e) Average of the percent /ʃi/, /ʃu/, and /ʃɑ/ responses (percent /ʃ/ responses) to stimuli $S^{0.5}$, S^1, and S^2. (f) Mean and SE of $DS^{2-0.5}$ and $DS^{1-0.5}$. *, $P<0.05$; **, $P<0.01$.

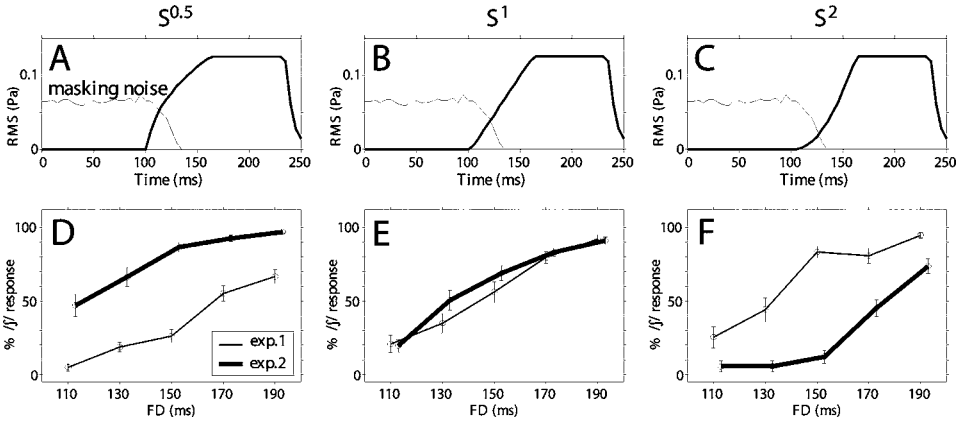


FIG. 5. Experiment 2. (a)–(c) The rms-envelopes of masking noise and stimuli $S^{0.5}$, S^1, and S^2 (FD=150 ms), respectively. (d)–(f) Percent /ʃ/ responses at stimuli $S^{0.5}$, S^1, and S^2, respectively. Thick lines, results from experiment 2; thin lines, results from experiment 1.

1997) and in the edited CV syllables including only the consonant part (Jongman, 1989). Second, it is clear that the rise slope pattern is also a cue for manner of articulation, as the percent /ʃ/ responses varied with different rise slope patterns [Figs. 4(b)–4(e)] in spite of the same rise time and the rise amplitude [Fig. 4(a)]. Third, we have shown that the relative importance of the rise slope pattern is a function of frication duration: the difference in percent /ʃ/ responses between the rise slope patterns was the maximum when frication duration was at 150 ms [Fig. 4(f)].

The stimuli used in experiment 1 were characterized by three different rise slope patterns. $S^{0.5}$ had a steep slope at onset gradually becoming gentle, S^1 had a linear slope during the rise phase, whereas S^2 had a gentle slope at onset gradually becoming steep. Which part of the slope pattern was effective for distinguishing between /ʃ/ and /tʃ/ in experiment 1? We assume that the rise slope at onset (but not at later phase) affects the distinction, because it is known that the neurons in the primary auditory cortex tend to respond to the onset slope of the sound envelope (Heil, 1997). The steep onset slope of $S^{0.5}$ may result in low percent /ʃ/ responses, while the gentle onset slope of S^2 may result in high percent /ʃ/ responses. If so, masking the onset envelope of $S^{0.5}$, which makes the unmasked part of onset slope gentle, would result in an increase of the percent /ʃ/ responses, whereas masking the onset envelope of S^2, which makes the unmasked part of onset slope steep, would result in a decrease of the percent /ʃ/ responses. Masking the onset envelope of S^1, which has linear slope, was not expected to change the percent /ʃ/ responses. We tested the onset slope hypothesis in experiment 2.

III. EXPERIMENT 2

As mentioned above, we assumed that the relative importance between rise slope phases was not even but weighted on the onset phase. In experiment 2, we tested if masking the onset phase of the rise slope results in a change in perceptual discrimination between /ʃ/ and /tʃ/.

A. Method

1. Stimulus

The following waveform editing procedures were used for all /ʃi/, /ʃu/, and /ʃɑ/ syllables. Gaussian white noise was generated with the use of Matlab (Mathworks Inc.), which was utilized for masking the onset portion of the frication. The white noise duration was varied from 340 (when FD was 190 ms) to 420 ms (when FD was 110 ms) with a rise time of 100 ms and a fall time of 30 ms. The fall time of the white noise overlapped the initial 30 ms of the frication [$S^{0.5}$, S^1, and S^2 in Figs. 5(a)–5(c), respectively]. Thus, the duration from the white noise onset to the frication offset was fixed at 500 ms. The rms amplitude of the white noise at the steady phase was set to half of MArms [Figs. 5(a)–5(c)].

2. Subjects and forced-choice tasks

The same 13 subjects in experiment 1 participated in experiment 2. The percent /ʃ/ responses was measured by giving 5 FDs × 3 slope patterns × 10 repetitions × 3 vowels =450 trials (random order) to each subject. The listeners' tasks were the same as those in experiment 1.

B. Results

Because the results were similar in the three vowel conditions of /ʃi/, /ʃu/, and /ʃɑ/, the three different vowel responses were averaged as the percent /ʃ/ responses. The percent of /ʃ/ responses (mean and SE of 13 subjects) from experiment 2 is shown [Figs. 5(d)–5(f), thick line] in addition to those from experiment 1 [Figs. 5(d)–5(f), thin line].

It was noted that (1) the percent of /ʃ/ responses to stimulus $S^{0.5}$ in experiments 1 and 2 significantly increased with an increase in FD [a two-way ANOVA, $F(4,48)=81.91$, $P<0.01$], (2) the percent of /ʃ/ responses to stimulus $S^{0.5}$ was significantly increased by masking the initial slopes [a two-way ANOVA, $F(1,12)=85.03$, $P<0.01$], and (3) there was a significant FD× noise-condition interaction [$F(4,48)=7.05$, $P<0.01$] [Fig. 5(d)].

Concerning the stimulus S^2, (1) the percent /ʃ/ responses to stimulus S^2 in experiments 1 and 2 significantly increased with an increase in FD [a two-way ANOVA, $F(4,48)=117.09$, $P<0.01$], (2) the percent /ʃ/ responses to stimulus S^2 were significantly decreased by the masking [a two-way ANOVA, $F(1,12)=79.42$, $P<0.01$], and (3) there was a significant FD× noise-condition interaction [$F(4,48)=20.34$, $P<0.01$] [Fig. 5(f)].

Concerning the stimulus S^1, (1) the percent /ʃ/ responses to stimuli S^1 in experiments 1 and 2 significantly increased with an increase in FD [a two-way ANOVA, $F(4,48)=85.02$, $P<0.01$], (2) the effect of the masking on the per-

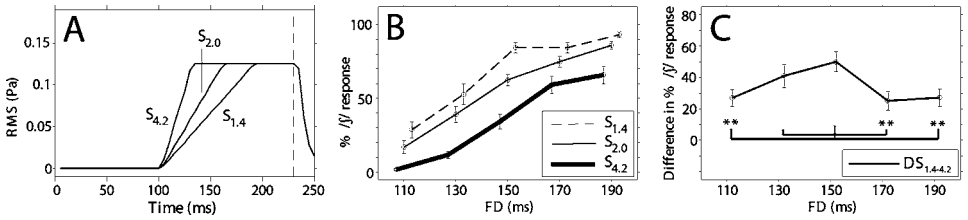


FIG. 6. Experiment 3. (a) The rms envelopes of stimuli $S_{1.4}$ (1.4 Pa/s), $S_{2.0}$ (2.0 Pa/s), and $S_{4.2}$ (4.2 Pa/s). (b) Mean and SE of the percent /ʃ/ responses to $S_{1.4}$, $S_{2.0}$, and $S_{4.2}$ stimuli. (c) Mean and SE of $DS_{1.4-4.2}$. **, $P<0.01$.

cent /ʃ/ responses was not significant [a two-way ANOVA, $F(1,12)=4.08, P>0.05$], and (3) there was a significant FD × noise-condition interaction [$F(4,48)=4.59, P<0.01$] [Fig. 5(e)].

C. Discussion

It is noteworthy that the masking experiments resulted in an opposite shift of percent /ʃ/ responses: it increased by masking the onset of $S^{0.5}$ [Fig. 5(d)]. Conversely, the percent /ʃ/ responses decreased by masking the onset of S^2 [Fig. 5(f)]. The opposite shift of the percent /ʃ/ responses may be explained by the shift of steepness of the perceived onset slope by the masking. Masking the steep onset slope of $S^{0.5}$ resulted in the gentle onset slope of unmasked portion [Fig. 5(a)]. Conversely, masking the gentle onset slope of S^2 resulted in a steep onset slope of unmasked portion [Fig. 5(c)]. Masking the onset of S^1 did not change the percent /ʃ/ responses [Fig. 5(e)]. This may be explained by fact that masking the onset of S^1 does not change the onset steepness of the unmasked portion [Fig. 5(b)]. Thus, all the findings in experiment 2 are in accord with the suggestion made in the discussion of experiment 1. It is concluded that the amplitude rise slope at onset is an additional acoustic cue responsible for discrimination between the fricative /ʃ/ and the affricate /tʃ/.

IV. EXPERIMENT 3

In experiments 1 and 2, we used linear and exponential functions to implement rise slope patterns and concluded that the amplitude rise slope at onset is a cue for discrimination between the fricative /ʃ/ and the affricate /tʃ/. It is difficult to quantitatively measure the amplitude rise slope at onset generated by exponential functions: the measure of pressure change per unit time of the exponential function is variable depending on the adopted unit of time, and the optimal time unit that the brain adopts is currently unknown. For a quantitative measure of the onset rise slope, we adjusted the rise slopes by linear functions in experiment 3. Naturally uttered fricatives have roughly linear rise slope patterns (Howell and Rosen, 1983). We then tested the relationship between the percent /ʃ/ responses and the steepness of the onset rise slope.

A. Method

1. Stimulus

The following waveform editing procedures were used for all /ʃi/, /ʃu/, and /ʃɑ/ syllables. The equalized sound wave amplitude was multiplied by differently accelerated linear rise functions that were 250 ms in duration. The equation is given by

$$\begin{aligned} f(t) &= 0 && \text{for } 0 \le t < t_0, \\ f(t) &= (t-t_0)/\mathrm{RT} && \text{for } t_0 \le t < \mathrm{RT}+t_0, \\ f(t) &= 1 && \text{for } \mathrm{RT}+t_0 \le t < 250 \text{ ms}, \end{aligned}$$

where t is the time from frication onset, t_0 is the onset time of rise slope, and RT is the rise time duration from t_0 to the time of the maximum amplitude (RT is varied from 30 to 90 ms in steps of 30 ms). The fricative duration (FD) was systematically shifted from 110 to 190 ms in steps of 20 ms (t_0 was correspondingly varied from 140 to 60 ms).

The above-mentioned manipulation resulted in different rates of linear change of sound pressure: $S_{1.4}$ has a gentle rise slope (1.4 Pa/s); $S_{2.0}$ has a medium rise slope (2.0 Pa/s); and $S_{4.2}$ has a steep slope (4.2 Pa/s). The rms of three different linear slopes at FD=150 ms are displayed in Fig. 6(a).

2. Subjects and forced-choice tasks

The same 13 subjects in experiments 1 and 2 participated in experiment 3. The percent /ʃ/ responses were measured by giving 5FDs × 3 slopes × 10 repetitions × 3 vowels=450 trials (random order) to each subject. The listeners' tasks were the same as those in experiments 1 and 2.

B. Result

Because the results were similar in the three vowel conditions of /ʃi/, /ʃu/, and /ʃɑ/, the three different vowel responses were averaged as the percent /ʃ/ responses. The percent /ʃ/ responses (mean and SE of 13 subjects) were plotted against FD [Fig. 6(b)]. It was noted that (1) the percent /ʃ/ responses significantly increased with an increase in FD [a two-way ANOVA, $F(4,48)=128.46$, $P<0.01$], (2) the effects of steepness of the linear slope on the percent /ʃ/ responses were significant [a two-way ANOVA, $F(2,24)=47.16$, $P<0.01$], and (3) there was a significant FD × slope interaction [$F(8,96)=4.85, P<0.01$].

The percent /ʃ/ response differences between $S_{1.4}$, $S_{2.0}$, and $S_{4.2}$ stimuli were significant in all pairs of combinations (Scheffe *posthoc* test: $P<0.05$) except for differences between $S_{2.0}$ and $S_{4.2}$ at FD=130, 170, and 190 ms (Scheffe *posthoc* test: $P>0.05$).

The difference in percent /ʃ/ responses between stimuli $S_{1.4}$ and $S_{4.2}$ was termed as $DS_{1.4-4.2}$. In order to investigate the effect of FD on $DS_{1.4-4.2}$, $DS_{1.4-4.2}$ was plotted against FD [Fig. 6(c)], making the $DS_{1.4-4.2}$-FD function. The function had rising and then falling slopes with increased FD, peaking at 150 ms of FD. The maximum difference (the

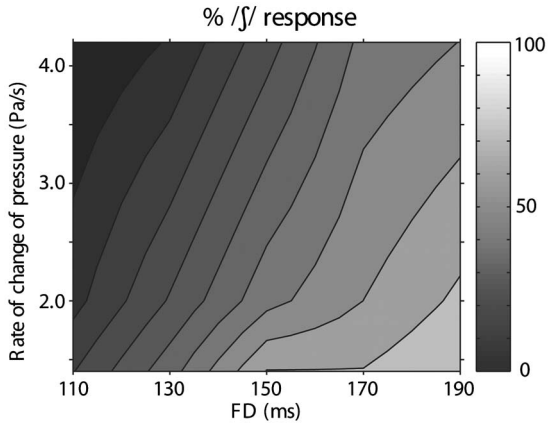


FIG. 7. Experiment 3. Percent /ʃ/ responses, shown by the degree of the brightness, were plotted against FD and amplitude rise slope.

maximum $DS_{1.4\text{-}4.2}$) was significantly different from the differences at other FD values (paired t test: $P<0.01$ at FD =110, 170, 190 ms).

The percent /ʃ/ responses were plotted in the two-dimensional plane with axes of FD and rate of change of sound pressure (Fig. 7). This contour diagram shows that the percent /ʃ/ responses are dependent on the two variables of FD and rate of change of sound pressure in the range of FD from 110 to 190 ms. The percent /ʃ/ responses become higher when the FD is longer and the rate of change of sound pressure is lower.

C. Discussion

The stimuli used in experiment 3 had a linear rate of change of sound pressure for quantification of the amplitude rise slope at onset [Fig. 6(a)]. The linear rate of change of sound pressure affected the distinction between the fricative /ʃ/ and the affricate /tʃ/ [Fig. 6(b)], and the relative importance of linear rate of change of pressure was affected by frication duration [Fig. 6(c)]. The difference in percent /ʃ/ responses between linear rate of change of sound pressure was maximum at 150 ms of FD. By quantification of the amplitude rise slope at onset in experiment 3, the percent /ʃ/ responses could be displayed in the two-dimensional contour plot with axes of FD and rate of change of pressure (Fig. 7), indicating that the identification of /ʃ/ depended upon the interaction of two acoustic cues. To our knowledge, these results are the first to show the relationship between FD and rate of change of sound pressure for distinction between /ʃ/ and /tʃ/ in psychophysical experiments.

V. GENERAL DISCUSSION

The aim of the current study was to clarify the role of two acoustic cues for the distinction between the fricative /ʃ/ and the affricate /tʃ/ in CV syllables. Previous experiments showed that frication duration is one cue for the distinction between /ʃ/ and /tʃ/ (Repp *et al.*, 1978; Dorman *et al.*, 1980; Kluender and Walsh, 1992). The results of our experiments also showed that frication duration is a salient acoustic cue for the distinction between /ʃ/ and /tʃ/ in agreement with previous studies.

In addition, rise time has also been considered as an important acoustic cue for separating /ʃ/ and /tʃ/ (Cutting and Rosner, 1974; Howell and Rosen, 1983). Kluender and Walsh (1992), however, indicated that the rise times were covaried with frication duration in the previous studies, so the effect of the rise time might be confounded by frication duration. They manipulated the rise time and frication duration independently, and suggested that the rise time cue was not sufficient for identification between /ʃ/ and /tʃ/. Their conclusion was also supported by Castleman and Diehl (1996) with the use of a discriminant function analysis.

Kluender and Walsh (1992) have pointed out that amplitude rise time is a less reliable guide to perception under real-world conditions with background noise, because perceptual rise time is quite different from physical rise time when the signal emerges from background noise. In our experiments, the stimuli of experiment 2 would have shorter perceptual rise times because the frication onset was masked in comparison to the unmasked stimuli of experiment 1 [compare Figs. 5(a)–5(c) with Fig. 4(a)]. We showed that the masking the $S^{0.5}$ stimuli [Fig. 5(a)] resulted in an increase of the percent /ʃ/ responses [Fig. 5(d)]. Masking the stimuli S^1 [Fig. 5(b)] did not result in a significant change in the percent of /ʃ/ responses [Fig. 5(e)]. Masking the stimuli S^2 [Fig. 5(c)] resulted in a decrease of the percent /ʃ/ responses [Fig. 5(f)]. Thus, the decrease of the rise time resulted in diverse changes in the percent /ʃ/ responses, unpredictable from the changes in rise time. These findings in experiments 1 and 2 do not suggest that rise time is the major acoustic cue for identification between /ʃ/ and /tʃ/.

In automatic speech recognition, Weigelt *et al.* (1990) proposed that the amplitude rise slope at the initial portion of frication was an acoustic cue for the distinction between fricatives and plosives (including affricates). The algorithm achieved a 96.8% correct finding for the overall rate of distinction. Nevertheless, there were no psychophysical studies investigating the effects of the amplitude rise slope on the distinction between fricatives and affricates up to the present, which motivated us to perform the current psychophysical experiments.

In experiment 1, we presented three amplitude rise slope patterns with rise time held constant and showed that the change of the amplitude rise slope patterns affected the distinction between the fricative /ʃ/ and the affricate /tʃ/ (Fig. 4). Experiment 2 was performed to investigate which portion of rise slope function affected the discrimination. By masking frication onset, we identified that the initial portion of the amplitude rise slope played an important role on the distinction (Fig. 5). In experiment 3, for quantification of the amplitude rise slope, linear rise slopes were used (Fig. 6), and we showed that the percent /ʃ/ responses could be represented in a two-dimensional contour plot with axes of frication duration and rate of change of sound pressure (Fig. 7). It is concluded that the amplitude rise slope at onset is also an acoustic cue for discrimination between the fricative /ʃ/ and the affricate /tʃ/.

Previous physiological studies have shown that the auditory neural system can discriminate between amplitude rise slopes of fricatives and affricates. Delguttle (1980) found

that a quick stimulus rise resulted in greater onset discharge rates than gradual stimulus rise in the auditory nerve fibers. A following study by Delgutte and Kiang (1984) showed that the onset discharge rates were greater for the affricate /tʃ/ stimuli than those for the fricative /ʃ/ stimuli, and suggested that there are the auditory nerve fibers whose response patterns reveal a clear distinction between /ʃ/ and /tʃ/. Furthermore, in the primary auditory cortex, Heil (1997) showed that most cortical neurons are sensitive to the onset slope of the sound envelope but not to the rise time.

We showed that frication duration and amplitude rise slope interact for the distinction between fricatives and affricates [Figs. 4(f), 6(c), and 7]. There are several studies that mentioned the interaction of multiple cues. Repp *et al.* (1978) demonstrated that a short interval of silence between two words "say" and "shop" caused the subjects to hear "say chop," indicating that a silent interval is one manner cue between fricatives and affricates. The shorter the duration of the noise, the shorter the silence necessary to convert the fricative into the affricate. Thus, Repp *et al.* (1978) showed that there is an interaction between two acoustic cues including the silent interval and frication duration.

A similar interaction between multiple acoustic cues was reported in some phonological perception experiments with fricative noises and transitions for distinction between and /s/ and /ʃ/ (Kunisaki and Fujisaki, 1997; Manrique and Massone, 1981; Whalen, 1991; Hedrick and Ohde, 1993; Hedrick and Younger, 2003). Temporal and spectral cues differentiated "slit" and "split" (Fitch *et al.*, 1980). Furthermore, Lisker (1986) suggested that there are at least 16 different acoustic parameters that can be used by a perceiver in making the voicing judgment on intervocalic stops.

Our experiments showed that there are multiple acoustic cues (frication duration and amplitude rise slope) for the distinction between the fricative /ʃ/ and the affricate /tʃ/, which interact with each other. It appears most efficient for the brain to generate phonological perception from multiple acoustic cues rather than a single cue when dealing with acoustic parameters with large overlap among the consonants (Fig. 3).

ACKNOWLEDGMENTS

We thank Professor T. Kumazawa and Professor T. Ando for providing facilities for the research, and N. Yaguchi for technical assistance. The work was supported by a Grant from the Ministry of the Education, Science, Culture, Sports and Technology, Japan.

Behrens, S., and Blumstein, S. E. (**1988**). "On the role of the amplitude of the fricative noise in the perception of place of articulation in the voiceless fricative consonant," J. Acoust. Soc. Am. **84**, 861–867.

Castleman, W. A., and Diehl, R. L. (**1996**). "Acoustic correlate of fricatives and affricates," J. Acoust. Soc. Am. **99**, 2546(A).

Cutting, J. E., and Rosner, B. S. (**1974**). "Categories and boundaries in speech and music," Percept. Psychophys. **16**, 564–570.

Delgutte, B. (**1980**). "Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers," J. Acoust. Soc. Am. **68**, 843–857.

Delgutte, B., and Kiang, N. Y. S. (**1984**). "Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics," J. Acoust. Soc. Am. **75**, 897–907.

Dorman, M. F., Raphael, L. J., and Isenberg, D. (**1980**). "Acoustic cues for a fricative-affricate contrast in word-final position," J. Phonetics **8**, 397–405.

Fitch, H. L., Halwes, T., Erickson, D. M., and Liberman, A. M. (**1986**). "Perceptual equivalence of two acoustic cues for stop-consonant manner," Percept. Psychophys. **27**, 343–350.

Furui, S. (**1986**). "On the spectral transition for speech perception," J. Acoust. Soc. Am. **80**, 1016–1025.

Grim, W. A. (**1966**). "Perception of segments of English-spoken consonant-vowel syllables," J. Acoust. Soc. Am. **40**, 1454–1461.

Hedrick, M. S. (**1997**). "Effect of acoustic cues on labeling fricatives and affricates," J. Speech Lang. Hear. Res. **40**, 925–938.

Hedrick, M. S., and Ohde, R. N. (**1993**). "Effect of relative amplitude of frication perception of place of articulation," J. Acoust. Soc. Am. **94**, 2005–2026.

Hedrick, M. S., and Younger, M. S. (**2003**). "Labeling of /s/ and /ʃ/ by listeners with normal and impaired hearing, revisited," J. Speech Lang. Hear. Res. **46**, 636–648.

Heil, P. (**1997**). "Auditoy cortical onset responses revisited. II. Response strength," J. Neurophysiol. **77**, 2642–2660.

Howell, P., and Rosen, S. (**1983**). "Production and perception of rise time in the voiceless affricate/fricative distinction," J. Acoust. Soc. Am. **73**, 976–984.

Jongman, A. (**1989**). "Duration of frication noise required for identification of English," J. Acoust. Soc. Am. **85**, 1718–1725.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.

Kluender, K. R., and Walsh, M. A. (**1992**). "Amplitude rise time and the perception of the voiceless affricate/fricative distinction," Percept. Psychophys. **51**, 328–333.

Kunisaki, O., and Fujisaki, H. (**1977**). "On the influence of context upon perception of voiceless fricative consonants," Annu. Bull. Res. Inst. Logopedics Phoniatrics **11**, 85–91.

Lisker, L. (**1986**). "'Voicing' in English: a catalogue of acoustic features signaling /b/ versus /p/ in trochees," Lang Speech **29**, 3–11.

Manrique, A. M. B., and Massone, M. I. (**1981**). "Acoustic analysis and perception of Spanish fricative consonants," J. Acoust. Soc. Am. **69**, 1145–1153.

Repp, B. H., Liberman, A. M., Eccardt, T., and Pesetsky, D. (**1978**). "Perceptual integration of acoustic cues for stop, fricative, and affricate manner," J. Exp. Psychol. **4**, 621–637.

Weigelt, L. F., Sadoff, S. J., and Miller, J. D. (**1990**). "Plosive/fricative distinction: the voiceless case," J. Acoust. Soc. Am. **87**, 2729–2737.

Whalen, D. H. (**1991**). "Perception of the English /s/-/ʃ/ distinction relies on fricative noises and transitions, not on brief spectral slices," J. Acoust. Soc. Am. **90**, 1776–1785.

Measurements and calculations on the simple up-down adaptive procedure for speech-in-noise tests

Cas Smits[a] and Tammo Houtgast
Department of Otolaryngology/Audiology, VU University Medical Center, Amsterdam, The Netherlands

(Received 22 March 2005; revised 21 April 2006; accepted 4 June 2006)

The simple up-down adaptive procedure is a common method for measuring speech reception thresholds. It is used by the Dutch speech-in-noise telephone screening test [National Hearing test; Smits and Houtgast Ear Hear. **26**, 89–95 (2005)]. The test uses digit triplets to measure the speech reception threshold in noise by telephone ($SRTT_n$). About 66 000 people took this test within four months of its introduction and details were stored of all individual measurements. Analyses of this large volume of data have revealed that the standard deviation of $SRTT_n$ estimates increases with hearing loss. This paper presents a calculation model which—using an intelligibility function as input—can determine the standard deviation of $SRTT_n$ estimates and the bias for the simple up-down procedure. The effects of variations in the slope of the intelligibility function, the guess rate, the starting level, the heterogeneity of the speech material, and the possibilities of optimizing $SRTT_n$ measurements were all explored with this model. The predicted decrease in the standard deviation of $SRTT_n$ estimates as a result of optimizing the speech material was confirmed by measurements in 244 listeners. The paper concludes by discussing possibilities for optimizing the development of comparable tests. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2221405]

PACS number(s): 43.71.Gv [DOS] Pages: 1608–1621

I. INTRODUCTION

The simple up-down adaptive procedure is applied in both clinical audiology and research programs. It is frequently used in speech-in-noise measurements to determine the ability to understand speech in noise. Often, it determines the speech reception threshold in noise (SRT_n), i.e., the signal-to-noise ratio that corresponds to 50% intelligibility. Although this procedure has been in use for a long time, it is still not fully understood how far the accuracy of the SRT_n estimate is effected by various factors. Perhaps this is partly due to the fact that a great many experiments would be needed in order to reduce uncertainties. The accuracy of the SRT_n estimate depends on several factors: first, the shape of the underlying intelligibility function (e.g., the slope of the function, lapse rate, guess rate); second, the characteristics of the measurement method (e.g., adaptive or fixed levels, starting level, step size, etc.); third, the number of presentations; and fourth, the calculation method (e.g., averaging presentation levels, maximum-likelihood fit, etc.). These four factors affect the standard deviation of the SRT_n estimates (precision) and can lead to bias (difference between the signal-to-noise ratio that corresponds to 50% intelligibility and the mean SRT_n estimate).

Adaptive psychophysical procedures have many advantages over fixed-level procedures and are widely used. They can be split into three general categories (Leek, 2001), viz. PEST procedures (parameter estimation by sequential testing; Taylor and Creelman, 1967), maximum-likelihood procedures, and staircase (simple up-down) procedures. Fixed-level and adaptive procedures are both used regularly in speech-in-noise experiments. The most common adaptive procedure in these experiments is the simple up-down method. Brand and Kollmeier (2002) propose an adaptive procedure with a decreasing step size in which each presentation level is based on the discrimination value obtained in the previous sentence. The SRT_n is calculated by applying a maximum-likelihood fit to the data. The effects of heterogeneity of stimuli and inattentiveness have been investigated for some procedures. For example, Green (1995) performed computer simulations and found that inattentiveness can generate a strong bias in the threshold estimate when using a maximum-likelihood procedure. Green (1990) also studied the effect of a mismatch between the assumed intelligibility function and the true intelligibility function. The latter is not relevant in staircase procedures because the only assumption for the underlying intelligibility function is that it increases monotonically.

Plomp and Mimpen (1979) developed an adaptive speech-in-noise test that uses 13 sentences per list. Later, a similar test, the HINT, was developed in the USA by Nilsson *et al.* (1994). Plomp and Mimpen's test has figured in numerous studies. For example, it was used by Festen and Plomp (1990) to examine the effect of fluctuating noise as opposed to stationary noise, and by Lyzenga *et al.* (2002) in studies on speech enhancement. The value of a speech-in-noise test depends mainly on its ability to detect differences between subjects or conditions (e.g., by using different hearing aids). Because the results of speech-in-noise experiments are not usually comparable, the absolute value of the test result is of lesser importance.

In 2004, Smits *et al.* developed an automatic telephone speech-in-noise screening test, similar to the sentence

[a]Author to whom correspondence should be addressed; electronic mail: c.smits@vumc.nl

speech-in-noise test of Plomp and Mimpen (1979). The aim was twofold: to meet the need for a functional self-test and to enhance public awareness of hearing loss. The test uses digit triplets to measure the speech reception threshold in noise by telephone ($SRTT_n$). Further details on the development, validation, and implementation of the test can be found in Smits *et al.* (2004) and Smits and Houtgast (2005). Briefly, digit triplets were uttered in Dutch by a trained female speaker and digitally recorded. Only monosyllabic digits were used: 0, 1, 2, 3, 4, 5, 6, 8 (/nɣl/, /en/, /twe/, /dri/, /vir/, /vɛif/, /zɛs/, /ɑxt/). Masking noise was constructed with a spectral shape similar to the mean spectra of the triplets. The intelligibility of the triplets was homogenized by applying level corrections. The final set consisted of 80 different triplets. Experiments revealed no significant differences in $SRTT_n$ between the telephones used. A validation study with normal-hearing and hearing-impaired listeners ($SRTT_n$s ranging from −9 to +4 dB) showed a correlation between the triplet $SRTT_n$ telephone test and the standard Dutch sentence test (Plomp and Mimpen, 1979) of 0.87. After correction for measurement error the actual correlation coefficient worked out at approximately 0.94, suggesting that the triplet $SRTT_n$ telephone test can be used to screen hearing disability. The test measures the $SRTT_n$ by applying an up-down procedure: the signal-to-noise ratio of a presentation increases by 2 dB after an incorrect response and decreases by 2 dB after a correct response. A fixed starting level is used. The test is implemented on an interactive voice-response system and is fully automatic. Forty parallel lines are available. The subject responds by pressing the telephone keys. A response qualifies as correct only when all three digits are correctly understood. A series of 23 triplets is chosen at random from the set of 80 triplets for each $SRTT_n$ measurement. The $SRTT_n$ is taken to be the average signal-to-noise ratio of the last 20 presentations (in which the signal-to-noise ratio based on the last response is not actually used in the test). The test was introduced as the National Hearing test on 1 January 2003. Publicity was generated and, in the first four months, the test was taken by 65 924 individuals. Exclusion criteria were applied with a view to further statistical analysis of the data: more than three instances of no-response, an incorrect response at the maximum signal-to-noise ratio of +8 dB, and the use of a mobile (cellular) or unknown type of telephone. The results were reported for the remaining 39 968 respondents (Smits and Houtgast, 2005). Detailed data were attained from all measurements, resulting in almost 40 000 $SRTT_n$s, around 800 000 triplet presentations (different signal-to-noise ratios and scores) and around 2 400 000 digit presentations.

Most parameters in the National Hearing test were adopted from the standard Dutch sentence SRT_n test. The large volume of data enables a thorough investigation of the test material and procedure, which is valuable for groups who are developing comparable tests in other languages. The aim of this study is to find out more about the four above-mentioned factors in the simple up-down adaptive procedure in speech-in-noise measurements and to quantify their contribution to measurement accuracy. The large number of $SRTT_n$ measurements enabled us to perform a detailed analysis and thereby identify properties of the intelligibility functions. The staircase procedure was analytically described by means of a calculation model in which the input parameters were step size, starting level, and an intelligibility function describing the relation between signal-to-noise ratio and performance. First, the effects of the procedure and the different properties of the intelligibility function on measurement accuracy were examined (slope, guessing, heterogeneity of the speech material, starting level). Next, the model was used in combination with data from the National Hearing test to explore the scope for optimizing the speech material and the measurement procedure of the National Hearing test. Experiments were performed to compare the optimized speech material with the original speech material. The last section of this paper discusses the results and sets out some general conclusions.

II. THE INTELLIGIBILITY FUNCTION

A. Basic concepts

The intelligibility function relates the physical intensity of a stimulus to the intelligibility of a stimulus in an intelligibility task. Intelligibility is expressed as the probability of a correct response. If the psychophysical task is a speech-in-noise test, the physical intensity is, in most cases, the signal-to-noise ratio. The performance can be, for instance, the percentage of sentences or words that meet with a correct response. Normally, performance increases monotonically with stimulus intensity. The intelligibility function may be written as

$$P(x) = \gamma + (1 - \gamma - \lambda)\Phi(x) \tag{1}$$

in which γ is the lower asymptote (or guess rate) and $1-\lambda$ is the upper asymptote of the function. The lapse rate (or miss rate), λ, reflects the rate at which incorrect responses are given regardless of the signal level. Ideally, the lapse rate is zero, but it has a nonzero value in most psychophysical experiments as a result of, amongst others, inattentive subjects. In forced-choice methods, the guess rate is simply related to the number of alternatives $(1/n)$. In speech-in-noise experiments it depends on the type of speech material and will effectively range from zero for open-set speech material to values related to the number of items in a closed set. $\Phi(x)$ can be any arbitrary S-shaped function between 0 and 1. Standard functions such as the logistic, Weibull, arctangent, and cumulative normal distribution can be used. In this study we used the cumulative normal distribution:

$$\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} \exp\left(-\frac{(\zeta - x_0)^2}{2\sigma^2}\right)d\zeta \tag{2}$$

in which $\Phi(x)=0.5$ at $x=x_0$, and the slope S (in dB^{-1}, when x represents signal-to-noise ratio) at $x=x_0$ can be derived from σ by

$$S = \frac{1}{\sigma\sqrt{2\pi}}. \tag{3}$$

It is important to realize that S represents the maximum slope of the cumulative normal distribution. The maximum slope of $P(x)$, the intelligibility function, is also found at x

$=x_0$ but, for an intelligibility function reduced by guess rate and lapse rate, the maximum slope at that point equals $(1-\gamma-\lambda)/\sigma\sqrt{2\pi}$ and $P(x_0)=0.5+0.5\gamma-0.5\lambda$. The point of 50% intelligibility $[P(x)=0.5]$ can be found via the inverse cumulative normal distribution and will be smaller than x_0 when $\gamma>\lambda$ and higher than x_0 when $\gamma<\lambda$. The slope at this point will be somewhat smaller than the maximum slope.

To avoid ambiguities we shall explicitly define some key concepts. $SRTT_n$ is the signal-to-noise ratio where intelligibility is 50%. The result of a particular experiment is often simply presented as the $SRTT_n$ where, in reality, it is only an estimate of the true $SRTT_n$. Whether the convergence point equals the true $SRTT_n$ depends on the measurement method and, in some experiments, on implicit assumptions about the shape of the intelligibility function. We therefore draw a sharp distinction between the *true* $SRTT_n$ and the *measured* $SRTT_n$ (or $SRTT_n$ *estimate*). The true $SRTT_n$, or target value, is the signal-to-noise ratio that corresponds to 50% intelligibility, while the measured $SRTT_n$ is the result of a measurement procedure. The measured $SRTT_n$ is prone to systematic and random errors. A systematic error (i.e., the difference between the mean measured $SRTT_n$ and the true $SRTT_n$) is called a *bias*. A random error denotes the imprecision of the measurement and is expressed as the standard deviation of $SRTT_n$ estimates.

It should be noted that the intelligibility function in speech-in-noise tests can be defined in different ways. First, it may represent the performance of a single observer, in which case the term "psychometric function" is frequently used. Second, it may represent the intelligibility of an item (word, sentence, digit, etc.) as a function of the signal-to-noise ratio. These intelligibility functions are often determined in order to create a homogeneous set of items for a test. Third, it may represent the mean performance for a group of listeners. When the intelligibility function is being determined for a group of listeners, the data of individual listeners are often shifted in order to align thresholds (i.e., align $SRTT_n$ estimates). In other words the data are corrected for interindividual differences in true $SRTT_n$ by use of the $SRTT_n$ estimate. So, the correction (or shift) is actually the sum of the measurement error and the true $SRTT_n$, whereas it should be limited to the true $SRTT_n$. In most cases, this procedure is not entirely correct and will result in a slope bias: the estimated slope will be higher than the true slope, unless the measurement error is zero. Also estimates of the guess rate and lapse rate are unreliable. The error arising from this procedure can be illustrated by a simple example. Suppose several subjects have exactly the same ability for understanding speech in noise (identical intelligibility functions). When performing $SRTT_n$ measurements, this should, ideally, give the same $SRTT_n$ value for each subject. However, due to the measurement error, some spread will be found around the mean $SRTT_n$. The intelligibility function that is determined after applying corrections for interindividual $SRTT_n$ differences will be steeper than the true intelligibility function which, in this hypothetical case, should be determined without applying corrections.

The intelligibility functions in the present study represent the intelligibility of an item (digit or triplet) or the mean performance for groups of listeners. Most intelligibility functions are determined after correction for interindividual differences in $SRTT_n$ by use of the $SRTT_n$ estimate. Consequently, the observed slopes of these functions are greater than the true (underlying) slopes. It may, however, be assumed that noted qualitative differences between intelligibility functions are still valid. This topic is further addressed in Sec. V A.

Although cumulative normal distribution often adequately describe results of speech-in-noise measurements, one should not forget that it is an approximation of the true underlying intelligibility function. This is reflected in the scoring method for the National Hearing test where the speech material consists of triplets of digits. Unlike words in a meaningful sentence, the digits in a triplet can be considered independent. The intelligibility function of a triplet can be described in two ways: first, by a single intelligibility function [triplet-intelligibility function, Eq. (1)], and second, by multiplication of three intelligibility functions that represent the three digits (product-intelligibility function):

$$P_{\text{triplet}}(x) = P_{\text{digit1}}(x) \cdot P_{\text{digit2}}(x) \cdot P_{\text{digit3}}(x). \tag{4}$$

For P_{digit1}, P_{digit2}, and P_{digit3} the values of γ, λ, σ, and x_0 must be determined separately. As the mathematical product of two or more cumulative normal distributions is not, in itself, a cumulative normal distribution, the intelligibility functions must be regarded as an approximation.

In this study intelligibility functions were determined by performing maximum-likelihood fits for the data. The only restriction on the parameters was that γ and λ were between 0 and 0.5.

B. Analysis of measurements: Effect of age and hearing loss

Data from the National Hearing test were analyzed to identify the properties of the intelligibility function. First, the mean intelligibility function was established for all subjects: the signal-to-noise ratios of the triplet presentations were corrected for interindividual differences in $SRTT_n$ (e.g., Smits *et al.*, 2004; Smoorenburg, 1992) for each individual by expressing the presentation levels relative to the measured $SRTT_n$. A maximum-likelihood fit performed on the data resulted in an intelligibility function with slope $S=0.158\ \text{dB}^{-1}$, guess rate $\gamma=0.029$, lapse rate $\lambda=0.043$, and $x_0=-0.29$ dB. The Pearson X^2 test was applied to assess goodness-of-fit and a near-perfect fit was found ($p \ll 0.001$).[1] The resulting function is shown in Fig. 1 along with the original data. The original data, represented by dots, show the percentage of correct responses for levels with at least 150 presentations. The fitting result yielded some important parameters. The guess rate, γ, equaled 0.029 and was higher than anticipated. When every digit has the same intelligibility at a given signal-to-noise ratio, the anticipated guess rate is approximately 0.001. This high value may be largely due to unreliable parameter estimates from intelligibility functions representing data which have been corrected for interindividual differences in $SRTT_n$ (Sec. II A), and to the fact that the parameters are meaningful only for the range of used signal-to-noise ratios.[1] The lapse rate, λ, was 0.043. Again,

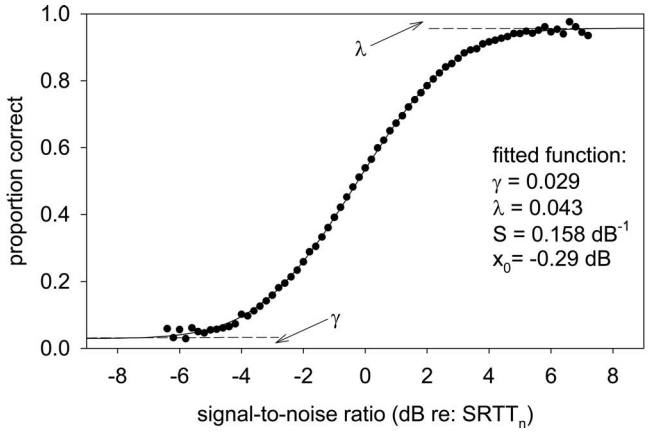


FIG. 1. Intelligibility score as a function of presentation level relative to the individual $SRTT_n$, averaged over 759 392 presentations (39 968 measurements). Data from the National Hearing test. The solid line represents the result of a maximum-likelihood fit to the data.

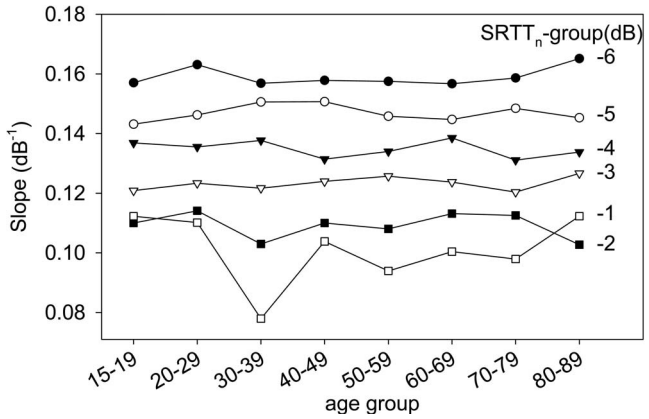


FIG. 2. Slope of the intelligibility function as a function of age and $SRTT_n$. For the purposes of clarity only data points that are based on at least 20 $SRTT_n$ measurements and $SRTT_n$ groups that have data points over the entire age range are shown.

because of the procedure, this value does not represent the true lapse rate for individuals. However, a lapse rate of more than 0 was expected because the subjects could not correct their response if they accidentally pressed the wrong key. The last parameter in the fitting function, x_0, was −0.29 dB. At first glance, this value comes across as something of a surprise, given that the signal-to-noise ratios of the presentation levels were corrected with the individual $SRTT_n$ estimates and, therefore, an $x_0$value of 0 dB could be expected. This discrepancy may be explained first by the fact that the $SRTT_n$ was calculated by averaging over 20 presentation levels. The last level was not, however, presented to the subject and no response was obtained, so it could not be included in the maximum-likelihood fitting procedure. As there is a (small) learning effect (Smits and Houtgast, 2005), the average value of the last presentation would be lower than the $SRTT_n$, causing a systematic shift. The second explanation lies in the asymmetric shape of the intelligibility function ($\gamma \neq \lambda$) and the third in the chosen starting level, which was about 4.6 dB higher than the average $SRTT_n$. Using the calculation model (see Sec. III) it was estimated that these three explanations account for <0.01, 0.06, and 0.13 dB, respectively. The effect of the starting level should be regarded as a rough approximation, given the broad range of $SRTT_n$s. The discrepancy may also be partly attributable to the fact that the fitted intelligibility function was only an approximation of the true intelligibility function. For instance, $\Phi(x)$ may be asymmetric or the intelligibility functions may differ for each subject.

As reported by Smits and Houtgast (2005), the greater the hearing loss, the higher the standard deviation of $SRTT_n$ estimates. Given that hearing deteriorates with age, the relationship that emerged could be due to a higher average age of subjects with higher $SRTT_n$s. To explore this aspect further, groups were created for a grid of ages and $SRTT_n$ values. The intelligibility function for each group was approximated with a cumulative normal distribution ($\gamma=\lambda=0$).[2] The results are shown in Fig. 2. Note that, because the data were corrected for interindividual differences in $SRTT_n$, the estimated slope values are greater than the true slope values (Sec. II A). Although linear regression lines reveal slopes that deviate significantly from zero (between −0.004 and +0.01 dB^{-1}/yr), it may be concluded that the decrease in S and, consequently, the increase in the standard deviation of $SRTT_n$ estimates with increasing $SRTT_n$,, were caused by hearing loss rather than by age. Results from an ANOVA showed that 97% of the variance in slopes can be explained by $SRTT_n$ values.

III. CALCULATION MODEL

A. Description of the model

The simple adaptive up-down method with a step size of 2 dB and 13 presentations is very common in speech-in-noise measurements and has figured in many different experiments since it was first proposed by Plomp and Mimpen (1979). The SRT_n is calculated by averaging the last 10 presentation levels (the last level is not actually presented to the subject). Direct calculation of the standard deviation of $SRTT_n$ estimates is not possible. Monte Carlo simulations are often used to explore the relationship between the intelligibility function and the accuracy of the SRT_n (e.g., Green, 1990; Brand and Kollmeier, 2002). However, we chose a more direct and exact calculation method, similar to the one used by Kollmeier *et al.* (1988). Although the National Hearing test consists of 23 presentations, we opted to start with a calculation model based on Plomp and Mimpen's method (1979), which comprises 13 presentations. We limited the presentations to 13, first because it was supposed that measurement error decreases by $1/\sqrt{n}$ for large n, making it easy to predict the properties of tests that use more presentations. A second—but more important—reason was the exponential increase in the number of calculations with increasing n (there are $2^{23}=8\,388\,608$ execution possibilities with 23 presentations). A fixed starting level (as in the National Hearing test) is assumed in the model. Any intelligibility function can be used as input (the calculation model is illustrated in Fig. 3). The first presentation is at the starting level; the response can be either correct or incorrect. The probability of a correct response is derived from the intelligibility function. The second presentation is at the starting level plus or minus 2 dB. The probability of a correct response to the second presenta-

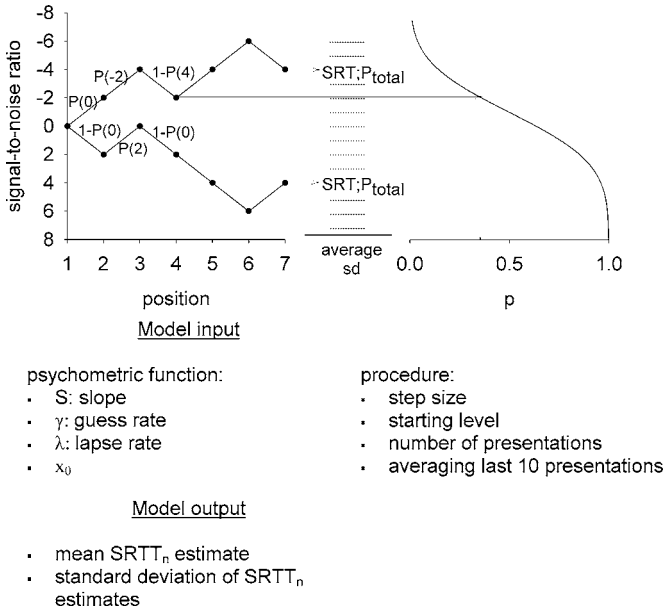


FIG. 3. Schematic presentation of the calculation model. The first presentation, represented by the outermost left dot, is at a signal-to-noise ratio of 0 dB (starting level). A track is followed depending on correct or incorrect responses. Two of the 2^{13} tracks are shown. The probability of a correct or incorrect response at any signal-to-noise ratio can be derived from the intelligibility function shown on the right. Each track results in a $SRTT_n$ and a probability. The mean $SRTT_n$ estimate and standard deviation of $SRTT_n$ estimates are determined from the model. The input parameters and the output of the model are summarized in the lower panel.

tion is again derived from the intelligibility function. Obviously, as this probability depends on the level, two different results are obtained. This procedure is repeated for the next presentations and it results in different tracks. The $SRTT_n$ is calculated for each track by averaging the last 10 presentation levels. The associated probability is calculated by multiplying the different probabilities in the track. A total of 13 presentations were used with two possibilities per presentation (correct or incorrect), yielding $2^{13}=8192$ different tracks. Not every track gives a different $SRTT_n$ or probability. The weighted mean and weighted standard deviation are calculated from the 8192 $SRTT_n$s and probabilities. Ideally, the weighted mean $SRTT_n$ (i.e., mean $SRTT_n$ estimate) corresponds to the 50% intelligibility point (true $SRTT_n$), or can include bias. The imprecision is represented by the spread in $SRTT_n$s, i.e., the standard deviation of $SRTT_n$ estimates.

B. Effect of changes in the intelligibility function on the standard deviation of $SRTT_n$ estimates: Model calculations

The calculation model makes it possible to investigate the relationships between, e.g., guess rate or starting level and bias or standard deviation of $SRTT_n$ estimates. The variables can be represented by changes in the intelligibility function and are thereby included in the calculation model. Parameters were set within a range that can be considered realistic, given the intelligibility function for the speech material of the National Hearing test (Figs. 1 and 2).

First, the effect of the slope of the intelligibility function on the standard deviation of $SRTT_n$ estimates was calculated.

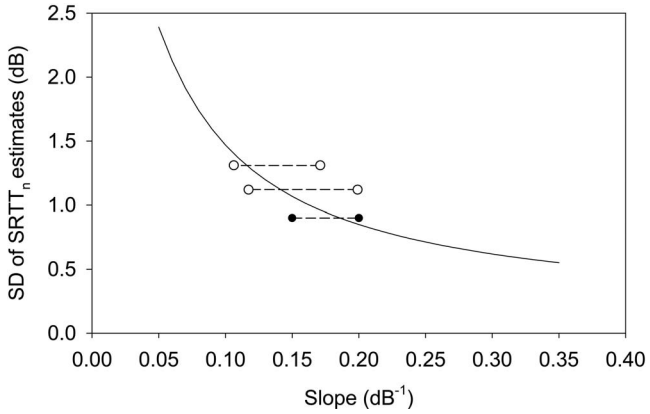


FIG. 4. The effect of the slope of the intelligibility function (cumulative normal distribution) on the model-predicted standard deviation of $SRTT_n$ estimates for a simple up-down adaptive procedure with a step size of 2 dB and a total of 13 presentations. (See the text for an explanation of the dots.)

Intelligibility functions were represented by simple cumulative normal distributions. Guess rate and lapse rate were set at zero. These intelligibility functions were used as input for the calculation model. The step size was 2 dB and the starting level was at the point of 50% intelligibility. As displayed in Fig. 4, the results indicate an approximately inversely proportional relationship between slope and standard deviation of $SRTT_n$ estimates. It is particularly noticeable in the case of speech material with relatively shallow intelligibility functions that an increase in slope results in a relatively strong decrease in the standard deviation of $SRTT_n$ estimates. For the intelligibility function derived from data from their adaptive sentence SRT_n test comprising 13 presentations, Plomp and Mimpen (1979) found a slope of 0.20 dB^{-1} and, without correction for interindividual differences in SRT_n, a slope of 0.15 dB^{-1}. They reported a standard deviation of SRT_n estimates of 0.9 dB based on test-retest measurements. These results are plotted in Fig. 4 as two solid dots joined by a dashed line. It may be assumed that the slope of the true intelligibility function is between 0.15 and 0.20 dB^{-1}, which is in accordance with the result from the calculation model.

Second, the effect of guess rate (γ) on the standard deviation of $SRTT_n$ estimates was investigated. Lapse rate (λ) was fixed at 0.04. The standard deviation of $SRTT_n$ estimates and mean $SRTT_n$ estimate were calculated for different values of γ for three different slopes, S (0.10, 0.14, and 0.18 dB^{-1}). As shown in the upper panel of Fig. 5, the standard deviation of $SRTT_n$ estimates increased with an increasing γ value. The middle panel shows the bias: this deviates from zero for higher values of γ but equals zero for $\gamma=0.04$, because the intelligibility function is then symmetric. In speech-in-noise measurements high γ values are found for speech material from a closed set of a few items. A value of about 0.1 was expected for a speech-in-noise test using single digits (10 different items). The lower panel of Fig. 5 shows that the intelligibility percentage corresponding to the average measured $SRTT_n$ deviates only slightly from 0.5.

Third, the effect of heterogeneity of the speech material was studied. The model assumed that all the presentations had the same intelligibility function: a cumulative normal distribution with equal slope and $\gamma=\lambda=0$. However, the x_0

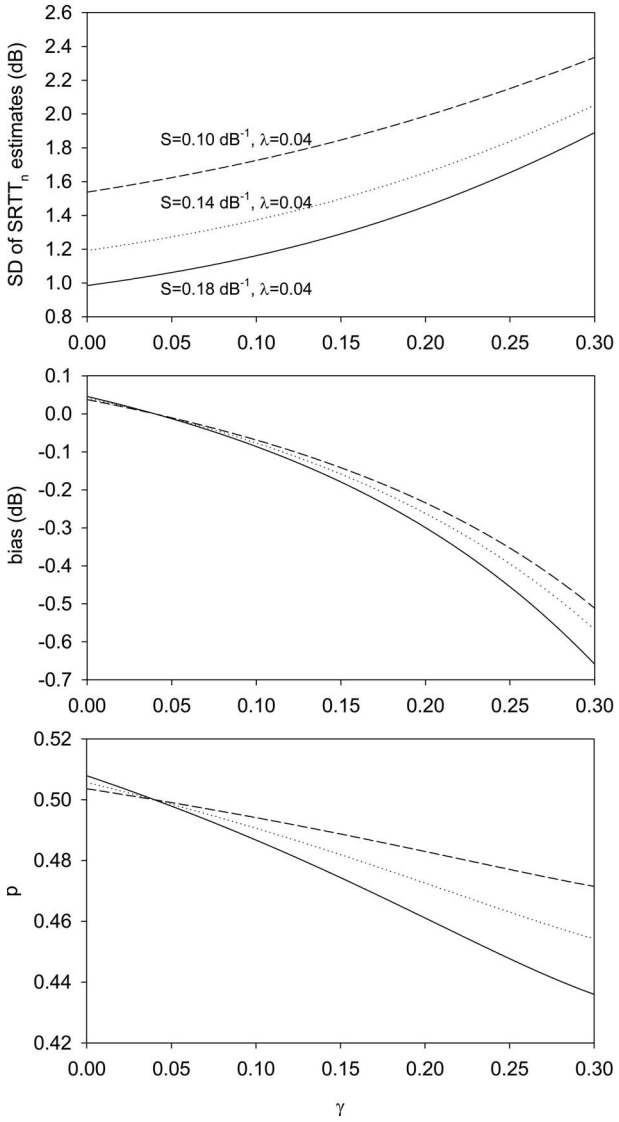


FIG. 5. The effect of guess rate (γ) on model-predicted standard deviation of $SRTT_n$ estimates (top panel), bias (middle panel), and corresponding intelligibility percentage (lowest panel) for a simple up-down adaptive procedure with a step size of 2 dB and a total of 13 presentations. The lapse rate (λ) was fixed at 0.04. The results are shown for three different slopes of the intelligibility function.

values did not coincide exactly but followed a normal distribution. Figure 6 shows the standard deviation of $SRTT_n$ estimates against the standard deviation of the normal distribution, for different slopes. As the result of the calculation model was dependent on the distribution of the different intelligibility functions over the presentations, the average result for 40 calculations is given. Mean values were zero, i.e., no bias, because the intelligibility functions were symmetric. Homogeneity of the speech material, though not exactly crucial, proved more important for intelligibility functions with steeper slopes.

Fourth, the effect of the starting level was explored. A fixed starting level (as in the National Hearing test; Smits and Houtgast, 2005) can affect the $SRTT_n$ estimate, because several presentations are needed to reach the level of approximately 50% intelligibility. The most important parameter is the difference between the starting level and the

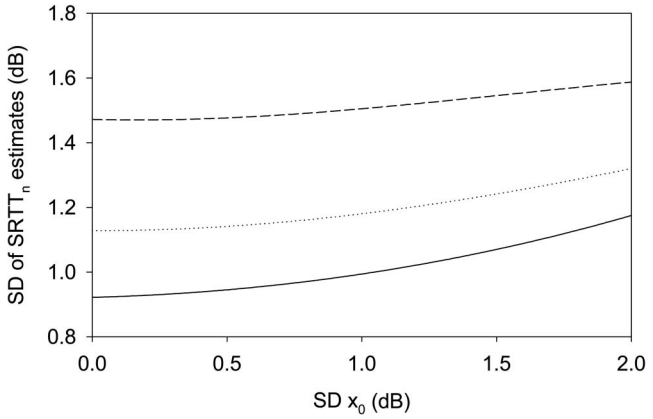


FIG. 6. Model-predicted standard deviation of $SRTT_n$ estimates vs standard deviation of the 50% intelligibility points of the intelligibility functions (i.e., heterogeneity of the speech material). The results relate to intelligibility functions with slopes of 0.10 dB^{-1} (dashed line), 0.14 dB^{-1} (dotted line), and 0.18 dB^{-1} (solid line), using a simple up-down adaptive procedure with a step size of 2 dB and a total of 13 presentations.

$SRTT_n$. If this difference is very large, the fifth presentation, which is the first to be used in calculating the $SRTT_n$, will still not be in the region of the $SRTT_n$. Figure 7 shows the (weighted) average of the signal-to-noise ratios for the different positions in the procedure. The results refer to an intelligibility function with a slope of 0.14 dB^{-1} and to starting levels relative to the $SRTT_n$ from 0 to 10 dB. Bias is shown on the right. This is calculated by averaging the last 10 presentation levels. The effect on the $SRTT_n$ was found to be very small ($<$0.1 dB) for starting levels of less than 5 dB from the $SRTT_n$ and the effect of the starting level turned out to be negligible for positions higher than 10 ($<$0.1 dB for a starting level 10 dB higher than $SRTT_n$). Note that the bias will decline when the number of presentations is increased to 23, as in the National Hearing test.

IV. INCREASING THE ACCURACY OF $SRTT_n$ MEASUREMENTS

As mentioned earlier, the accuracy of a $SRTT_n$ measurement depends on different factors. The calculation model

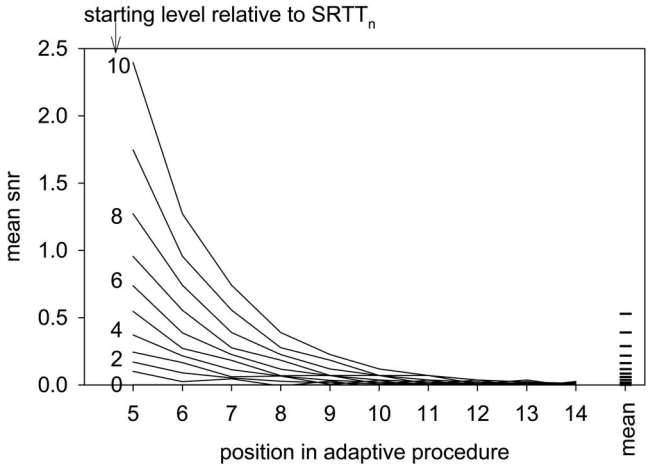


FIG. 7. Weighted average of the signal-to-noise ratios of the presentations for the different positions in the procedure. The results are shown for different starting levels relative to the $SRTT_n$. The mean values on the right represent bias.

makes it possible to examine and optimize these factors. In the following, factors that might be relevant to the National Hearing test are explored.

A. Adjusting the speech material

1. Homogenizing the triplets

Homogeneous test material is important in psychophysical experiments (Fig. 6). In many experiments, such as tone detection, one stimulus is sufficient and only the intensity is changed. Speech intelligibility experiments, on the other hand, require different stimuli. The term homogeneity (and heterogeneity) is used here to indicate equality of the signal-to-noise ratios associated with the target point (50% intelligibility for symmetric intelligibility functions). Homogeneity does not therefore mean equality of the steepness of the triplets intelligibility functions. Homogeneity was achieved for the triplets in the National Hearing test by applying level corrections to individual triplets (Smits *et al.*, 2004). As much more data are now available, these corrections could be refined. After correction for interindividual differences in $SRTT_n$, the intelligibility function was determined for each triplet in the total of 80 by fitting the data (about 9255 data points per triplet). To detect any possible interaction between the amount of hearing loss and heterogeneity of the speech material, the same procedure was performed separately on data from two $SRTT_n$ groups with an interval width of 1 dB. $SRTT_n$ groups −7 dB (typical normal hearing) and −4 dB (mild hearing loss) were used (about 1034 and 1660 data points per triplet, respectively). The parameters of the intelligibility functions were used as input for the calculation model and the modeled mean $SRTT_n$ was calculated for each triplet. These values represent the refined level corrections that should be applied to create "truly" homogeneous triplets. The standard deviation of these values around the mean was 1.23, 1.14, and 1.32 dB for the group that included all the measurements, the $SRTT_n$ group of −7 dB and the $SRTT_n$ group of −4 dB, respectively. The correlation coefficients, over all triplets, between the level corrections derived from the group that included all the measurements and the $SRTT_n$ groups of −7 dB and −4 dB were 0.93 and 0.99, respectively. As shown in Fig. 6, the refined level corrections lead only to a slight decrease in the standard deviation of $SRTT_n$ estimates. Moreover, level corrections derived from measurements for listeners with impaired hearing and normal hearing are nearly the same, implying that the decrease in the standard deviation of $SRTT_n$ estimates with increasing $SRTT_n$ is not due to the heterogeneity of the speech material.

2. Optimizing the intelligibility functions for individual triplets

As each triplet consists of three digits, the intelligibility function of the triplet is determined by the intelligibility of the digits separately and in relation to each other [Eq. (4)]. The slope for the triplet can be changed by raising or lowering the level of the individual digits. It should be noted that, in most cases, the optimal intelligibility function for the triplet is not reached for the situation in which the x_0 values of the digits coincide, even when γ and λ are equal for each digit. This would occur only if the slopes of the intelligibility functions were the same for each digit. When, for instance, the slope of the intelligibility function of one digit is much steeper than those of the other two digits, the optimal intelligibility function of the triplet will be reached when the two digits with the shallow slopes are always correctly understood. In that case, the intelligibility function of the triplet equals the intelligibility function of the digit with the steep slope. Needless to say, changing the intelligibility of the digits by, for example, making one digit easy to understand can influence the guess/lapse rate of the triplet. This is taken into account by the calculation model. Essentially, the output of the calculation model (standard deviation of $SRTT_n$ estimates) was minimized by changing the input (intelligibility of the individual digits, represented by the product-intelligibility function). Three steps were taken to optimize the intelligibility functions of the triplets.

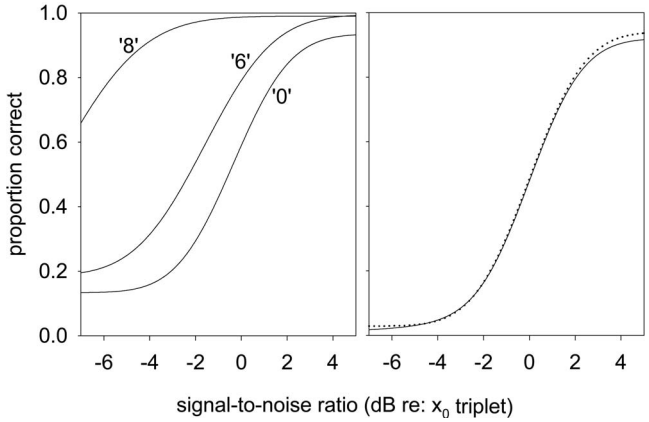


FIG. 8. Left panel: individual intelligibility functions of the three digits that form the triplet 0-6-8. Right panel: product of the three intelligibility functions of the digits forming the product-intelligibility function of the triplet (solid line) and the intelligibility function of the whole triplet (dotted line).

First, the intelligibility functions of all the digits were determined. As each of the 80 triplets was uttered as a whole, every digit was unique. Accordingly, 240 intelligibility functions were determined. The average guess rate (γ) was 0.146. The enormous spread shown by the x_0 values means that, at a given overall signal-to-noise ratio (i.e., average level of all triplets minus average noise level), some digits are very hard to understand, whilst others are very easy.

The second step confirmed that multiplying the three intelligibility functions [product-intelligibility function, Eq. (4)] gives essentially the same intelligibility function as the one based on the triplets [Eq. (1)]. An example is presented in Fig. 8. Although both intelligibility functions look very similar, it is important to establish that they deliver the same result in the adaptive procedure. This was done by calculating the standard deviation of $SRTT_n$ estimates for every triplet with the model, using both the triplet-intelligibility function and the product-intelligibility function as input. The correlation coefficient between the two standard deviations of $SRTT_n$ estimates was 0.92. About 94% of the differences between the corresponding standard deviations of $SRTT_n$ estimates were within 0.1 dB. Note that the differences between the product-intelligibility functions and the triplet-intelligibility functions stem from differences between the

fitted intelligibility functions and the true intelligibility functions, and from mathematical differences between the product-intelligibility function and the triplet-intelligibility function (see Sec. II A).

The third and final step consisted of an optimizing procedure. The standard deviation of $SRTT_n$ estimates was minimized for every triplet by changing the x_0 values of the three underlying digit-intelligibility functions. However, two restrictions were applied: first, the mean modeled $SRTT_n$ of the product-intelligibility function (i.e., the intelligibility of the triplet) needed to remain unchanged and, second, changes in x_0 values between the first and second, and between the second and third digits were limited to 3 dB to maintain natural speech. This procedure resulted in an average reduction factor of 0.91 (values between 1.00 and 0.75) in the standard deviation of $SRTT_n$ estimates.

3. Selecting the best triplets and expected decrease in standard deviation of $SRTT_n$ estimates

Finally, after optimizing the intelligibility functions and homogenizing the triplets, 60 triplets with the smallest standard deviation of $SRTT_n$ estimates were chosen from the original 80. As a very small set was undesirable, it was decided to limit the number of different triplets to 60. The expected decrease in the standard deviation of $SRTT_n$ estimates due to optimization of the speech material can be approximated with the calculation model. A reduction factor of standard deviation of $SRTT_n$ estimates of 0.84 was found by using the intelligibility functions of the original speech material and the optimized speech material as input.

B. Adjusting the measurement procedure

In addition, the standard deviation of $SRTT_n$ estimates can be reduced by adjusting the measurement procedure. Some procedures have already been discussed in Sec. I. However, most of them are difficult to implement on an IVR system because of the complicated nature of the calculations (e.g., maximum-likelihood estimates) and/or the strong increase in the number of sound files (e.g., adjusting the step size). Three adjustments to the measurement procedure were explored: increasing the number of presentations, using single digits or digit pairs, and changing the step size. The model was extended to be able to perform the necessary calculations. As a straightforward calculation of the weighted average and weighted standard deviation for $SRTT_n$ measurements consisting of 23 presentations would result in an extremely large number of calculations ($2^{23}=8\,388\,608$ tracks), an approximation was made. The number of tracks was limited to 16 384 (2^{14}). Every track in the calculation model described in Sec. III was randomly extended. To avoid interaction between the resulting tracks and other parameters and to minimize the effect of not using all possible tracks, the track extensions were chosen randomly for each calculation.

1. Number of presentations

The expectation was that the standard deviation of $SRTT_n$ estimates would decrease by approximately $1/\sqrt{n}$.

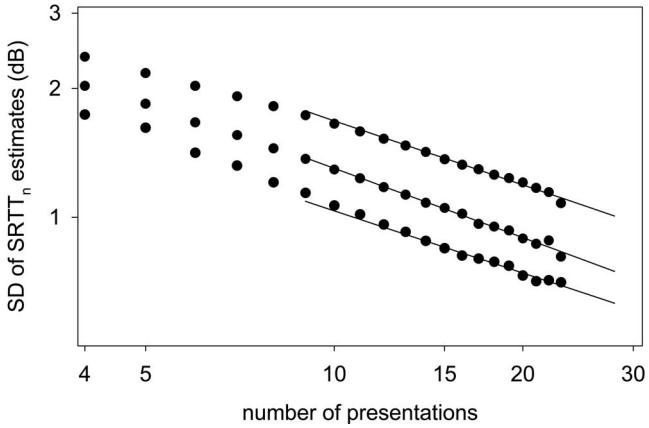


FIG. 9. Standard deviation of $SRTT_n$ estimates as a function of the number of presentations for three different intelligibility functions with slopes of 0.10 dB^{-1} (top line), 0.14 dB^{-1} (middle line), and 0.18 dB^{-1} (bottom line). The lines are the results of a linear fit on the data points from presentation 13 up to 23. The results are shown on a log-log scale.

Therefore, increasing the number of presentations offers a simple way to enhance measurement precision. The dependence of the standard deviation of $SRTT_n$ estimates on the number of presentations was calculated with the extended calculation model. Figure 9 shows the results for intelligibility functions with different slopes (0.10, 0.14, and 0.18 dB^{-1}). The results are shown on a log-log scale. The data points from $n=13$ up to $n=23$ were fitted with a linear equation. A near-perfect relationship was found ($r=-0.99$ for all curves). The slopes of the curves differ only slightly from $-1/2$ $(-0.50,-0.54,-0.48)$, which means that the standard deviation of $SRTT_n$ estimates decreases by approximately $1/\sqrt{n}$. When, for instance, the number of presentations is increased from 23 to 33, the standard deviation of $SRTT_n$ estimates decreases by a factor of 0.83.

2. Number of independent items

In the National Hearing test, triplets were used as speech material and a response counted as correct only if it was correct for all digits. A lot of time and effort could have been saved by using single digits, especially in the development phase, but using digits instead of triplets entails two serious effects which cause the standard deviation of $SRTT_n$ estimates to increase: the guess rate increases and the slope of the intelligibility function decreases. The impact of using single digits, digit pairs, or triplets on the standard deviation of $SRTT_n$ estimates was examined with the calculation model. Some assumptions were made: the intelligibility function of every digit was the same, the guess rate (γ) for one digit was 0.146 and the lapse rate was 0. Three different values for the slope S of the digit-intelligibility function were taken: 0.074, 0.104, and 0.134 dB^{-1}. With these values the slope of the triplet-intelligibility function is 0.10, 0.14, and 0.18 dB^{-1}, respectively. It can be shown that, for example, $P_{\text{digit}}(x)^3=(0.146+(1-0.146)\Phi(x,S=0.104))^3\approx P_{\text{triplet}}(x-2.69)=\Phi(x-2.69,S=0.14)$. The results in Fig. 10 show that using digit pairs instead of single digits brings about a sharp decrease in the standard deviation of $SRTT_n$ estimates. Adding an extra digit to form triplets has only a small extra effect. It is important to note that the use of single digits,

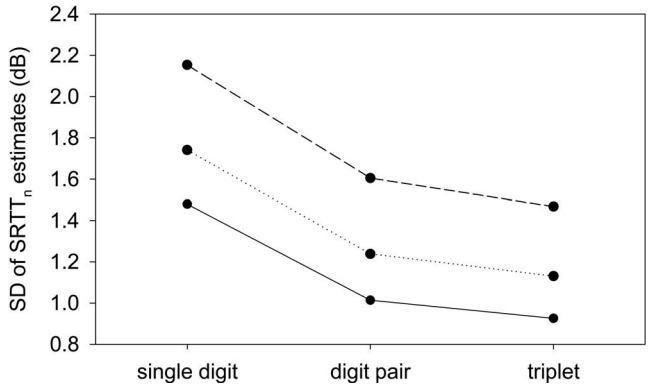


FIG. 10. Standard deviation of $SRTT_n$ estimates as a function of number of independent items: 1, 2, or 3 for a single digit, digit pair, and triplet, respectively. The results are shown for a simple up-down adaptive procedure with a step size of 2 dB and a total of 13 presentations, for intelligibility functions with slopes of 0.10 dB^{-1} (dashed line), 0.14 dB^{-1} (dotted line), and 0.18 dB^{-1} (solid line) for the triplets.

digit pairs, or triplets with the adaptive procedure results in different $SRTT_n$s. The target value is 50% intelligibility for the complete item. According to the product rule, a triplet score of 50% corresponds to a digit score of 79.4%. When using digit pairs or single digits, about 70.7% and 50% of the digits will be understood correctly.

3. Step size and starting level

Smaller steps in an adaptive procedure will, in general, lead to smaller standard deviations of $SRTT_n$ estimates. Reducing the step size will create a negative effect because it causes a higher bias (Fig. 7). An added advantage of a larger step size is that the subjects will often understand the speech more easily and feel motivated. Reducing the step size could also create practical problems as a result of the increase in the number of sound files (from 880 to 1680 when the step size is changed from 2 to 1 dB).

Calculations were performed with the calculation model (13 presentations) and the extended calculation model (23 presentations). 345 combinations of step sizes between 0.1 and 3.0 dB and true $SRTT_n$s from +0.5 to −10.5 dB, with a fixed starting level at 0 dB were used. The starting level relative to the true $SRTT_n$ is the important factor because, for instance, a starting level of 0 dB and a true $SRTT_n$ of −7 dB would give the same result as a starting level of +2 dB and a true $SRTT_n$ of −5 dB. The slope of the intelligibility function, S, was 0.14 dB^{-1} and the guess rate and lapse rate were 0.03 and 0.04, respectively. Some smoothing was applied in the graphical representation of the results from the extended calculation model. The upper panels of Fig. 11 show the standard deviation of $SRTT_n$ estimates. The lowest values are in the upper left corner because, in these cases, the step size was small and the starting level was much higher than the $SRTT_n$. Consequently, almost every response was correct and the spread minimal. The middle panel shows the bias for $SRTT_n$ estimates. As expected, bias decreases as step size and $SRTT_n$ increase. As stated in Sec. I, it is more important for a speech-in-noise test to distinguish between different conditions than to give an exact value of the $SRTT_n$. Hence, bias in the $SRTT_n$ is not a major problem in itself. The middle panel of Fig. 11 does, however, show that, for a certain step size, bias is not constant but depends on the $SRTT_n$s for a fixed starting level. Because of this effect the differences between the measured $SRTT_n$s will be somewhat smaller than between the true $SRTT_n$s which will make it more difficult to separate them. To take account of this effect each local standard deviation of $SRTT_n$ estimates (upper panel) was divided by the difference between measured $SRTT_n$ values for a true difference of 1 dB:

$$\sigma'(\mathrm{SRTT_n}) = \frac{1 \cdot \mathrm{SD\ of\ SRTT_n\ estimates}}{\mathrm{measured\ difference}|_{1\ \mathrm{dB\ true\ difference}}}. \quad (5)$$

The smaller σ', the better the test can distinguish true $SRTT_n$s differences. The results are shown in the lower panels of Fig. 11. When, for instance, 13 presentations are used with a step size of 2 dB and a $SRTT_n$ of −4 dB, the local standard deviation is 1.24 dB (the values in this example are represented by dots in Fig. 11). The difference between the measured $SRTT_n$s corresponding to true $SRTT_n$s of 3.5 and 4.5 dB is 0.96 (1 minus difference in bias), therefore the value of σ' is 1.29 (1.24/0.96). Interestingly, the lower panel reveals that, for a certain $SRTT_n$, there is an optimal choice (minimum σ') for the step size given the starting level of 0 dB. The optimal step size is indicated by the positions of the tops of the iso-σ' curves. The lower right panel of Fig. 11 allows us to check whether the parameters for the National Hearing test (step size 2 dB, fixed starting level 0 dB) were optimal. As about 80% of the $SRTT_n$s were between −7 and −3 dB (Smits and Houtgast, 2005) these parameters seem to have been fairly well chosen, although a step size of approximately 1.5 dB would have resulted in a somewhat smaller σ' for most $SRTT_n$s.

C. Measurements

Two experiments were set up to compare the results of measurements with the original speech material with the results of measurements with the optimized speech material. The optimized speech material is described in Sec. IV A. The level corrections that were applied to the individual digits ranged from −5.1 to +5.0 dB with a standard deviation of 1.7 dB. The aim of the first experiment was to confirm that the average $SRTT_n$ had remained unchanged with the optimized speech material. The aim of the second experiment was to establish whether the optimized speech material gives a smaller standard deviation of $SRTT_n$ estimates

1. Experiment 1

a. Subjects Sixteen subjects participated in the experiment. They reported no otological problems or hearing difficulties. As the intention was to compare different conditions and not to collect norm data, pure-tone audiometry was not performed. The subjects used the ear that they normally used for telephoning.

b. Apparatus A computer program was developed to simulate the characteristics of a telephone and a telephone network (a "simulated hearing test"). It included signal filtering and signal compression and decompression based on A-Law (ITU-T Recommendation P.830), the European tele-

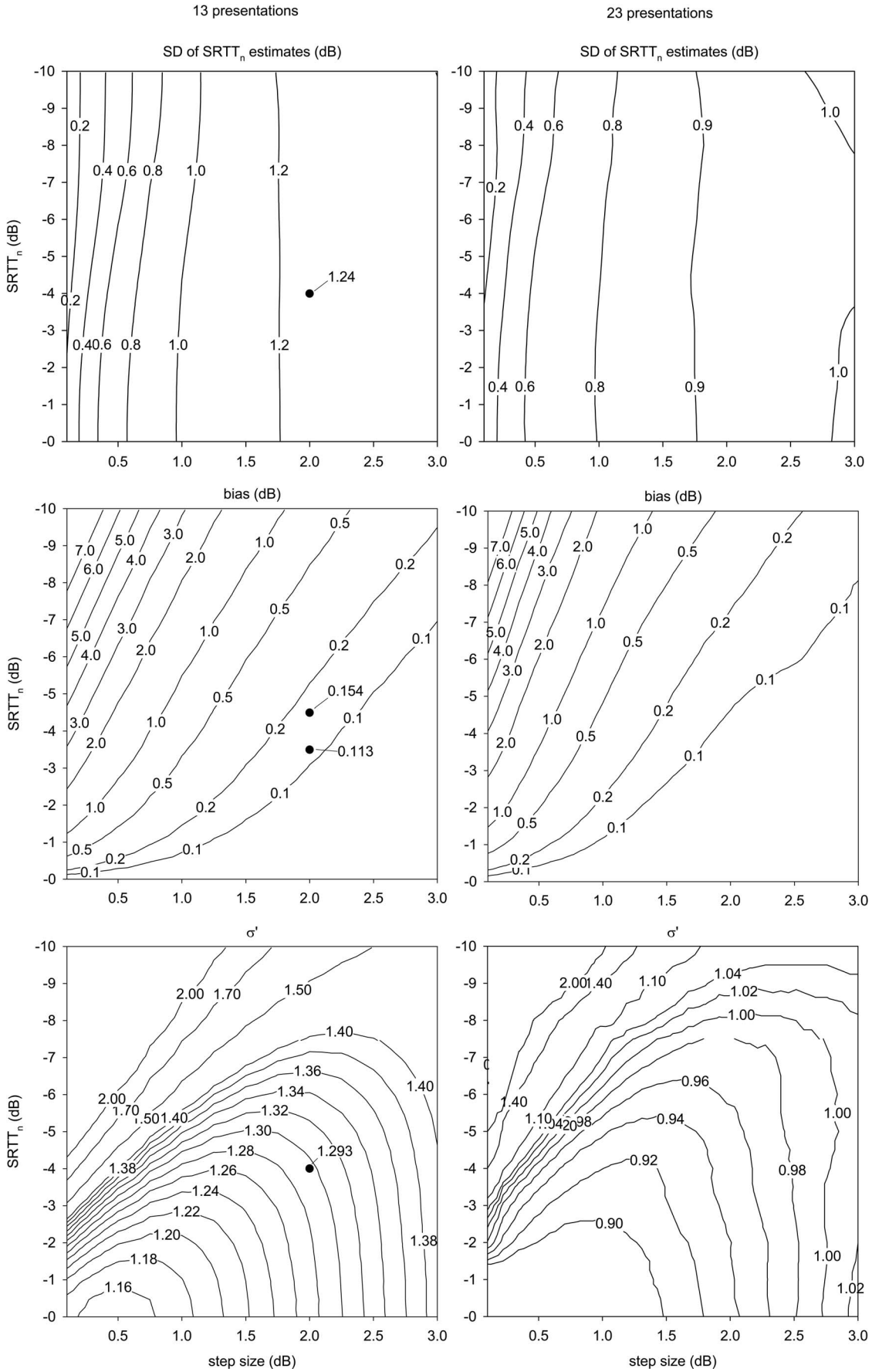


FIG. 11. Effect of true $SRTT_n$ (relative to a starting level of 0 dB) and step size on standard deviation of $SRTT_n$ estimates (top panels), bias (middle panels), and σ' (lower panels). The left panels show the results from the calculation model with 13 presentations per measurement and the right panels show the results from the extended calculation model with 23 presentations per measurement. σ' can be considered the best measure of the accuracy of the test because it includes the combined effects of standard deviation of $SRTT_n$ estimates and bias (see the text). The dots in the left panels represent the data used in an example (see the text).

TABLE I. Results of $SRTT_n$ measurements with 16 subjects. Each $SRTT_n$ measurement consisted of 23 presentations. Average $SRTT_n$s for different setups and for the original and the optimized speech material is displayed. No significant effect of condition was found.

			$SRTT_n$ (SD)
National Hearing test		Telephone	−6.6 (1.6)
National Hearing test		Headphones	−6.9 (1.1)
Simulated hearing test with original speech material	Test	Headphones	−6.5 (1.3)
	Retest		−7.0 (1.3)
Simulated hearing test with optimized speech material	Test	Headphones	−6.4 (1.3)
	Retest		−6.5 (1.5)

TABLE II. Results of $SRTT_n$ measurements for 244 subjects. Each $SRTT_n$ measurement consisted of 13 presentations. No significant effect of speech material was found. The decrease in the standard deviation of $SRTT_n$ estimates is significant (F-test, $p=0.08$). An intelligibility function was determined by fitting the data (after correction for interindividual differences in $SRTT_n$) with a cumulative normal distribution. The slope of the intelligibility function is shown in the last column.

	$SRTT_n$ (average of test and retest)	SD of $SRTT_n$ estimates	Slope (dB^{-1})
Simulated hearing test with original speech material	−6.9	1.31	0.171
Simulated hearing test with optimized speech material	−7.0	1.12	0.199

phony standard. Signals were played by a standard sound card and presented monaurally through headphones. In the computer program it was possible to chose between original speech material and optimized speech material. To enable comparisons, the National Hearing test was performed as implemented on an interactive voice response system (Smits and Houtgast, 2005).

c. Measurement procedure The measurement procedure in the computer program was exactly the same as the procedure for the National Hearing test (Smits and Houtgast, 2005). Each subject performed eight different $SRTT_n$ measurements, six of which are relevant here: National Hearing test by telephone, National Hearing test by headphones (monaural), simulated hearing test with original speech material by headphones (test and retest) and simulated hearing test with optimized speech material by headphones (test and retest).

d. Results The results are summarized in Table I. The last column presents the mean and standard deviation over the 16 subjects. First the $SRTT_n$s were analyzed in an analysis of variance (ANOVA) with measurement condition as a within-subjects factor. For the simulated hearing tests, only the first measurements (test condition) were used. The effect of measurement condition was not significant, $F(3,45)=0.982$, $p=0.41$. Then the results from the simulated hearing tests were explored in more detail by a 2×2 ANOVA with speech material (original versus optimized) and measurement (test versus retest) as within-subject factors. The effects of speech material $[F(1,15)=2.22, p=0.16]$ and measurement $[F(1,15)=0.02, p=0.96]$ were not significant. It may be concluded that, despite the manipulations of the speech material, the restriction that the measured $SRTT_n$s remain unchanged was met.

2. Experiment 2

a. Subjects A total of 244 medical students participated in this experiment, which formed part of a practice exercise on hearing.

b. Apparatus and measurement procedure The setup (simulated hearing test) was the same as in the first experiment. However, the signals were presented diotically. Each subject performed two $SRTT_n$ measurements (test and retest). In this experiment a single $SRTT_n$ measurement consisted of 13 presentations. The type of speech material (original or optimized) was chosen randomly for each subject.

c. Results The results are summarized in Table II. The $SRTT_n$s were analyzed in an ANOVA with measurement (test versus retest) as a within-subject factor and speech material (original versus optimized) as a between subjects factor. The ANOVA yielded a significant effect for measurement; no significant interaction between measurement and speech material was found. The difference between test and retest values is probably due to a training effect, which diminishes when the number of presentations increases to 23. The main effect for speech material was not significant $[F(1,242)=0.586, p=0.45]$. The standard deviation of $SRTT_n$ estimates was derived from the test-retest differences (i.e., standard deviation of the differences between test and retest, divided by $\sqrt{2}$). A reduction factor of the standard deviation of 0.85 was found for the $SRTT_n$ estimates (i.e., the ratio of both standard deviations of $SRTT_n$ estimates). This reduction factor comes very close to the estimated reduction factor of 0.84. A one-tailed F-test revealed a significant decrease in the standard deviation of the $SRTT_n$ estimate ($p=0.08$).

V. DISCUSSION

With almost 40 000 $SRTT_n$ measurements the database containing the National Hearing test results is unique in terms of size. These data and the calculation model presented in this paper enabled a thorough investigation of the measurement procedure and the speech material.

The calculation model can be applied to estimate the accuracy of the simple up-down adaptive procedure. As it uses all possible tracks in the adaptive procedure to calculate the average $SRTT_n$ and the standard deviation of $SRTT_n$ estimates, it is preferable to Monte Carlo simulations. However, when the number of presentations increases, the number of tracks becomes so high that it is impossible to use them all. Hence, an approximation must be made. The ability to perform exact calculations is particularly important if the calculation model is used in an optimizing procedure in which its output is minimized.

Optimizing the speech material resulted in a reduction factor of 0.85 in the standard deviation of $SRTT_n$ estimates. This may not seem particularly impressive, but it is equivalent to an increment from 23 to 32 in the number of presentations. The optimizing method applied in this paper, however, is highly time-consuming and requires many $SRTT_n$ measurements to determine the intelligibility functions of the individual digits.

A. Slope bias of the intelligibility function

An important issue raised in Sec. II A was the inability to determine the exact form of the intelligibility function. Kaernbach (2001) demonstrated a very large slope bias when data from individual tracks in a simple up-down adaptive procedure were fitted with a maximum-likelihood procedure (i.e., slope of the psychometric function of a single observer). Kaernbach (2001) and Klein (2001) maintain that the only way to determine the slope of these psychometric functions without bias is to apply a procedure aimed at different points of the psychometric function. In the present study no attempts were made to determine psychometric functions from individual subjects. However, the slope estimates from the intelligibility functions for groups of subjects have a bias as well (Sec. II A).[3]This can also be demonstrated by using data from Sec. IV C 2. Table II shows the slopes of the intelligibility functions for the original speech material and the optimized speech material. The slope estimates of 0.171 and 0.199 dB^{-1} were determined after the data had been corrected with individual $SRTT_n$ estimates. They were also determined without the corrections. Slope estimates of 0.106 and 0.117 dB^{-1} were found, respectively. The true slope values should lie somewhere between these values and the values reported in Table II. These true slope values can be derived from Fig. 4 via the standard deviation of $SRTT_n$ estimates from Table II. The data are plotted in Fig. 4. Using this figure the true slope values were estimated at 0.116 and 0.142 dB^{-1}, respectively. It should be noted here that the slope bias will be lower for intelligibility functions based on adaptive procedures that use 23 presentations. This is because the standard deviation of $SRTT_n$ estimates is smaller, and consequently the difference between the true $SRTT_n$ and the $SRTT_n$ estimate is smaller.

B. Effect of guess rate, lapse rate, and heterogeneity of the speech material

The effect of the guess rate and lapse rate needs to be ascertained for two reasons. First, it is necessary to confirm that the measured $SRTT_n$ actually represents the point of 50% intelligibility. Second, lapses cannot always be avoided. For instance, in the National Hearing test a response cannot be corrected when a wrong key is pressed accidentally. Figure 5 shows the effect of the guess rate on the measured $SRTT_n$. As expected, an increase was found in the standard deviation of $SRTT_n$ estimates and a difference was found between the true $SRTT_n$ and the measured $SRTT_n$. However, even for a guess rate as high as 20%, the bias is less than 0.3 dB and the corresponding intelligibility is higher than 45%. This implies that the simple up-down procedure is relatively insensitive to the guess rate or unknown lapse rate.

The contribution of homogeneity of the speech material toward reliable measurements was investigated (Fig. 6). It may be concluded that, even for steep intelligibility functions (0.18 dB^{-1}), a standard deviation of 1 dB in 50% intelligibility points has very little effect on the standard deviation of $SRTT_n$ estimates. When the standard deviation in 50% intelligibility points equals the step size of 2 dB, the standard deviation of $SRTT_n$ estimates increases from 0.92 to 1.18 dB. The adaptive procedure works less effectively in such cases because, in about 16% of the presentations, the signal-to-noise ratio will be higher (or lower) than the preceding presentation although the response was correct (or incorrect). Wagener *et al.* (1999) have devised a formula to calculate the slope of a test list based on the distribution of the individual slopes of the items. This formula can be used to determine the effect of heterogeneity of the speech material[4] (Fig. 6); it delivers essentially the same results as those delivered by the calculation model. It was checked out by calculating the slope of the "mean" intelligibility function [4] for different values of the standard deviation of the points of 50% intelligibility. The mean intelligibility function was used as input when the standard deviation of $SRTT_n$ estimates was calculated with the model. The correlation between the standard deviation of $SRTT_n$ estimates found with this procedure and the values presented in Fig. 6 was almost 1.

C. Efficiency of the adaptive procedure

Brand and Kollmeier (2002) estimated the SRT_n by applying an adaptive procedure with a decreasing step size. They concluded that word-scoring is far more efficient than sentence-scoring because of the increase in the number of independent items per sentence. Hagerman and Kinnefors (1995) demonstrated the applicability of an adaptive procedure in which the step size was based on the number of correct words in a sentence. Such a procedure could also be used for the triplet speech material. Probably, a smaller standard deviation of $SRTT_n$ estimates will be found with the same speech material. However, digit-scoring will be ambiguous if only one or two digits are understood. The position of the digit that was not understood must then be known, otherwise the scoring method fails.

The calculations confirmed the experimental findings that speech-in-noise measurements which use relatively few presentations and a simple up-down procedure with sentence-scoring result in a low standard deviation of $SRTT_n$ estimates. With only 13 sentences Plomp and Mimpen (1979) found a standard deviation of $SRTT_n$ estimates of 0.9 dB, and Versfeld *et al.* (2000) reported an error of 1.1 dB. Versfeld *et al.* found in both their experimental results and Monte Carlo simulations that calculating the $SRTT_n$ by averaging presentation levels gives a smaller standard deviation of $SRTT_n$ estimates than fitting the data with an intelligibility function. We recently confirmed this finding by analyzing data from adaptive speech-in-noise tests. When using the simple up-down adaptive procedure, it is therefore recommended to calculate the $SRTT_n$ by simply averaging the presentation levels. Probably, more accurate results can be gained with more sophisticated adaptive procedures and calculation methods (e.g., Brand and Kollmeier, 2002; Zera, 2004).

To compare the efficiency of adaptive procedures Brand and Kollmeier (2002) used the normalized standard deviation of threshold estimates ($\hat{\sigma}$). This is defined as the standard deviation of threshold estimates of the specific procedure, $\sigma_{\text{procedure}}$, divided by the theoretical minimal standard devia-

TABLE III. Normalized standard deviation of threshold estimates for three different intelligibility functions. The triplet intelligibility functions are represented by cumulative normal distributions. It is assumed that the procedure determines the $SRTT_n$, i.e., $P=0.5$ for triplets or $P=0.79$ for individual digits.

Slope (dB^{-1})	$\sigma_{procedure}/\sigma_{min}=\hat{\sigma}$	
	One independent item per presentation	Three independent items per presentation
0.10	1.47/1.39=1.06	1.47/1.31=1.12
0.14	1.13/0.99=1.14	1.13/0.93=1.21
0.18	0.92/0.77=1.19	0.92/0.72=1.28

tion of threshold estimates, σ_{min} (Taylor, 1971; Green, 1995). For a certain target threshold (e.g., $P=0.5$ for the $SRTT_n$), the normalized standard deviation of threshold estimates can be approximated as follows:

$$\hat{\sigma}=\sigma_{procedure}\Bigg/\frac{\sqrt{P_{threshold}\cdot(1-P_{threshold})}}{\left.\frac{dP}{d\mathrm{SNR}}\right|_{\mathrm{SNR=threshold}}\cdot\sqrt{n}}. \quad (6)$$

Calculations were performed to determine $\hat{\sigma}$ for the simple up-down adaptive procedure. Thirteen presentations and a step size of 2 dB were used in the calculations. Three different intelligibility functions were applied: cumulative normal distributions with slopes of 0.10, 0.14, and 0.18 dB^{-1}. The normalized standard deviation of threshold estimates is shown in Table III. These calculations assumed that there was only one independent item per presentation, but there are, in principle, three. This could be taken into account when calculating the theoretical minimal standard deviation of threshold estimates. Because the intelligibility function of a single digit differs from the intelligibility function of a triplet (shallower slope and a nonzero guess rate), and because the target threshold differs ($P=0.79$ for single digits to measure $P=0.5$ for triplets), the decrease in the theoretical minimal standard deviation of threshold estimates will be less than $1/\sqrt{3}$. The intelligibility function of the digits was taken from Sec. IV B 2.[5] The results are shown in Table III. Note that the theoretical minimal standard deviation assumes three independent items whereas the standard deviation of threshold estimates assumes just one independent item. It must be concluded that the combination of a simple up-down adaptive procedure and a calculation method that averages presentation levels is highly efficient. As both the theoretical minimal standard deviation of threshold estimates and the standard deviation of threshold estimates for the simple up-down adaptive procedure are proportional to $1/\sqrt{n}$ when n is not too small, the normalized standard deviation of threshold estimates does not depend on the number of presentations.

The simple up-down procedure is only highly efficient when the first presentation level is not too far from the $SRTT_n$. There are two simple ways of achieving this: first, the procedure proposed by Plomp and Mimpen (1979), i.e., repeat the first presentation with an increasing signal-to-noise ratio until the response is correct. Second, choose a starting level somewhere in the middle of the range of $SRTT_n$s. As this range is about 15 dB, the maximum difference between the starting level and $SRTT_n$ will not be greater than 8 dB and the bias will be small for most $SRTT_n$s (middle panel Fig. 11).

D. Triplet speech material versus digit speech material

The difference between triplet speech material and digit speech material was also investigated with the calculation model. It should be noted that the number of independent items increases by a factor of 3, but the slope of the intelligibility function decreases, the guess rate increases and the proportion of correctly repeated digits decreases from about 0.79 to 0.50 (see Sec. IV B 2). The presentation of 10 triplets with slopes of the intelligibility functions of the triplets of 0.14 dB^{-1} (cf. Sec. IV B 2) was compared with the presentation of 30 single digits. The standard deviation of $SRTT_n$ estimates decreased from 1.29 to 1.19 dB. This means that the benefit from increasing the number of presentations is greater than the combined loss from the decreasing slope and the increasing guess rate. On the other hand, it probably makes the test less user-friendly.

The experimental findings and the mathematical results enable us to review the development process and the procedure of the National Hearing test. In general terms, it can be said that the development process yielded homogeneous triplets and the parameters in the measurement procedure were well chosen. The desired standard deviation of $SRTT_n$ estimates can be controlled by the number of presentations. However, some suggestions are presented for developing a comparable test (e.g., in other languages). In the National Hearing test digit triplets were enunciated as a whole to promote naturalness of speech. A great many measurements were therefore necessary to obtain homogeneity between the different triplets. In addition, the advantage of using three digits, and consequently, of being able to create very steep slopes of the intelligibility function of the triplets, was not fully utilized because selection applied only to the triplets and not to the digits. It might be useful to combine single digits into optimal triplets or digit pairs with the aid of the calculation model.

VI. CONCLUSIONS

This study explored the up-down adaptive procedure in speech-in-noise measurements. Almost 40 000 $SRTT_n$ measurements were used from the Dutch speech-in-noise telephone test (National Hearing test). The findings are as follows:

(1) The intelligibility function for the speech material from the National Hearing test can be described by a cumulative normal distribution, a lapse rate, and a guess rate. The intelligibility function of a triplet can be constructed from the intelligibility functions of the individual digits.

(2) The standard deviation of $SRTT_n$ estimates increases with hearing loss. This is not age-related or due to a connection between heterogeneity of the speech material and $SRTT_n$.
(3) The calculation model presented in this study can be used to examine the influence of the characteristics of the speech material and the measurement method on the standard deviation of $SRTT_n$ estimates. It can also be used to optimize the speech material.
(4) When using the simple up-down adaptive procedure, the guess rate or lapse rate has only a minor effect on the intelligibility percentage that corresponds to the measured $SRTT_n$. There is, of course, a negative effect on the standard deviation of $SRTT_n$ estimates.
(5) A fixed starting level can be used if chosen in the middle of the range of $SRTT_n$s.
(6) Theoretically, optimizing the speech material of the National Hearing test by homogenizing the triplets, performing level corrections to individual digits and selecting 60 out of the original 80 different items was expected to lead to a reduction factor of about 0.84 in the standard deviation of $SRTT_n$ estimates. This was confirmed by $SRTT_n$ measurements in 244 subjects.
(7) The usefulness of the speech-in-noise test is defined by the standard deviation of $SRTT_n$ estimates, bias, and how they interact. For the National Hearing test the starting level of 0 dB and a step size of 2 dB turned out to be good choices.

ACKNOWLEDGMENTS

The authors acknowledge Annemarie de Backer for her contribution to experiment 1 (Sec. IV C 1). They further wish to thank the two anonymous reviewers for their valuable comments and the third reviewer, C. Kaernbach, for the fruitful discussion on the topic of slope bias.

[1]The data were also fitted with a logistic function. Again, a near-perfect fit was found. Parameters were: $S=0.155\ dB^{-1}$, $\gamma=0.010$, $\lambda=0.025$, $x_0=-0.29$. As a cumulative normal distribution approaches the asymptotic values quicker than a logistic function with the same slope, different values were found for the guess rate and lapse rate.

[2]It was decided to use a simple cumulative normal distribution as an approximation of the intelligibility function in order to directly compare fitting results. This was not possible when fitting the data with the general function, Eq. (1), because the standard deviation of $SRTT_n$ estimates depends on several properties of the intelligibility function: S, γ, and λ. Later, it was possible to verify, by applying the calculation model, that the effect of using $\gamma=\lambda=0$ on the standard deviation of $SRTT_n$ estimates is very small because it is compensated for by a shallower slope.

[3]As mentioned in Sec. II A this slope bias arises from the procedure in which data of individual listeners are shifted in order to align thresholds. In that procedure the $SRTT_n$ estimate is used whereas the true $SRTT_n$ should be used. It is important to realize that the slope bias is not caused by the adaptive procedure itself. This was supported by additional calculations: mean intelligibility functions were calculated from the data (see Fig. 1), but now only one trial from every individual measurement (track) was used. When comparing the intelligibility functions that are based on the fifth, thirteenth or twenty-third trial to the original intelligibility function, based on all the data points, only slightly different slope values were found: 0.152, 0.161, and 0.173 dB^{-1}, respectively. The mean slope of those three intelligibility functions was 0.162 dB^{-1}, almost identical to the value of 0.158 dB^{-1} as reported in Fig. 1. It can be verified mathematically that the intelligibility function that is derived from pooled adaptive data approximates the true intelligibility function when the number of tracks increases and there is no variance in true $SRTT_n$ values (this means that the variance in $SRTT_n$ estimates is only due to the measurement error, and consequently, no correction for interindividual differences should be applied).

[4]The formula can be written as: $S_{overall}=S/[1+(\text{s.d. of } 50\% \text{ points})^2/\sigma^2]^{1/2}$ where $\sigma=1/S\sqrt{2\pi}$.

[5]The theoretical minimal standard deviation of threshold estimates for the triplet intelligibility function with a slope of 0.14 dB^{-1} was calculated by using the corresponding digit intelligibility function (Sec. IV B 2) and $n=3\cdot13:\sigma_{min}=[0.79\cdot(1-0.79)]^{1/2}/(1-0.146)\cdot0.104\cdot e^{-\pi(0.104\cdot2.69)^2}\cdot(3\cdot13)^{1/2}$.

Brand, T., and Kollmeier, B. (**2002**). "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," J. Acoust. Soc. Am. **111**, 2801–2810.
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.
Green, D. M. (**1990**). "Stimulus selection in adaptive psychophysical procedures," J. Acoust. Soc. Am. **87**, 2662–2674.
Green, D. M. (**1995**). "Maximum-likelihood procedures and the inattentive observer," J. Acoust. Soc. Am. **97**, 3749–3760.
Hagerman, B., and Kinnefors, C. (**1995**). "Efficient adaptive methods for measuring speech reception threshold in quiet and in noise," Scand. Audiol. **24**, 71–77.
International Telecommunications Union (**1996**). "Subjective performance assessment of telephone-band and wideband digital codecs," ITU-T Recommendation, P. 830.
Kaernbach, C. (**2001**). "Slope bias of psychometric functions derived from adaptive data," Percept. Psychophys. **63**, 1389–1398.
Klein, S. (**2001**). "Measuring, estimating, and understanding the psychometric function: A commentary," Percept. Psychophys. **63**, 1421–1455.
Kollmeier, B., Gilkey, R. H., and Sieben, U. K. (**1988**). "Adaptive staircase techniques in psychoacoustics: A comparison of human data and a mathematical model," J. Acoust. Soc. Am. **83**, 1852–1862.
Leek, M. R. (**2001**). "Adaptive procedures in psychophysical research," Percept. Psychophys. **63**, 1279–1292.
Lyzenga, J., Festen, J. M., and Houtgast, T. (**2002**). "A speech enhancement scheme incorporating spectral expansion evaluated with simulated loss of frequency selectivity," J. Acoust. Soc. Am. **112**, 1145–1157.
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.
Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18**, 43–52.
Smits, C., and Houtgast, T. (**2005**). "Results from the Dutch speech-in-noise screening test by telephone," Ear Hear. **26**, 89–95.
Smits, C., Kapteyn, T. S., and Houtgast, T. (**2004**). "Development and validation of an automatic speech-in-noise screening test by telephone," Int. J. Audiology **43**, 15–28.
Smoorenburg, G. F. (**1992**). "Speech recognition in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram," J. Acoust. Soc. Am. **91**, 421–137.
Taylor, M. M. (**1971**). "On the efficiency of psychophysical measurements," J. Acoust. Soc. Am. **49**, 505–508.
Taylor, M. M., and Creelman, C. D. (**1967**). "PEST: Efficient estimates on probability functions," J. Acoust. Soc. Am. **41**, 782–787.
Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). "Method for the selection of sentence materials for efficient measurements of the speech reception threshold," J. Acoust. Soc. Am. **107**, 1671–1684.
Wagener, K., Brand, T., and Kollmeier, B. (**1999**). "Entwicklung und evaluation eines satztests für die deutsche Sprache. Teil II. Optimierung des Oldenburger satztests," Z. Audiol. **38**, 44–56.
Zera, J. (**2004**). "Speech intelligibility measured by adaptive maximum-likelihood procedure," Speech Commun. **42**, 313–328.

Temporal characteristics of nasalization in children and adult speakers of American English and Korean during production of three vowel contexts[a)]

Seunghee Ha[b)]
Department of Audiology and Speech Pathology, University of Tennessee, Knoxville, Tennessee 37996

David Kuehn
Department of Speech and Hearing Science, University of Illinois, Urbana-Champaign, Champaign, Illinois 61820

(Received 15 June 2005; revised 14 June 2006; accepted 19 June 2006)

The purpose of this study was to identify and compare the temporal characteristics of nasalization in relation to (1) languages, (2) vowel contexts, and (3) age groups. Two distinct acoustic energies from the mouth and nose were recorded during speech production (/pamap, pimip, pumup/) using two microphones to obtain the absolute and proportional measurements on the acoustic temporal characteristics of nasalization. Twenty-eight normal adults (14 American English and 14 Korean speakers) and 28 normal children (14 American English and 14 Korean speakers) participated in this study. In both languages, adults showed shorter duration of nasalization than children within all three vowel contexts. The high vowel context revealed longer duration of nasalization than the low vowel context in both languages. There was no significant difference of temporal characteristics of nasalization between American English and Korean. Nasalization showed different timing characteristics between children and adults across vowel contexts. The results are discussed in association with developmental coarticulation and the relationship between acoustic consequences of articulatory events and vowel height. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2225382]

PACS number(s): 43.72.Ar, 43.70.Ep, 43.70.Kv [AL] Pages: 1622–1630

I. INTRODUCTION

Sequential movements of the velum and tongue or lips are essential in normal production of nasal consonants. Velar lowering for nasal sounds begins before the tongue or lips begin the movement toward occlusion. Opening of the velopharyngeal port is already achieved at the time the oral tract becomes closed (Kent *et al.*, 1974; Kuehn, 1976). The velopharyngeal port then keeps opening while the oral occlusion is released and becomes open. Such an early onset and delayed offset of velar movement occurs before and after oral cavity occlusion and causes the vowel preceding and following nasal consonants to be nasalized for certain durations. Therefore, unless velopharyngeal opening and closing are appropriately coordinated with oral closure and opening, nasalization will deviate from the normal range and excessive nasal resonance may be perceived.

In our previous study we investigated the temporal characteristics of nasalization in children with and without cleft palate using acoustic temporal measures derived from both oral and nasal acoustic energies that were simultaneously recorded (Ha *et al.*, 2004). Children with repaired cleft palate in that study showed significantly longer durations of nasalization than children without cleft palate. The results of the study suggest that children with cleft palate exhibit greater anticipatory and carryover nasal coarticulation than normal children. The study also demonstrated that the acoustic temporal measures utilized in that study might provide useful information on nasal coarticulation in speech production and nasalization that deviates from normal.

Several investigators have reported that the temporal features of nasalization are influenced by vowel contexts, age, and language using a variety of measurements (e.g., Flege, 1988; Solé, 1995; Young *et al.*, 2001). The relationship between vowel height and articulatory movements of the velum, tongue, and jaw has been extensively investigated (e.g., Bell-Berti and Krakow, 1991; Kuehn, 1976; Kuehn and Moon, 1998). Nasal airflow and nasal sound pressure levels which result from modifying the velum, tongue, or jaw position have been found to be different across vowel contexts (e.g., Lubker and Moll, 1965; Young *et al.*, 2001). A direct association between nasal airflow levels and vowel height has been reported. Nasal airflow is substantially greater for high vowels than for low vowels in vowel-nasal consonant contexts. Hajek (1997), among others, pointed out that increased tongue height causes greater constriction in the oral cavity and consequently produces increased impedance to oral airflow. As a result, more air would be directed to the nasal cavity in the production of high vowels.

In addition, acoustic consequences of velopharyngeal opening have been investigated. Nasalization is primarily

[a)]Portions of this work were presented at American Cleft Palata-Craniofacial Association meeting in Myrtle Beach, SC, April 2005 and at American Speech-language Hearing Association Annual meeting in San Diego, CA, November 2005.
[b)]Electronic mail: sha@utk.edu

characterized by a decrease in the amplitude of the first formant (F1). Perceptual experiments of synthetic stimuli showed that F1 amplitude lowered by 6–8 dB is necessary to achieve a significant level of nasality perception (House and Stevens, 1956; Stevens *et al.*, 1988). The reduced F1 amplitude is associated with an increase in formant bandwidths and upward shifts in formant frequencies. Also, a high-energy nasal formant is typically present in the region of 600–1000 Hz and is accompanied by low energy of the upper formants due to the presence of antiformants. However, the spectral characteristics of nasalization vary across speakers and phonetic contexts, especially vowel types. High and low vowels show different spectral changes for the same amount of velopharyngeal opening. This current study investigates whether the relationship between acoustic consequences of articulatory events and vowel height exists in the temporal domain of nasalization and whether their relationship can be quantitatively described using the temporal acoustic measures in our previous study (Ha *et al.*, 2004).

Developmental changes of coarticulation have been a topic of interest in children's speech compared to adults' speech. Several acoustic studies have investigated whether the temporal domain of coarticulation is the same in the speech of children and adults (Flege, 1988; Katz *et al.*, 1991; Repp, 1986; Turnbaugh *et al.*, 1985). The results of these studies tend to be inconsistent. Some studies suggest that coarticulatory effects are more prominent in the speech of children than adults (e.g., Repp, 1986). In contrast, Kent (1983) suggested that children may show a restriction in the extent of coarticulation as they may be more likely than adults to show "synchronous" patterns of interarticulator timing in which the multiple gestures needed for a phone coincide. With regard to various results on coarticulatory patterns in the speech of children and adults, Repp (1986) suggested that some forms of coarticulation are an indication of advanced speech production skills whereas others may be a sign of articulatory immaturity, and yet others are neither of these because they simply cannot be avoided. In the present study, acoustic temporal measures of nasalization have been obtained from the speech of children and adults to investigate developmental changes of nasalization.

Another purpose of this study is to determine whether differences between American English and Korean exist in the temporal extent of nasalization. Clumeck (1976) in a study investigating patterns of soft palate movements in six languages suggested that the degree of coarticulated nasality is language-specific. Solé (1995) analyzed cross-linguistic data on coarticulatory nasalization of vowels preceding a nasal consonant in American English and Spanish. The author found that the two languages showed different patterns of timing, magnitude, duration, and velocity of velopharyngeal movements. To our knowledge, no study has investigated features of nasalized vowels or nasal coarticulation in the Korean language. As in English, the Korean language has the same three nasal consonants (/m, n, ŋ/) and no phonemically nasalized vowels or diphthongs. Given that the Korean and English languages are phonemically and phonologically similar with regard to the nasalization feature, it is hypothesized that the nasal duration segments measured in the study are also similar between the two languages.

TABLE I. Age means and ranges of each group.[a]

Group	KC	AC	KA	AA
Mean	5;8	5;9	27	29
Range	4;5–7;2	4;6–7;11	24-33	19–47

[a]KC=Korean children; AC=American children; KA=Korean adults; AA=American adults.

The present study provides information regarding the possible influences of vowel context, age, and language on acoustic measures of nasalization.

II. METHOD

A. Participants

To determine the temporal aspects of nasalization for English versus Korean and for age, 28 normal adults (14 American English and 14 Korean speakers) and 28 normal children (14 American English and 14 Korean speakers) participated in this study. All speakers reported American English or Korean as their first language. Participants were divided into four groups in terms of their age and first language. Each group consisted of 7 males and 7 females. The age mean and range among groups are presented in Table I. Child groups ranged in age from 4 to 7 years and adult groups ranged in age from 19 to 47 years. Participants or their parents reported no history involving speech, language, or hearing impairments, nor orofacial anomalies. Informed parental or adult written consents were obtained for all participants. Children's oral assents also were obtained before data collection.

B. Speech sample and data collection

All recordings were obtained in conditions that attenuated background noise as much as possible. All participants repeated speech tasks that included bilabial nasal and oral stops in carrier phrases. The carrier phrase for English speakers was "say —— again." The Korean carrier phrase was "Uri —— irago hayo." ("We say ——."). To determine whether the temporal measures of nasalization are different among vowel contexts, the speech tasks consisted of three different vowels, /i/, /a/, and /u/. The speech tasks were /pimip/, /pamap/, and /pumup/. These speech samples contain phonotactically permissible sequences in both languages. No speakers had difficulty in producing the samples.

All participants produced each speech task five times with a specially devised microphone set to obtain two distinct acoustic signals from the oral and nasal cavities. The microphone set consisted of two microphones and was the same as that used in a previous study (Ha *et al.*, 2004). The mouth microphone (Shure, WH30XLR) was positioned at a fixed distance of 3 cm from the right side of the subject's mouth and the nose microphone (Shure, MX184) was positioned just below the right nostril. Two distinct acoustic energies from the mouth and nose were recorded simulta-

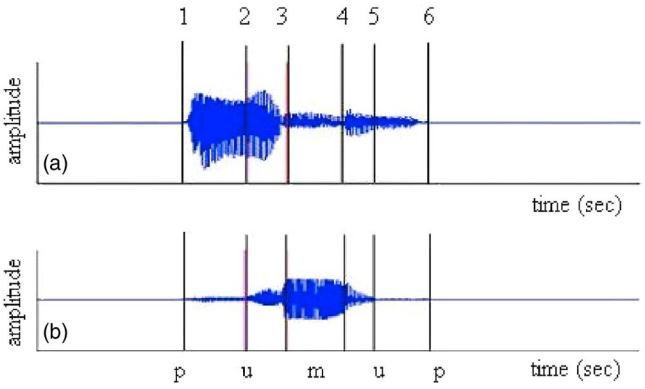


FIG. 1. (Color online) Example of wave forms from (a) oral microphone and from (b) nasal microphone for /pumup/ spoken by a Korean child. Timing parameters measured: 2-3, nasal onset interval; 3-4, acoustic period of the nasal consonant; 4-5, nasal offset interval; 2-5, total nasalization duration; 1-6 total utterance duration [from Ha *et al.*, Cleft Palate Craniofac. J. **41**, 535–543 (2007), reprinted with permission].

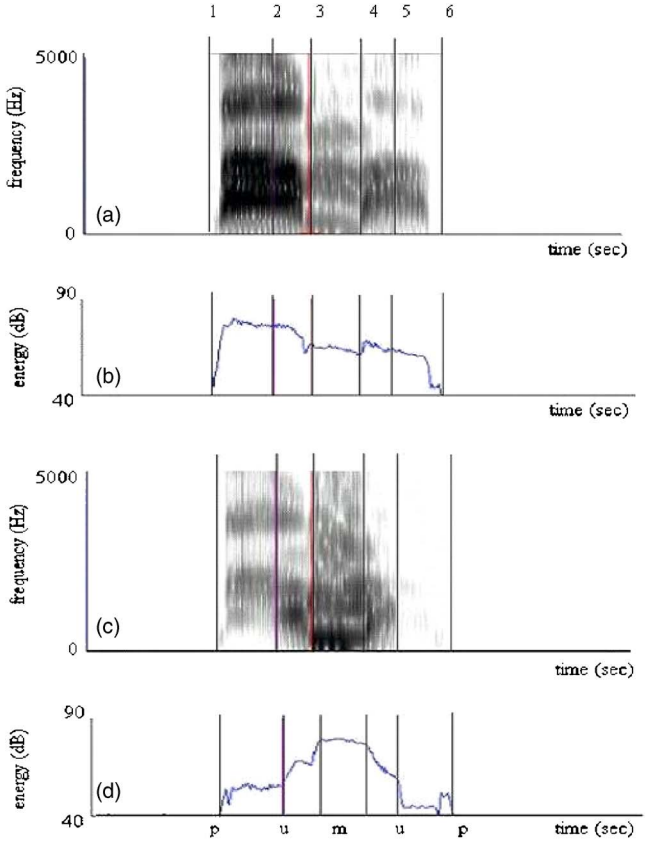


FIG. 2. (Color online) Example of traces for /pumup/ spoken by a Korean child. (a) Spectrogram from oral microphone; (b) energy contour from oral microphone; (c) spectrogram from nasal microphone; (d) energy contour from nasal microphone. Timing events 1, 2, 3, 4, 5, 6 are the same as those indicated in Fig. 1 [from Ha *et al.*, Cleft Palate Craniofac. J. **41**, 535–543 (2007), reprinted with permission].

neously through two channels of a preamplifier (M-Audio, Audio Buddy), which controlled the amplitude of the two acoustic energies at an appropriate level for analysis.

All subjects practiced the test words enough to articulate the speech sample correctly before recording. Each subject produced the test words at a conversational pitch and loudness level. Each subject was given an auditory model of the test words at conversational pitch and loudness. To control the effect of pitch and loudness of the voice, stress, and the rate of speech on the production of nasalization as much as possible, Korean subjects were asked to repeat the auditory model of S.H. at every test word production. American subjects repeated the auditory model of one graduate student who speaks American English as a first language.

C. Data analysis

The two distinct acoustic signals corresponding to the speech samples were analyzed using the Computerized Speech Lab (KayPENTAX™, model 4300). The oral acoustic signal and the nasal acoustic signal were measured utilizing two channels within the software. Nasal onset interval, nasal offset interval, and total nasalization duration, which reflect timing characteristics of nasalization, were measured on the basis of features revealed in the acoustic wave forms, spectrograms, and energy contours. The definition of timing measurements and the procedure used to obtain the measurements were identical to the ones described in Ha *et al.* (2004). The timing measurements are defined in the following and displayed in Figs. 1 and 2.

Nasal onset interval: Measured from a point showing a clear increase of nasal acoustic energy and a decrease of oral acoustic energy to the starting point of acoustic period of the nasal consonant. This interval is between time points "2" and "3" in Figs. 1 and 2.

Acoustic period of the nasal consonant: Refers to the period between time points "3" and "4" in Figs. 1 and 2. This acoustic period is characterized by formants with decreased intensity on spectrograms from the oral microphone.

Nasal offset interval: Measured from the end of the acoustic period of the nasal consonant to the increase of oral acoustic energy and the decrease of nasal acoustic energy indicating the oral opening for the production of the following vowel. This interval is between time points "4" and "5" in Figs. 1 and 2.

Total nasalization duration: The sum of nasal onset interval, acoustic period of the nasal consonant, and nasal offset interval. This duration is between time points "2" and "5" in Figs. 1 and 2.

Each measure was taken on the basis of timing points showing increase-decrease of nasal and oral acoustic energies when nasal consonant, /m/, was produced before and after vowels. The timing points of increase-decrease of nasal and oral acoustic energies reflect the sequential opening/closing movements of the lips and velum. Therefore, it is assumed that the three measures, nasal onset interval, nasal offset interval, and total nasalization duration are associated with onset-offset timing of lip and velum movements. We also measured total utterance duration which is between time points "1" and "6" in Figs. 1 and 2.

In addition to the above-described absolute duration measures, proportional duration measures of nasalization also were computed. These proportional measurements were obtained to control the influence of different speech rate on temporal characteristics of nasalization between children and

TABLE II. Means, standard deviations, and coefficients of variation in milliseconds for each temporal period by group.[a]

		/pamap/				/pimip/				/pumup/			
		Children		Adults		Children		Adults		Children		Adults	
		English	Korean	English	Korean	English	Korean	English	Korean	English	Korean	English	Korean
Onset	Mean	33.24	31.48	24.24	17.86	55.14	51.26	38.24	27.60	61.05	65.12	36.52	26.71
	s.d.	8.3	8.45	7.23	3.30	13.97	17.12	14.19	8.77	12.10	33.18	9.51	5.65
	CV	0.25	0.27	0.30	0.19	0.25	0.33	0.37	0.32	0.20	0.51	0.26	21
Nasal	Mean	55.74	55.76	51.98	49.57	78.14	71.14	60.45	54.43	77.86	70.33	59.05	50.95
	s.d.	16.56	15.95	9.85	9.42	15.71	18.15	9.94	8.43	19.99	17.47	7.16	10.06
	CV	0.30	0.29	0.19	0.19	0.20	0.26	0.16	0.16	0.26	0.25	0.12	0.20
Offset	Mean	32.17	31.26	23.29	21.60	85.17	105.14	74.86	64.71	89.71	91.19	60.76	59.64
	s.d.	15.03	18.10	6.64	6.27	24.53	38.03	19.86	23.82	23.79	42.87	23.87	20.62
	CV	0.47	0.58	0.29	0.29	0.29	0.36	0.27	0.37	0.27	0.47	0.39	0.35
Total nasal.	Mean	121.14	118.50	99.50	89.02	218.45	227.55	173.55	146.74	228.62	234.17	156.33	137.31
	s.d.	20.18	29.12	15.97	13.19	34.09	51.27	31.91	30.12	44.93	86.80	34.95	24.38
	CV	0.17	0.25	0.16	0.15	0.16	0.23	0.18	0.21	0.20	0.37	0.22	0.18
Total Utter	Mean	381.88	369.02	325.19	289.55	382.79	390.40	315.19	283	420	404.55	306.17	273.21
	s.d.	38.23	38.71	26.05	20.16	34.44	47.84	26.21	24.79	42.78	74.68	25.43	28.11
	CV	0.10	0.11	0.08	0.07	0.11	0.12	0.8	0.9	0.10	0.19	0.8	0.10

[a]Onset=nasal onset interval; Nasal=acoustic period of nasal consonant; Offset=nasal offset interval; Total Nasal=total nasalization duration; Total utter =total utterance duration; CV=coefficient of variation.

adults and speakers of the two different languages. Ratios were measured by dividing each interval into the total utterance duration and then multiplying by 100.

Proportional duration measures of nasalization were computed as follows.

Nasal onset interval ratio=nasal onset interval/total utterance duration ×100; *Nasal offset interval ratio*=nasal offset interval/total utterance duration ×100; *Total nasalization duration ratio*=total nasalization duration/total utterance duration ×100.

D. Statistical analyses

A split-plot factorial design was used to determine whether the absolute and proportional temporal measures were significantly different among the three variables: language, age, and vowel contexts. Repeated measure ANOVAs were used to investigate the differences of absolute as well as proportional nasal onset interval, nasal offset interval, and total nasalization duration among the four groups and within vowel contexts.

Effect size of each repeated measure by group, age, and language was calculated by partial eta squared (η^2). Fisher-Hayter multiple comparison tests were used to test for differences of the three timing parameters among /i/, /u/, and /a/ vowel contexts within each subject group.

E. Measurement reliability

Intra- and interobserver reliability were calculated by means of a Pearson correlation coefficient analysis. A subset of the data which comprised approximately 10% of the entire data set was measured by two observers. For intraobserver reliability, the subset was remeasured by the same investigator (S. H.) after the original measurements were completed.

All Pearson *r* values of the intraobserver measures ranged from 0.85 to 0.99 and the *r* values between observers ranged from 0.77 to 1.0. Among the measures, nasal offset interval showed relatively low correlations between the two observers because in some acoustic data the entire duration of vowels following nasal consonants tended to be nasalized and the end points of nasal offset interval (i.e., decrease of nasal acoustic energy) were not clear.

III. RESULTS

The group means, standard deviations, and coefficients of variation of nasal onset interval, acoustic period of nasal consonant, nasal offset interval, total nasalization duration, and total utterance duration for each test word are presented in Table II. Overall, children from both language groups showed larger means as well as standard deviations of all temporal measurements than adults. The coefficient of variation is the standard deviation of a distribution of scores divided by the mean for the distribution. It is a measure of relative variability and is sensitive to deviations from proportionality of the standard deviation to the mean (Carlton and Newell, 1993). The coefficients of variation revealed that children generally showed larger variability than adults in Tables II and III. Each temporal period associated with nasalization in the test words produced by the four groups is revealed graphically in Fig. 3. Each segment of the bars represents nasal onset interval, acoustic period of the nasal consonant, and nasal offset interval from bottom to top. The total length of the bars represents total nasalization duration. As shown in Fig. 3, children who speak American English and children who speak Korean exhibited relatively longer nasal onset interval, acoustic period of the nasal consonant, nasal offset interval, and total nasalization duration than adults of the two language groups in all the test words. In particular,

TABLE III. Means, standard deviations, and coefficients of variation of three proportional durations of nasalization.[a]

		/pamap/				/pimip/				/pumup/			
		Children		Adults		Children		Adults		Children		Adults	
		English	Korean	English	Korean	English	Korean	English	Korean	English	Korean	English	Korean
Onset ratio	Mean	8.7	8.5	7.5	6.2	14.4	13.1	12.1	9.8	14.5	16.1	11.9	9.8
	s.d.	1.9	1.6	2.1	1.0	3.4	4.2	4.1	3.1	2.3	7.3	2.6	2.1
	CV	0.22	0.19	0.28	0.16	0.24	0.32	0.34	0.32	0.16	0.45	0.22	0.21
Nasal ratio	Mean	14.8	15.0	15.9	17.1	20.5	18.1	19.2	19.3	18.6	17.6	19.3	18.7
	s.d.	4.9	3.6	2.2	3.1	3.8	3.9	2.5	2.6	4.3	4.1	2.5	3.7
	CV	0.33	0.24	0.14	0.18	0.19	0.22	0.13	0.13	0.23	0.23	0.13	0.20
Offset ratio	Mean	8.4	8.6	7.2	7.5	14.5	17.8	16.0	15.5	21.2	22.1	19.8	21.6
	s.d.	3.6	4.7	2.0	2.1	4.4	5.5	3.8	5.1	4.5	7.3	7.7	6.6
	CV	0.43	0.55	0.28	0.28	0.30	0.31	0.24	0.33	0.21	0.33	0.39	0.31
Total nasal ratio	Mean	31.9	32.1	30.6	30.8	57.1	57.9	54.9	51.9	54.4	56.8	51.0	50.2
	s.d.	4.8	5.9	3.9	4.0	7.6	8.5	6.9	9.0	7.6	12.1	10.7	7.0
	CV	0.15	0.18	0.13	0.13	0.13	0.15	0.13	0.17	0.14	0.21	0.21	0.14

[a]Onset ratio=nasal onset interval ratio; Nasal ratio=nasal acoustic segment ratio; Offset ratio=nasal offset interval ratio; Total nasal ratio=total nasalization duration ratio.

nasal onset interval and nasal offset interval for the high vowel, /u/ and /i/ contexts revealed the greatest difference between children and adult groups. The low vowel, /a/, context exhibited relatively small differences between the age groups for the nasal onset interval, nasal offset interval, and total nasalization duration compared to the other vowel contexts. Regarding the timing variables across vowel contexts, high vowel contexts revealed longer nasal onset interval, nasal offset interval and total nasalization duration than low vowel contexts for all the groups.

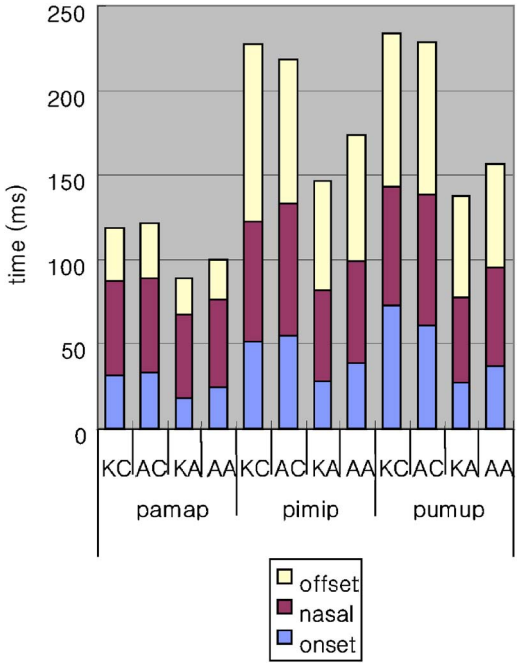


FIG. 3. (Color online) The length of three temporal measures in the test words produced by Korean/American children and adults. KC=Korean children; AC=American children; KA=Korean adults; AA=American adults.

A repeated measure ANOVA was conducted to determine whether there were differences in nasal onset interval among age groups and language groups within /i/, /a/, and /u/ contexts. The analysis revealed that the child group showed significantly longer nasal onset interval than the adult group ($F_{1,52}=71.599, p<0.0001; \eta^2=0.579$). Significant differences also were found within the three vowel contexts ($F_{2,80}=34.157, p<0.0001; \eta^2=0.396$). Regarding interaction among variables, only interaction between age and vowel contexts showed a significant difference ($F_{2,80}=9.196, p<0.005; \eta^2=0.150$). The results of the Fisher-Hayter multiple comparison test revealed that all pairwise contrasts among the three vowel contexts were significantly different in child groups of both languages. However, the adult Korean speaker group showed no differences among the three vowel contexts. The adult American English speaker group showed significant difference only in nasal onset intervals between /a/ and /i/ vowel contexts.

In addition, the child group showed a significantly longer acoustic period of the nasal consonant than the adult group ($F_{1,52}=19.627, p<0.0001; \eta^2=0.274$). There were significant differences among the three vowel contexts ($F_{2,90}=31.336, p<0.0001; \eta^2=0.376$). The results also showed significant interaction between age and vowel contexts ($F_{2,80}=9.425, p<0.0001; \eta^2=0.153$). The results of the Fisher-Hayter test indicated that the nasal acoustic segments of the two high vowel contexts were significantly longer than those of the low vowel /a/ context in the child groups of Korean and American English speakers. Adult American English speakers showed significantly longer nasal acoustic segments of the /i/ vowel context than those of the /a/ vowel context.

Regarding nasal offset interval, the results of the ANOVA analysis indicated that the child group in both lan-

guages showed significantly longer nasal offset interval than the adult group ($F_{1,52}=17.944, p<0.0001; \eta^2=0.257$). There were significant differences within the three vowel contexts ($F_{2,104}=152.865, p<0.0001; \eta^2=0.746$). The interaction between age and vowel contexts also showed a significant difference ($F_{2,104}=5.071, p<0.01; \eta^2=0.089$). The results of the Fisher-Hayter test indicated that the child Korean speaker group and adult American English speaker group showed significant differences among all pairwise contrasts of the three vowel contexts. Nasal offset interval of /a/ vowel context showed significant differences from those of /i/ and /u/ in adult Korean speakers and child American English speakers. That is, nasal offset interval showed no statistically significant difference between /i/ and /u/ vowel contexts in the two groups.

The child group showed significantly longer total nasalization duration than the adult group ($F_{1,52}=45.322, p<0.0001; \eta^2=0.466$). There were significant differences within the three vowel contexts ($F_{2,104}=160.906, p<0.0001; \eta^2=0.756$). There was significant interaction between age and vowel contexts ($F_{2,104}=15.484, p<0.0001; \eta^2=0.229$). The results of the Fisher-Hayter test indicated that total nasalization durations of /a/ vowel context was significantly shorter than those of /i/ and /u/ in all four groups. The differences of total nasalization durations between /i/ and /u/ were not statistically different in all four groups.

There were no significant differences of nasal onset intervals, nasal offset intervals, and total nasalization duration between the two languages, American English and Korean.

Table II shows the group means and standard deviations of the absolute durations of the phrases for each of the test words. Table III shows the group means, standard deviations, and coefficients of variation of nasal onset interval ratio, nasal acoustic segment ratio, nasal offset interval ratio, and total nasalization duration ratio for each test word. The ANOVA revealed that for nasal onset interval ratio there was a significant difference between the age groups ($F_{1,52}=36.168$, $p<0.0001$; $\eta^2=0.410$), and significant difference within the vowel contexts ($F_{2,89}=45.979$, $p<0.0001$; $\eta^2=0.469$). Also, interaction between age and vowel contexts showed a significant difference ($F_{2,89}=3.293$, $p<0.05$; $\eta^2=0.060$). The results of the Fisher-Hayter test indicated that nasal onset interval ratio of /a/ vowel context was significantly shorter than those of /i/ and /u/ vowel contexts in all groups. With regard to relative duration of acoustic nasal segment to total utterance duration (nasal acoustic segment ratio), the results showed no significant differences in age and language variables. However, there were significant differences among the three vowel contexts ($F_{2,93}=26.648$, $p<0.0001$; $\eta^2=0.339$). The Fisher-Hayter test indicated that nasal acoustic segment ratio of the /a/ vowel context was significantly shorter than those of /i/ and /u/ in the three groups, child groups of both language, and adult group of English.

Children in both language groups showed larger nasal offset interval ratio than adults in their corresponding language groups, but the difference was not statistically significant. Nasal offset interval ratio within the vowel contexts showed significant difference ($F_{2,104}=145.596, p<0.0001; \eta^2=0.737$). There were no significant interactions between variables. The results of the Fisher-Hayter test indicated that the Korean child speaker group and American English adult speaker group showed significant differences among all pairwise contrasts of the three vowel contexts. Korean adult speakers and American English child speakers showed significantly shorter nasal offset interval ratio of /a/ vowel context than those of /i/ and /u/ vowel contexts.

The analysis revealed that for total nasalization duration ratio there was a significant difference between the age groups ($F_{1,52}=4.954, p<0.05; \eta^2=0.087$), and significant difference within the vowel contexts ($F_{2,104}=256.728, p<0.0001; \eta^2=0.832$). There were no significant interactions between variables. As the results of absolute total nasalization duration, total nasalization duration ratio of /a/ vowel context was significantly smaller than those of /i/ and /u/ vowel contexts in child and adult groups.

IV. DISCUSSION

This study investigated the temporal characteristics of nasalization in relation to two languages, three vowel contexts, and two age groups using absolute and proportional temporal measurements. The absolute measurements of nasalization revealed that adults showed significantly shorter nasal onset interval, nasal offset interval, and total nasalization duration than children for all three vowel contexts in American English and Korean. It is assumed that the difference of absolute nasalization duration between children and adults might be partially due to different speech rates between the two groups. With regard to vowel contexts, the high vowel contexts revealed significantly longer durations of the three measurements on nasalization than the low vowel context within adults and children groups and in the two languages. In contrast, and in support of our stated hypothesis, there were no significant differences of temporal characteristics of nasalization between American English and Korean. The two languages showed similar temporal patterns of nasalization in adults versus children and among the three vowels. These findings suggest that the two languages which are similar phonemically and phonologically with regard to the nasalization feature are also similar with regard to nasal acoustic segment durations.

The proportional duration measures of nasalization that were obtained to control the possible influence of different speech rates among the four groups showed interesting results compared to those of the absolute measurements. Nasal offset interval ratio showed no statistically significant difference between children and adults in both language groups in contrast with significant difference of its corresponding absolute measurement. However, nasal onset interval ratio and total nasalization duration ratio showed statistically significant differences between children and adults in both language groups, which are consistent with the results of the absolute measurements. Significant difference of total nasalization duration ratio between children and adults resulted from different nasal onset interval ratio between age groups. From the perspective of coarticulation, nasal onset interval

ratio represents anticipatory nasal coarticulation and nasal offset interval ratio corresponds to carryover nasal coarticulation. The results of this study suggest that children tend to exhibit greater temporal domain of anticipatory, but not carryover nasal coarticulation than adults. The results suggested that there might be distinctive underlying mechanisms of anticipatory and carryover nasal coarticulation. Bell-Berti (1993) reported that consensus among coarticulation studies has been that anticipatory coarticulation reflects reorganization of the motor system for segmental articulations, while carryover coarticulation has usually been attributed to mechanical and inertial forces acting on the articulators. Mechanical velar inertia resulting in carryover nasal coarticulation might be constant for both adults and children whereas anticipatory nasal coarticulation might show developmental patterns of reorganization of the motor system for segmental articulations and, therefore, adults and children show different temporal domains of anticipatory nasal coarticulation.

The results regarding anticipatory coarticulation can be discussed in association with Kent (1983) hypothesis. Kent (1983) suggested that children may show more synchronous patterns of interarticulator timing than adults. In support of Kent's hypothesis, Thomson and Hixon (1979) suggested that the temporal domain of anticipatory nasal coarticulation increases with age. On the other hand, Flege (1988) found that there were no significant differences of nasalization durations between children and adults. The inconsistent results among studies might be due to the difference in measurement techniques and composition of the speech sample. Thomson and Hixon (1979) focused on nasal airflow at the midpoint of the initial vowel in /ini/. In this study two separate but simultaneous acoustic signals from the oral and nasal cavities were analyzed to investigate temporal patterns of anticipatory and carryover coarticulation. The acoustic signals may derive from *articulatory* movements including both velopharyngeal opening and oral closing. Children may accomplish oral closing following the velum lowering for nasal consonants later than adults, which may result in longer nasal onset interval in this study.

In addition, Thomson and Hixon (1979) used the carrier phrase, "Say____again." That would substantially lengthen the vocalic elements preceding the nasal consonant /n/ in /ini/ compared to the current study in which a single vowel preceded the nasal consonant segments. The current study also used longer speech samples than those used in Thomson and Hixon (1979) which would tend to shorten the duration of all segments. Developmentally, speakers may learn to take advantage of the longer vocalic elements and begin velar lowering earlier as they mature. Children, on the other hand, might be more constrained physiologically, without such learning, and require relatively longer time to lower the velum if the nasal consonant is preceded by only a single vowel and an obstruent consonant such as the speech sample used in the current study. Flege (1988) used the carrier phrase, "A half____for Sam." In that sample for the test word /nin/, for example, the velum must lower immediately after the /f/ in half and then raise immediately before the /f/ in for. For these rapid maneuvers, both young children and mature speakers might be equally constrained by biomechanical factors. In that situation, the learning component would be expected to be minimal, which could help to explain the lack of differences in nasalization durations between adults and children in the Flege study. Future studies must take into account these possible influences of phonetic context.

The results of the present study can be interpreted in light of development of speech motor control. Several investigators have found that children show longer segmental durations, greater temporal and spectral variability, less stable and less segmentally differentiated speech movements than adults (Lee *et al.*, 1999; Goffman and Smith, 1999; Smith and Goffman, 1998; Goodell and Studdert-Kennedy, 1993; Nittrouer *et al.*, 1989). Indeed, children showed longer durations for all temporal measurements in this study: nasal onset interval, acoustic period of nasal consonants, nasal offset interval, total nasalization duration, and total utterance duration. This trend also was observed in the proportional temporal measurements although the extent of differences between children and adults were small due to control of different speech rates. In addition, children generally showed greater variability than adults in all temporal measurements in this study. Coefficients of variation for each temporal measurement by group were obtained to compare temporal variability between children and adults in Tables II and III. The coefficients of variation revealed that children generally showed larger variability than adults.

The different temporal characteristics of nasalization between children and adults might reflect different speech motor control skills of children and adults with regard to coordinated movements of the velum and lips. The longer nasalization duration of children might indicate prolonged interarticulatory intervals of the speech mechanism. Improvements in speed of interarticulatory movements with age might be concomitant with systematic reduction of duration.

This possible interpretation on developing coordinated speech movements in children's speech might be closely related to temporal characteristics of speech in speakers with cleft palate (Forner, 1983;Ha *et al.*, 2004). Forner (1983) found that speech segment durations in children with cleft palate were significantly longer than normal in the single word and nasal sentence contexts. Also, the results in Forner (1983) revealed that the group with cleft palate produced longer plosive and affricate stop gaps. The results in Ha *et al.* (2004) also suggest that children with cleft palate exhibit longer temporal durations of nasalization than normal children. Speakers with cleft palate might show developmental delay or deviance of speech motor control skill due to their anatomical differences in the structure of the palate. Hoopes *et al.* (1993) found that speed of velar movement during speech is slower for cleft than noncleft subjects. Speakers with cleft palate might show some difficulty with normal rate and range of movement and interarticulatory timing. The requirements of repeated opening and closing of the velopharyngeal valve add increased motor complexity to an utterance. Forner (1983) suggested that the added requirement of rapid velopharyngeal valving contributed to the lengthened nasal sentence. The author speculated that subjects with clefts might have difficulty tolerating the rapid velar valving action without some prolongation of the segments.

Kuehn and Moon (1995) showed that individuals with cleft palate use relatively higher levator veli palatini activation levels for speech compared to normal individuals (Kuehn and Moon, 1994). This suggests that individuals with cleft palate may require greater effort to achieve velopharyngeal closure to avoid hypernasality. As a result, individuals with cleft palate may avoid a situation of fatigue by reducing both the magnitude and velocity of velar displacements, both of which could prolong the nasalization duration.

The assumption is that children with cleft palate might begin opening the velopharyngeal port earlier because doing so requires less effort than opening the velopharyngeal port rapidly soon before the oral or lingual constriction (Flege, 1988; Ha *et al.*, 2004). Children with cleft palate might show earlier velopharyngeal opening than children without cleft palate to perform this activity without greater effort and thus avoiding a potentially fatiguing activity (Kuehn and Moon, 1995). This manipulation of interarticulatory timing appears to be more noticeable in speakers with cleft palate than without cleft palate and in children than adults. Therefore, temporal measurements of nasalization duration might decrease along with development of motor control skill.

Regarding the vowel contexts, the results are consistent with Ha *et al.* (2004). It is assumed that the degree of the oral constriction could possibly account for the results of this study. Increased tongue height of high vowels results in greater oral constriction than low vowels. Greater constriction in the oral cavity results in increased impedance to oral airflow. Consequently, more air will be directed to the nasal cavity in the production of high vowels (Hajek, 1997). For a given small velopharyngeal opening in vowel contexts, amount of nasal coupling must depend on oral constriction. Indeed, Lubker and Moll (1965) found that anticipatory nasal airflow increased although the size of the velopharyngeal orifice remained constant during production of a vowel-nasal sequence. They interpreted that this finding might be due to an increase in oral cavity constriction. Young *et al.* (2001) found that female speakers exhibited greater ratios of nasal to oral-plus-nasal airflow during production of /ini/ than during productions of /ana/. They suggested that anticipatory nasal airflow may be determined by the configuration of the oral cavity.

In addition, different configurations of the oral and nasal tracts in terms of vowel types have varying influences on the acoustic characteristics of vowel nasalization. Many studies showed that the acoustic consequences of oral-nasal coupling are not identical across all vowel types (e.g., House and Stevens, 1956; Maeda, 1993). Small levels of velopharyngeal opening have a substantial effect on high vowel spectra, while low vowels show little change at the same levels of velopharyngeal opening and required a far greater amount of velopharyngeal opening to achieve the same degree of acoustic alteration as in high vowels. These varying acoustic consequences of oral-nasal coupling indicate that vowel nasalization is complicatedly influenced by degrees of oral constriction as well as velopharyngeal opening. Increased oral constriction in the high vowel contexts might contribute to longer nasal onset interval, nasal offset interval, and total nasalization duration than those in the low vowel context.

V. CONCLUSION

Using two distinct acoustic signals from the oral and nasal cavities, it was found that the Korean speakers in this study showed similar temporal patterns of nasalization to those of the English speakers. It was also found that high vowel contexts exhibited longer durations of nasalization than low vowel contexts. This result suggests that the acoustic temporal measurements reflect changes of oral and nasal acoustic signals in accordance with varying vowel contexts. In addition, absolute temporal measures of nasalization were found to be longer in children's speech than in adults' speech. It is assumed that different temporal characteristics of nasalization between children and adults might reflect different speech motor control skills of children and adults with regard to coordinated movements of the velum and lips. The results from proportional temporal measures showed that there are significant differences in anticipatory nasal coarticulation between children and adults but not in carryover nasal coarticulation. This result suggests that anticipatory and carryover nasal coarticulation have distinctive underlying mechanisms.

Bell-Berti, F. (**1993**). *Veli Motor Control: Segmental Context in Phonetics and Phonology* Vol. **5**, edited by M. K. Huffman and R. A. Krakow (Academic, San Diego), pp. 63–85.

Bell-Berti, F., and Krakow, R. A. (**1991**). "Anticipatory velar lowering: A coproduction account," J. Acoust. Soc. Am. **90**(1), 112–123.

Carlton, L. G., and Newell, K. M. (**1993**). "Force variability and characteristics of force production," in *Variability and Motor Control*, (Human Kinetics, Champaign) pp. 15–36.

Clumeck, H. (**1976**). "Patterns of soft palate movements in six languages," J. Phonetics, **4**, 337–351.

Flege, J. (**1988**). "Anticipatory and carry-over nasal coarticulation in the speech of children and adult," J. Speech Hear. Res. **31**, 525–536.

Forner, L. (**1983**). "Speech segment durations produced by five and six year old speakers with and without cleft palates," Cleft Palate J. **20**, 185–198.

Goodell, E. W., and Studdert-Kennedy, M. (**1993**). "Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: A longitudinal study," J. Speech Hear. Res. **36**, 707–727.

Goffman, L., and Smith, A. (**1999**). "Development and phonetic differentiation of speech movement patterns," J. Exp. Psychol. **25**, 649–660.

Ha, S., Sim, H., Zhi, M., and Kuehn, D. P. (**2004**). "An acoustic study of the temporal characteristics of nasalization in children with and without cleft palate," Cleft Palate Craniofac J. **41**(5), 535–543.

Hajek, J. (**1997**). *Universals of Sound Change in Nasalization* (The Philological Society, Oxford).

Hoopes, J. E., Dellon, A. L., Fabrikant, J. I., Edgerton, M. T., Jr., and Soliman, A. H. (**1970**). "Cineradiographic definition of the functional anatomy and apathophysiology of the velopharynx," Cleft Palate J. **7**, 443–454.

House, A. S., and Stevens, K. N. (**1956**). "Analog studies of the nasalization of vowels," J. Speech Hear Disord. **21**, 218–232.

Katz, W. F., Kripke, C., and Tallal, P. (**1991**). "Anticipatory coarticulation in the speech of adults and young children: Acoustic, perceptual, and video data," J. Speech Hear. Res. **34**(6), 1222–1232.

Kent, R. (**1983**). "The segmental organization of speech," in *The Production of Speech*, edited by P. MacNeilage (Springer, New York), pp. 57–90.

Kent, R. D., Carney, P. J., and Severeid, L. R. (**1974**). "Velar movement and timing: Evaluation of a model for binary control," J. Speech Hear. Res. **17**(3), 470–488.

Kuehn, D. P. (**1976**). "A cineradiographic investigation of velar movement variables in two normals," Cleft Palate J. **13**, 88–103.

Kuehn, D. P., and Moon, J. B. (**1994**). "Levator veli palatini muscle activity in relation to intraoral air pressure variation," J. Speech Hear. Res. **37**(6), 1260–1270.

Kuehn, D. P., and Moon, J. B. (**1995**). "Levator veli palatini muscle activity in relation to intraoral air pressure variation in cleft palate subjects," Cleft Palate Craniofac J. **32**(5), 376–381.

Kuehn, D. P., and Moon, J. B. (**1998**). "Velopharyngeal closure force and levator veli palatini activation levels in varying phonetic contexts," J. Speech Lang. Hear. Res. **41**(1), 51–62.
Lee, S., Potamianos, A., and Narayanan, S. (**1999**). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**(3), 1455–1468.
Lubker, J. F., and Moll, K. L. (**1965**). "Oral-nasal air flow measurements and cinefluorographic observations during speech production," Cleft Palate J. **2**, 257–272.
Maeda, S. (**1993**). "Acoustics of Vowel Nasalization and Articulatory Shifts in French Nasal Vowels," in *Phonetics and Phonology*, Vol. **5**, edited by M. K. Huffman and R. A. Krakow (Academic, San Diego), pp. 147–167.
Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (**1989**). "The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults," J. Speech Hear. Res. **32**(1), 120–132.
Repp, B. H. (**1986**). "Some observations on the development of anticipatory coarticulation," J. Acoust. Soc. Am. **79**(5), 1616–1619.
Smith, A., and Goffman, L. (**1998**). "Stability and patterning of speech movement sequences in children and adults," J. Speech Lang. Hear. Res. **41**(1), 18–30.
Solé, M. J. (**1995**). "Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization," Lang Speech **38**(1), 1–23.
Stevens, K. N., Andrade, A., and Ceu Viana, M. (**1988**). "Perception of vowel nasalization in VC contexts: A cross-language study," J. Acoust. Soc. Am. **82**, S119.
Thomson, E., and Hixon, T. J. (**1979**). "Nasal air flow during normal speech production," Cleft Palate J. **16**, 412–420.
Turnbaugh, K., Hoffman, P., Daniloff, R. D., and Absher, R. (**1985**). "Stop-vowel coarticulation in 3-year-olds, 5-years-olds, and adults," J. Acoust. Soc. Am. **77**, 1256–1258.
Young, L. H., Zajac, D. J., Mayo, R., and Hooper, C. R. (**2001**). "Effects of vowel height and vocal intensity on anticipatory nasal airflow in individuals with normal speech," J. Speech Lang. Hear. Res. **44**(1), 52–60.

Speech utterance clustering based on the maximization of within-cluster homogeneity of speaker voice characteristics

Wei-Ho Tsai[a)]
Department of Electronic Engineering, National Taipei University of Technology, Taipei, Taiwan

Hsin-Min Wang[b)]
Institute of Information Science, Academia Sinica, Taipei, Taiwan

(Received 2 June 2005; revised 19 June 2006; accepted 19 June 2006)

This paper investigates the problem of how to partition unknown speech utterances into a set of clusters, such that each cluster consists of utterances from only one speaker, and the number of clusters reflects the unknown speaker population size. The proposed method begins by specifying a certain number of clusters, corresponding to one of the possible speaker population sizes, and then maximizes the level of overall within-cluster homogeneity of the speakers' voice characteristics. The within-cluster homogeneity is characterized by the likelihood probability that a cluster model, trained using all the utterances within a cluster, matches each of the within-cluster utterances. To attain the maximal sum of likelihood probabilities for all utterances, the proposed method applies a genetic algorithm to determine the cluster in which each utterance should be located. For greater computational efficiency, also proposed is a clustering criterion that approximates the likelihood probability with a divergence-based model similarity between a cluster and each of the within-cluster utterances. The clustering method then examines various legitimate numbers of clusters by adapting the Bayesian information criterion to determine the most likely speaker population size. The experimental results show the superiority of the proposed method over conventional methods based on hierarchical clustering.
[DOI: 10.1121/1.2225570]

PACS number(s): 43.72.Pf, 43.72.Fx [DOS] Pages: 1631–1645

I. INTRODUCTION

For more than two decades, automatic speaker recognition based on vocal characteristics[1–3] has received a tremendous amount of attention in research that facilitates human-machine communications and biometric applications. Nowadays, as speech is being exploited as an information source, the utility of recognizing speakers' voices is increasingly in demand for indexing and archiving the mushrooming amount of spoken data available universally. Traditional approaches to speaker recognition assume that some prior information or speech data is available about the speakers concerned. Thus, speaker-specific models can be trained using the labeled speech data, and the likelihoods of unknown test utterances can then be computed from the models, thereby determining the identity of a speaker (*speaker identification*), or determining if a speaker is who he/she claims to be (*speaker verification*). However, for indexing or archiving, the basic strategy needs to be expanded to distinguish between speakers when neither information about the speakers' voices nor the speaker population size is available. As a result, unsupervised classification of speech data based on speakers' voice characteristics has emerged as a new and challenging research problem;[4–8] however, the solutions to this problem require further investigation.

Classifying speech data by speaker is generally associated with two processes. One is to segment speech data into homogeneous utterances that contain only one speaker's voice. The other is to group together homogeneous utterances from the same speaker into a cluster. The former is usually referred to as *speaker segmentation*,[8,9] while the latter is referred to as *speaker clustering*.[10,7] A joint process consisting of speaker segmentation and clustering, called *speaker diarization*,[11–14] was recently defined by the NIST Speech Group.[15] It is hoped that by locating speech segments from the same speaker, the human effort required for indexing speech data can be greatly reduced from having to listen to every long audio recording to only having to check a few utterances in each cluster. In addition, by locating speech utterances from speakers with similar voices,[16–18] transcription or recognition of speech messages can be carried out more effectively by adapting acoustic models on a per cluster basis, which exploits more adaptation data than on a per utterance basis.

In this paper, we concentrate on the problem of speaker clustering. Given N unlabeled speech utterances, each of which is assumed to be from one of P speakers, where $N \geq P$ and P is unknown, speaker clustering is defined as the partitioning of N utterances into M clusters, such that $M = P$, where each cluster consists exclusively of utterances from only one speaker. For utterances that contain multiple speakers, the partitioning is preferably performed after the utterances are presegmented into speaker-homogeneous regions. However, in order to focus on the fundamental tech-

[a)]Electronic mail: whtsai@en.ntut.edu.tw
[b)]Electronic mail: whm@iis.sinica.edu.tw

niques for speaker clustering, this study does not investigate the speaker-segmentation problem, but only deals with utterances containing a single speaker.

Currently, most speaker-clustering methods follow a hierarchical clustering framework,[7,10,17,19–25] which consists of three major components: computation of interutterance similarities, generation of a cluster tree, and determination of the number of clusters. Similarity computation is designed to produce larger values for similarities between utterances of the same speaker and smaller values for similarities between utterances of different speakers. Several similarity measures, such as the arithmetic harmonic sphericity (AHS),[9] Kullback Leibler (KL) distance,[25] the cross likelihood ratio (CLR),[10] and the generalized likelihood ratio (GLR),[17,22,7] have been examined and compared in much of the literature, with GLR being the most prevalent similarity measure. The generation of a cluster tree can be performed in either a bottom-up (agglomerative) or a top-down (divisive) fashion, according to some criteria drawn from the similarity measure. The bottom-up approach starts with each utterance as a single cluster, and then successively merges the most similar pairs of clusters until one cluster contains all the utterances. The similarities between clusters are usually derived from the interutterance similarities, based on so-called *complete linkage, single linkage*, or *average linkage*. In the top-down approach, all utterances start in a single cluster, which is split into two dissimilar clusters. This procedure is then repeated for each of the dissimilar clusters, until each cluster contains exactly one utterance. The resulting cluster tree is then cut via an estimation of the number of clusters to retain the best partitioning. Representative methods for estimating the optimal number of clusters are based on the BBN Metric[7] and the Bayesian information criterion.[19]

In essence, the effectiveness of a speaker clustering system depends on whether or not the generated clusters are related to speakers, rather than other acoustic classes. In the hierarchical clustering framework, interutterance similarity computation plays a crucial role in determining if the clusters are formed on the basis of speakers. However, existing similarity measures, based on AHS, KL distance, CLR, or GLR, are performed entirely on spectrum-based features, such as Mel-scale frequency cepstral coefficients (MFCCs) and perceptual linear prediction (PLP) cepstral coefficients. These features are known to carry various types of information besides a speaker's voice characteristics, for example, phonetic and environmental information. Although feature normalization techniques, such as cepstral mean subtraction[26] and RASTA,[27] may be applied to alleviate the interference from the channels and noise, these techniques also run the risk of removing the target speakers' voice characteristics, especially when the utterances are short. As a result, there is no guarantee that the similarities between same-speaker utterances will always be larger than the similarities between different-speaker utterances. Since interutterance similarity computation is independent of cluster tree generation, and the latter trusts the former completely, the inevitable errors of interutterance similarity can propagate down the whole process, which severely limits the clustering performance.

To compensate for the imperfection of interutterance similarity computation, more sophisticated speaker-clustering methods[11,7] have tried to improve the measurement of intercluster similarities by concatenating all the utterances within each cluster into one long utterance and then computing the similarities between long utterances. Analogously, recent speaker-diarization systems, such as Refs. 28 and 29, further apply modeling and matching techniques in speaker identification to evaluate intercluster similarities. Specifically, each cluster is represented as a Gaussian mixture model (GMM) by using the so-called GMM-UBM method.[3] Then, the similarity between a pair of clusters is computed by accumulating the likelihoods of one cluster's utterances testing against another cluster's model. On the other hand, in the context of acoustic model adaptation for speech recognition, Ref. 30 proposes using a MLLR-adapted likelihood as a criterion, instead of the intercluster similarity measurement, to determine which speech data should be grouped together and handled by the same MLLR transforms, such that the mismatch between speech-recognition models and test data can be minimized. In addition, Ref. 30 and the subsequent Cambridge speaker-diarization systems[28] use gender and bandwidth classification to preprocess speech utterances, which allows data from different-gender speakers or different types of channel to be processed separately, thereby reducing the confusion as well as the load on clustering. Meanwhile, in Ref. 29, a feature warping technique[31] is used to further reduce the effects of the acoustic environment.

However, one unresolved problem in most existing systems is the propagation of errors in hierarchical clustering. Taking agglomerative clustering as an example, during the merging process, the utterances from different speakers may be misgrouped into a cluster. Since the misgrouped utterances will never be separated in the subsequent merging operations, such errors will proliferate as more clusters are merged. On the other hand, cluster tree generation based on either top-down or bottom-up hierarchical clustering usually uses a certain neighborhood selection rule, e.g., nearest or furthest neighbor, to determine which utterances should be assigned to which clusters. However, the neighborhood selection rule is applied in a cluster-by-cluster or pairwise manner, rather than in a global manner that considers all the clusters simultaneously. As a consequence, hierarchical clustering can only make each individual cluster as homogeneous as possible, but cannot attain the ultimate goal of maximizing the overall homogeneity.

To overcome the limitations of the hierarchical speaker-clustering framework, this study proposes a new clustering method, with the goal of finding the best partitioning of speech utterances by integrating interutterance similarity computation and cluster tree generation into a unified process. The process iteratively assigns speech utterances to a set of clusters and creates a stochastic model for each cluster, which attempts to maximize the similarity or agreement between each cluster model and the within-cluster utterances. In contrast to a similar idea proposed in Ref. 30, which uses a top-down split-and-merge framework to achieve the goal of the maximum likelihood of adapted data, we apply a *genetic*

algorithm,[32] together with a model similarity comparison method, to search for the best partitioning. In addition, the proposed clustering method further adapts the Bayesian information criterion to determine how many clusters should be created.

The remainder of this paper is organized as follows. Section II reviews a specific implementation of hierarchical clustering, which is the most popular method of speaker clustering. Section III introduces our proposed speaker clustering method, called maximum likelihood clustering, with the goal of maximizing the within-cluster homogeneity of voice characteristics. In Sec. IV, we present an alternative speaker-clustering solution, called minimum divergence clustering, which aims to improve the efficiency of maximum likelihood clustering. Section V discusses the problem of how to automatically determine the appropriate number of clusters. Section VI summarizes the configuration of our speaker-clustering system and analyzes its computational complexity. Section VII presents our experimental results. Finally, in Sec. VIII, we present our conclusions and discuss the direction of future works.

II. REVIEW OF HIERARCHICAL CLUSTERING

To cluster speech utterances by speaker, it is necessary to distinguish between utterances belonging to the same speaker and those belonging to different speakers. A common strategy for this process is to measure the similarities of voice characteristics between utterances and then determine which utterances are similar enough to be considered as being from the same speaker. This section details a specific implementation of this strategy, which serves as a baseline solution in the current study.

A. Interutterance similarity computation

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N$ denote N speech utterances to be clustered, each of which is represented by a certain spectrum-based feature, e.g., the cepstral feature. The similarities between utterances are measured on the basis of the generalized likelihood ratio (GLR).[17] For any pair of utterances, $\mathbf{X}_n$ and $\mathbf{X}_k$, the GLR is computed by

$$\mathrm{GLR}(\mathbf{X}_n, \mathbf{X}_k) = \frac{\Pr(\mathbf{X}_n|\lambda_{nk})\Pr(\mathbf{X}_k|\lambda_{nk})}{\Pr(\mathbf{X}_n|\lambda_n)\Pr(\mathbf{X}_k|\lambda_k)}, \tag{1}$$

or, equivalently,

$$\begin{aligned}\mathrm{GLR}(\mathbf{X}_n, \mathbf{X}_k) = {} & \log \Pr(\mathbf{X}_n|\lambda_{nk}) + \log \Pr(\mathbf{X}_k|\lambda_{nk}) \\ & - \log \Pr(\mathbf{X}_n|\lambda_n) - \log \Pr(\mathbf{X}_k|\lambda_k),\end{aligned} \tag{2}$$

where λ_n, λ_k, and λ_{nk} are stochastic models, e.g., Gaussian mixture models (GMMs), trained using $\mathbf{X}_n$, $\mathbf{X}_k$, and a concatenation of $\mathbf{X}_n$ and $\mathbf{X}_k$, respectively. These stochastic models are designed to capture the relevant aspects of voice characteristics underlying speech utterances. Implicit in Eqs. (1) and (2) is the presumption that if utterances $\mathbf{X}_n$ and $\mathbf{X}_k$ are from the same speaker, model λ_{nk} should be able to cover the voice characteristics of the individual utterances appropriately; hence, the likelihood probabilities $\Pr(\mathbf{X}_n|\lambda_{nk})$ and $\Pr(\mathbf{X}_k|\lambda_{nk})$ would be large, compared to the case where utterances $\mathbf{X}_n$ and $\mathbf{X}_k$ are from different speakers. This gives a large value of $GLR(\mathbf{X}_n, \mathbf{X}_k)$ when utterances $\mathbf{X}_n$ and $\mathbf{X}_k$ are from the same speaker, and a small value otherwise.

B. Cluster generation

After computing the interutterance similarities, the next step is to assign the utterances deemed similar to each other to the same cluster. This is commonly done by an agglomerative hierarchical clustering method,[33] which consists of the following procedure:

1. `begin initialize` $M \leftarrow N$, `and form clusters` $c_i \leftarrow \{\mathbf{X}_i\}, i=1,2,\ldots,N$
2. `do`
3. `find the most similar pair of clusters, say` c_i `and` c_j
4. `merge` c_i `and` c_j
5. $M \leftarrow M-1$
6. `until` $M=1$
7. `end`

The similarities between a pair of clusters, say c_i and c_j, can be derived from the interutterance similarities, according to one of the following heuristic measures:

(1) complete linkage:

$$\mathcal{S}(c_i, c_j) = \min_{\mathbf{X}_n \in c_i, \mathbf{X}_k \in c_j} \mathrm{GLR}(\mathbf{X}_n, \mathbf{X}_k), \tag{3}$$

(2) single linkage:

$$\mathcal{S}(c_i, c_j) = \max_{\mathbf{X}_n \in c_i, \mathbf{X}_k \in c_j} \mathrm{GLR}(\mathbf{X}_n, \mathbf{X}_k), \tag{4}$$

or

(3) average linkage:

$$\mathcal{S}(c_i, c_j) = \frac{1}{\#(i,j)} \sum_{\mathbf{X}_n \in c_i, \mathbf{X}_k \in c_j} \mathrm{GLR}(\mathbf{X}_n, \mathbf{X}_k), \tag{5}$$

where $\#(i,j)$ denotes the number of utterance pairs involved in the summation. Alternatively, the similarities between clusters can be measured by concatenating all the utterances within each cluster into a long utterance, and then computing the GLR between the concatenated utterances. The outcome of the agglomeration procedure is a cluster tree. The final partition of the utterances is then determined by pruning the tree that only has the desired number of leaves left.

III. MAXIMUM LIKELIHOOD CLUSTERING (MLC)

Although the above hierarchical clustering method is popular for speaker clustering, it is far from optimal in a number of respects. First, the similarities between utterances are measured in a pairwise manner, which only considers information about one pair of utterances at a time, and ignores the fact that out-of-pair information can benefit similarity computation for every pair of utterances. Obviously, a better solution would be to characterize the similarities between all the utterances to be clustered in a global fashion, rather than in a piecemeal manner. Second, hierarchical clustering only attempts to make the voice characteristics within

a newly generated cluster as homogeneous as possible. However, it cannot guarantee that the homogeneity for all the clusters can be summed to reach a maximum, since its decision does not consider the interaction between the new cluster to be generated and existing clusters. Consequently, some misclustering errors, arising from grouping different-speaker utterances together, can propagate down the whole process and hence limit the clustering performance. To overcome these shortcomings, we present a new clustering method based on the integration of similarity computation and cluster generation, which aims to maximize overall within-cluster homogeneity.

A. Principle

Recall that in GLR-based similarity measurement, it is assumed that the voice characteristics of a pair of utterances can be well represented by using a single model instead of two utterance-individual models, if both utterances are from the same speaker. Likewise, if several utterances are from the same speaker, they can be pooled to form a joint model without distorting their individual voice characteristics. In other words, if a model trained using a group of utterances is capable of characterizing the utterances consistently well, then these utterances are very likely produced by the same speaker. Therefore, we can formulate speaker clustering as a problem of determining which utterances should be grouped together such that the resultant models can best characterize the grouped utterances.

The proposed method begins by specifying a certain number of clusters to be generated. For any given number of clusters, M, the task of speaker clustering is to assign N utterances $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N$ to M clusters $c_1, c_2, \ldots, c_M$. Let g_n denote the index of the cluster that an utterance, $\mathbf{X}_n$, is assigned to, where g_n is an integer between 1 and M. The goal of optimal clustering is, therefore, to produce a set of cluster indices, $\mathbf{G}^* = \{g_1^*, g_2^*, \ldots, g_N^*\}$, satisfying $g_n^* = g_k^*$ for any utterances $\mathbf{X}_n$ and $\mathbf{X}_k$ from the same speaker, and $g_n^* \neq g_k^*$ for utterances $\mathbf{X}_n$ and $\mathbf{X}_k$ from different speakers. Toward this end, we first create a Gaussian mixture model $\lambda^{(m)}$ for each cluster $c_m, 1 \leq m \leq M$, by using all the feature vectors of the utterances assigned to c_m. Then, a certain level of agreement that the utterances assigned to the same cluster come from the same speaker is characterized by computing the likelihood probability $\Pr(\mathbf{X}_n|\lambda^{(m)})$ for every $g_n = m$. Conceivably, the larger the value of $\Pr(\mathbf{X}_n|\lambda^{(m)})$, the more suitable cluster c_m will be for utterance $\mathbf{X}_n$. Thus, by taking the likelihood probabilities for all the utterances into account, $\mathbf{G}^*$ can be determined by

$$\mathbf{G}^* = \underset{\mathbf{G}}{\operatorname{argmax}} \sum_{m=1}^{M} \sum_{n=1}^{N} \log \Pr(\mathbf{X}_n|\lambda^{(m)}) \delta(g_n, m), \tag{6}$$

where $\delta(\cdot)$ is a Kronecker delta function. We refer to this process as maximum likelihood clustering (MLC). As Eq. (6) is equivalent to

$$\mathbf{G}^* = \underset{\mathbf{G}}{\operatorname{argmax}} \sum_{m=1}^{M} \sum_{n=1}^{N} [\log \Pr(\mathbf{X}_n|\lambda^{(m)}) - \log \Pr(\mathbf{X}_n|\lambda_n)] \delta(g_n, m), \tag{7}$$

in which the term $\log \Pr(\mathbf{X}_n|\lambda_n)$ is a constant that is independent of clustering, MLC can be viewed as the maximization of the overall within-cluster GLRs, given a certain number of clusters.

B. Optimization via the genetic algorithm

Although the solution to Eq. (6) exists, no close form can be derived from this equation directly. Moreover, since the cluster indices are not scalar objects, we cannot use a gradient-based optimization in this scenario. It is also infeasible to perform an exhaustive search, which examines all possible solutions to determine the best one, because there are M^N possible combinations of cluster indices, and this task is an NP-complete problem. Recognizing these difficulties, we propose applying the genetic algorithm (GA)[32] to find $\mathbf{G}^*$ by virtue of its global scope and parallel searching power.

The basic operation of the GA is to explore a given search space in parallel by means of iterative modification of a population of chromosomes. Each chromosome, encoded as a string of alphabets or real numbers called genes, represents a potential solution to a given problem. In our task, a chromosome is exactly a legitimate $\mathbf{G}$, and a gene corresponds to a cluster index associated with an utterance. However, since the index of one cluster can be interchanged with that of another cluster, multiple chromosomes may reflect an identical clustering result. For example, the chromosomes $\{1,1,1,2,2,3,3\}$, $\{1,1,1,3,3,2,2\}$, $\{2,2,2,1,1,3,3\}$, $\{2,2,2,3,3,1,1\}$, $\{3,3,3,2,2,1,1\}$, and $\{3,3,3,1,1,2,2\}$ represent the same clustering result of grouping seven utterances into three clusters. Such a nonunique representation of the solution would significantly increase the GA search space and could lead to an inferior clustering result. To avoid this problem, we limit the inventory of chromosomes to conform to a *baseform* representation defined as follows.

Let $\mathcal{I}(c_m)$ be the lowest index of the utterance in the mth cluster, $c_m = \{\mathbf{X}_i | g_i = m, 1 \leq i \leq N\}$. A chromosome is a baseform iff

$$\forall c_m \text{ and } c_l, \quad \text{if } m < l, \quad \text{then } \mathcal{I}(c_m) < \mathcal{I}(c_l). \tag{8}$$

As the above example shows, chromosome $\{1,1,1,2,2,3,3\}$ is a baseform, since the lowest indices of the utterances in the first, second, and third clusters are 1, 4, and 6, respectively, which satisfies Eq. (8). In contrast, chromosome $\{1,1,1,3,3,2,2\}$ is not a baseform, since the lowest indices of the utterances in the first, second, and third clusters are 1, 6, and 4, respectively, which does not satisfy Eq. (8). Likewise, the other chromosomes, $\{2,2,2,1,1,3,3\}$, $\{2,2,2,3,3,1,1\}$, $\{3,3,3,2,2,1,1\}$, and $\{3,3,3,1,1,2,2\}$ are not baseforms. Even so, it is conceivable that all the nonbaseform chromosomes could be converted into a unique baseform representation by interchanging the clusters' indices. For example, interchanging index "2" with "1" in chromosome $\{2,2,2,1,1,3,3\}$ gives the baseform $\{1,1,1,2,2,3,3\}$.

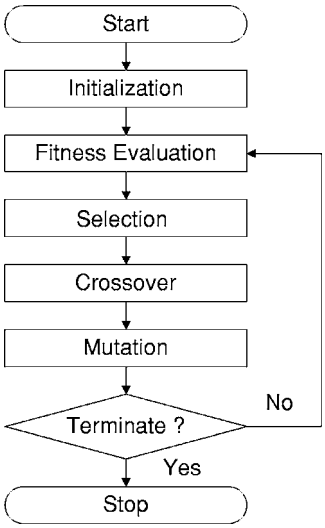


FIG. 1. Flow diagram of the genetic algorithm.

Figure 1 shows a block diagram of GA-based optimization. It starts with a random generation of chromosomes according to a certain population size Z, say 200. The fitness of all chromosomes is then evaluated and ranked on the basis of the overall model likelihood, i.e.,

$$\mathcal{L}(\mathbf{G}) = \sum_{m=1}^{M} \sum_{n=1}^{N} \log \Pr(\mathbf{X}_n|\lambda^{(m)})\delta(g_n,m). \tag{9}$$

As a result of this evaluation, a particular group of chromosomes is selected from the population to generate offspring by subsequent recombination. The selection reflects the fact that chromosomes with superior fitness have a higher probability of being included in the next generation than those that are inferior. To prevent premature convergence of the population, this study employs the linear ranking selection scheme,[34] which sorts chromosomes in increasing order of fitness, and then assigns the expected number of offspring according to their relative ranking. Note that after this operation, chromosomes with large fitness values will produce several copies, while chromosomes with tiny fitness values may be eliminated; hence, the total chromosome population size does not change.

Next, crossover among the selected chromosomes proceeds by exchanging the substrings of two chromosomes between two randomly selected crossover points. For example, the crossover made for chromosomes $\{1,1,\underline{1,2,2,3},3\}$ and $\{1,2,\underline{3,1,2,3},2\}$ generates $\{1,1,3,1,2,3,3\}$ and $\{1,2,1,2,2,3,2\}$, if the selected crossover points are 2 and 6, as the underlines indicate. However, as this example shows, the resulting chromosomes, such as $\{1,1,3,1,2,3,3\}$, may not conform to Eq. (8). Therefore, the procedure for interchanging the clusters' indices has to be performed again to ensure all the offspring are baseforms. In this example, the chromosome $\{1,1,3,1,2,3,3\}$ is converted to $\{1,1,2,1,3,2,2\}$ by swapping index "2" with "3." In addition, a crossover probability is assigned to control the ratio of the number of offspring produced in each generation to the chromosome population size.

After crossover, a mutation operator is used to introduce random variations into the genetic structure of the chromosomes. This is done by generating a legitimate random number and then replacing one gene of an existing chromosome with this random number according to a mutation probability. The resulting chromosomes are converted to the baseform representations again, if necessary. Then, the procedure of fitness evaluation, selection, crossover, and mutation is repeated continuously, following the principle of survival of the fittest, to produce better approximations of the optimal solution. Accordingly, it is hoped that the overall model likelihood will increase from generation to generation. When the maximum number of generations (iterations) Q, say 4000, is reached, the best chromosome in the final population is taken as the solution $\mathbf{G}^*$.

C. MAP estimation of the cluster model

As the above optimization procedure requires the creation of $M \times Z$ GMMs during each GA iteration, the computational complexity can be too high to implement properly if the parameters of the GMMs are estimated via the expectation-maximization (EM) algorithm.[35] To overcome this problem, we propose applying model adaptation techniques to generate cluster GMMs, instead of training them from scratch. The basic strategy, stemming from the GMM-UBM method for speaker recognition,[3] is to create a cluster-independent GMM using all the utterances to be clustered, followed by an adaptation of the cluster-independent GMM for each of the clusters using maximum *a posteriori* (MAP) estimation.

Let $\lambda=\{\omega_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, 1 \le j \le J\}$ denote the parameter set of a cluster-independent GMM having J mixture Gaussian components, where ω_j is the mixture weight, $\boldsymbol{\mu}_j$ is the mean vector, and $\boldsymbol{\Sigma}_j$ is the covariance matrix. These parameters are estimated via the EM algorithm. For each utterance $\mathbf{X}_n$, with T_n feature vectors $\{\mathbf{x}_{n,1}, \mathbf{x}_{n,2}, \ldots, \mathbf{x}_{n,T_n}\}$, we compute the *a posteriori* probability of each feature vector $\mathbf{x}_{n,t}$ in the jth mixture of GMM λ as follows:

$$\Pr(j|\mathbf{x}_{n,t}) = \frac{\omega_j \mathcal{N}(\mathbf{x}_{n;t}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_{l=1}^{J} \omega_l \mathcal{N}(\mathbf{x}_{n,t}; \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)}, \tag{10}$$

where $\mathcal{N}(\cdot)$ is a Gaussian density function. Then, the following parameters are computed and stored in a look-up table:

$$\zeta_{n,j} = \sum_{t=1}^{T_n} \Pr(j|\mathbf{x}_{n,t}), \tag{11}$$

$$E_{n,j}(\mathbf{x}) = \sum_{t=1}^{T_n} \Pr(j|\mathbf{x}_{n,t})\mathbf{x}_{n,t}, \tag{12}$$

$$E_{n,j}(\mathbf{x}\mathbf{x}') = \sum_{t=1}^{T_n} \Pr(j|\mathbf{x}_{n,t})\mathbf{x}_{n,t}\mathbf{x}'_{n,t}, \tag{13}$$

where prime ($'$) denotes a vector transpose. Whenever a set of cluster indices, $g_1, g_2, \ldots, g_N$, is assigned to N utterances, the cluster GMMs, $\lambda^{(m)}=\{\omega_j^{(m)}, \boldsymbol{\mu}_j^{(m)}, \boldsymbol{\Sigma}_j^{(m)}, 1 \le j \le J\}$, $1 \le m \le M$, can be updated by

$$\omega_j^{(m)} = \left[\frac{\tau_j^{(m)}}{\tau_j^{(m)} + \epsilon} \varpi_j^{(m)} + \frac{\epsilon}{\tau_j^{(m)} + \epsilon} \omega_j \right] \vartheta, \tag{14}$$

$$\boldsymbol{\mu}_j^{(m)} = \frac{\tau_j^{(m)}}{\tau_j^{(m)} + \epsilon} E_j^{(m)}(\mathbf{x}) + \frac{\epsilon}{\tau_j^{(m)} + \epsilon} \boldsymbol{\mu}_j, \tag{15}$$

$$\boldsymbol{\Sigma}_j^{(m)} = \frac{\tau_j^{(m)}}{\tau_j^{(m)} + \epsilon} E_j^{(m)}(\mathbf{x}\mathbf{x}') + \frac{\epsilon}{\tau_j^{(m)} + \epsilon} (\boldsymbol{\mu}_j^{(m)} \boldsymbol{\mu}_j'^{(m)} + \boldsymbol{\Sigma}_j) - \boldsymbol{\mu}_j^{(m)} \boldsymbol{\mu}_j'^{(m)}, \tag{16}$$

where ϑ is a scale factor that ensures all the mixture weights sum to unity; ε is a relevance factor that controls how much new data should be observed in a mixture; and $\tau_j^{(m)}$, $\varpi_j^{(m)}$, $E_j^{(m)}(\mathbf{x})$, and $E_j^{(m)}(\mathbf{x}\mathbf{x}')$ are computed using

$$\tau_j^{(m)} = \sum_{n=1}^{N} \zeta_{n,j} \delta(g_n, m), \tag{17}$$

$$\varpi_j^{(m)} = \frac{\tau_j^{(m)}}{\sum_{n=1}^{n} T_n \delta(g_n, m)}, \tag{18}$$

$$E_j^{(m)}(\mathbf{x}) = \frac{1}{\tau_j^{(m)}} \sum_{n=1}^{N} E_{n,j}(\mathbf{x}) \delta(g_n, m), \tag{19}$$

$$E_j^{(m)}(\mathbf{x}\mathbf{x}') = \frac{1}{\tau_j^{(m)}} \sum_{n=1}^{N} E_{n,j}(\mathbf{x}\mathbf{x}') \delta(g_n, m), \tag{20}$$

respectively. Note that although the adaptation can be carried out iteratively, empirical evidence shows that the performance of a single-iteration adaptation is often comparable to that of a multiple-iteration adaptation. Therefore, without relying on the iterative estimation as required by the EM algorithm, MAP-adapted cluster models can be created rapidly whenever a set of cluster indices is reassigned to the utterances.

IV. MINIMUM DIVERGENCE CLUSTERING (MDC)

In addition to training cluster GMMs, another issue concerning the realization of Eq. (6) is the considerable complexity of likelihood computation. Specifically, the standard procedure for computing likelihood $\Pr(\mathbf{X}_n|\lambda^{(m)})$ is

$$\Pr(\mathbf{X}_n|\lambda^{(m)}) = \prod_{t=1}^{T_n} \sum_{j=1}^{J} \omega_j^{(m)} \mathcal{N}(\mathbf{x}_{n,t}; \boldsymbol{\mu}_j^{(m)}, \boldsymbol{\Sigma}_j^{(m)}), \tag{21}$$

which requires $T_n \times J$ computations of Gaussian density $\mathcal{N}(\cdot)$. Thus, each GA iteration involves $Z \times J \times (\Sigma_{n=1}^{N} T_n)$ computations of Gaussian density. When the number of utterances to be clustered is large, the whole clustering process can be extremely time consuming. To overcome this problem, we further propose a clustering method based on an approximation of the likelihood by a computationally more tractable metric, called *divergence*.[36]

Recall that the likelihood $\Pr(\mathbf{X}_n|\lambda^{(m)})$ represents how well the cluster GMM $\lambda^{(m)}$ fits the distribution of the feature vectors of $\mathbf{X}_n$. If we characterize the distribution of the feature vectors of $\mathbf{X}_n$ by utterance GMM λ_n, the computation of $\Pr(\mathbf{X}_n|\lambda^{(m)})$ should be roughly equivalent to a certain similarity measurement between GMMs $\lambda^{(m)}$ and λ_n. Let $\{\omega_{n,i}, \boldsymbol{\mu}_{n,i}, \boldsymbol{\Sigma}_{n,i}, 1 \le i \le J\}$ be the parameters of λ_n estimated via MAP adaptation from GMM λ. The similarity between GMMs $\lambda^{(m)}$ and λ_n can be measured by[37]

$$\mathcal{S}(\lambda^{(m)}, \lambda_n) = \sum_{j=1}^{J} \sum_{i=1}^{J} \omega_j^{(m)} \omega_{n,i} \exp\left[- \mathcal{D}(\boldsymbol{\mu}_j^{(m)}, \boldsymbol{\Sigma}_j^{(m)}; \boldsymbol{\mu}_{n,i}, \boldsymbol{\Sigma}_{n,i}) \right], \tag{22}$$

and

$$\begin{aligned} \mathcal{D}(\boldsymbol{\mu}_j^{(m)}, \boldsymbol{\Sigma}_j^{(m)}; \boldsymbol{\mu}_{n,i}, \boldsymbol{\Sigma}_{n,i}) = {} & \frac{1}{2} (\boldsymbol{\mu}_j^{(m)} - \boldsymbol{\mu}_{n,i})' (\boldsymbol{\Sigma}_j^{(m)^{-1}} + \boldsymbol{\Sigma}_{n,i}^{-1}) \\ & \times (\boldsymbol{\mu}_j^{(m)} - \boldsymbol{\mu}_{n,i}) \\ & + \frac{1}{2} \mathrm{Tr}[(\boldsymbol{\Sigma}_j^{(m)^{1/2}} \boldsymbol{\Sigma}_{n,i}^{-1/2}) \\ & \times (\boldsymbol{\Sigma}_j^{(m)^{1/2}} \boldsymbol{\Sigma}_{n,i}^{-1/2})'] \\ & + \frac{1}{2} \mathrm{Tr}[(\boldsymbol{\Sigma}_j^{(m)^{-1/2}} \boldsymbol{\Sigma}_{n,i}^{1/2}) \\ & \times (\boldsymbol{\Sigma}_j^{(m)^{-1/2}} \boldsymbol{\Sigma}_{n,i}^{1/2})'] - R, \end{aligned} \tag{23}$$

where $\mathcal{D}(\boldsymbol{\mu}_j^{(m)}, \boldsymbol{\Sigma}_j^{(m)}; \boldsymbol{\mu}_{n,i}, \boldsymbol{\Sigma}_{n,i})$ is the divergence between Gaussian distributions $\mathcal{N}(\boldsymbol{\mu}_j^{(m)}, \boldsymbol{\Sigma}_j^{(m)})$ and $\mathcal{N}(\boldsymbol{\mu}_{n,i}, \boldsymbol{\Sigma}_{n,i})$; $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix; and R is the dimension of the feature vectors. Note that the divergence can also be replaced by other measurements between two Gaussian densities, such as arithmetic harmonic sphericity and arithmetic geometric sphericity, which are discussed in Refs. 38 and 39.

For greater computational efficiency, we keep the mixture weights unchanged during MAP adaptation, i.e., $\omega_j^{(m)} = \omega_{n,j} = \omega_j, 1 \le j \le J$. Since the mixture components of $\lambda^{(m)}$ and λ_n are aligned, Eq. (22) can be simplified as

$$\mathcal{S}(\lambda^{(m)}, \lambda_n) = \sum_{j=1}^{J} \omega_j \exp\left[- \mathcal{D}(\boldsymbol{\mu}_j^{(m)} \boldsymbol{\Sigma}_j^{(m)}; \boldsymbol{\mu}_{n,j}, \boldsymbol{\Sigma}_{n,j}) \right]. \tag{24}$$

Note that $0 \le \mathcal{S}(\lambda^{(m)}, \lambda_n) \le 1$, in which the upper bound reflects that $\lambda^{(m)}$ and λ_n are identical. A large value of $\mathcal{S}(\cdot)$ signifies a high degree of homogeneity between the utterances within a cluster. Thus, speaker clustering can be converted into a problem of finding a set of cluster indices $\mathbf{G}^* = \{g_1^*, g_2^*, \ldots, g_N^*\}$ that satisfies

$$\mathbf{G}^* = \underset{\mathbf{G}}{\mathrm{argmax}} \sum_{m=1}^{M} \sum_{n=1}^{N} \log \mathcal{S}(\lambda^{(m)}, \lambda_n) \delta(g_n, m). \tag{25}$$

We refer to this clustering method as *minimum divergence clustering* (MDC). Since Eq. (24) is not dependent on the length of utterance, the computational complexity can be dramatically reduced, compared to that of MLC. If the covariance matrices are set to be diagonal, the computational complexity is approximately reduced by the factor $(\Sigma_{n=1}^{N} T_n)/N$, i.e., the average length of utterance.

V. ESTIMATION OF THE NUMBER OF SPEAKERS

The proposed speaker-clustering methods described above are based on specifying a certain number of clusters to be generated in advance. In general, the greater the number of clusters specified, the higher the level of homogeneity within a cluster. However, if too many clusters are generated, a single speaker's utterances would spread over multiple clusters; hence, the speaker clustering would not be complete. Clearly, the optimal number of clusters is equal to the speaker population size, which is unknown and must be estimated.

Consider a collection of N speech utterances to be partitioned into M clusters. The optimal value of M must be an integer between 1 and N. Thus, if we produce a set of possible partitionings, in which the number of clusters ranges from 1 to N, the task of determining the optimal value of M would amount to selecting one of the N partitionings that achieves the level of within-cluster homogeneity as high as possible with the number of clusters as small as possible. To realize such a selection, we adapt the Bayesian information criterion (BIC)[40] to score each of the possible partitionings, thereby identifying the best one.

The BIC is a model selection criterion that assigns a value to a parametric model based on how well the model fits a data set, and how simple the model is:

$$\mathrm{BIC}(\Lambda) = \log \Pr(\mathbf{O}|\Lambda) - \frac{1}{2}\gamma \#(\Lambda) \log |\mathbf{O}|, \tag{26}$$

where γ is a penalty factor generally equal to one, $\#(\Lambda)$ denotes the number of free parameters in model Λ, and $|\mathbf{O}|$ is the size of the data set $\mathbf{O}$. The larger the value of BIC (Λ), the better model Λ will perform.

By treating each of the possible partitionings as a model for characterizing speaker information in the utterances, we can evaluate a partitioning with M clusters via the following BIC-motivated score:

$$\mathcal{B}(M) = \sum_{m=1}^{M}\sum_{n=1}^{N} \log \mathcal{S}(\lambda^{(m)^*}, \lambda_n)\delta(g_n^*, m) - \frac{1}{2}\gamma M \log N, \tag{27}$$

where g_n^* denotes the index of the cluster in which utterance $\mathbf{X}_n$ is located according to the GA optimization for Eq. (25), and $\lambda^{(m)^*}$ is the resulting GMM of cluster c_m after optimization. In Eq. (27), we use the divergence-based similarity measurement, $\Sigma_{m=1}^{M}\Sigma_{n=1}^{N} \log \mathcal{S}(\lambda^{(m)^*}, \lambda_n)\delta(g_n^*, m)$, to represent how well the model fits the data. This approximates the log probability $\log \Pr(\mathbf{O}|\Lambda)$. Here, the data is actually a set of N utterance GMMs, which is further "modeled" by M cluster GMMs, if M clusters are generated. Hence, the size of data can be considered as the number of utterance GMMs, i.e., $|\mathbf{O}| = N$, which does not depend on the utterance duration. Moreover, since the configuration of the data (utterance GMMs) are the same as that of the model (cluster GMMs), the number of free parameters in model Λ can be considered independent of the number of Gaussian densities used and the dimensionality of feature vectors. This indicates that $\#(\Lambda) \approx M$.

The value of $\mathcal{B}(M)$ should increase with the increase in the value of M initially, but it will decline significantly after an excess of clusters is created. Thus, a reasonable number of clusters can be determined by choosing the partitioning that produces the largest value of $\mathcal{B}(M)$, i.e.,

$$M^* = \arg \max_{1 \le M \le N} \mathcal{B}(M). \tag{28}$$

Note that in the original definition of $\mathrm{BIC}(\Lambda)$, the term $\log \Pr(\mathbf{O}|\Lambda)$ can also be represented by $\mathcal{L}(\mathbf{G}^*)$ obtained with MLC. However, due to the high computational complexity, we find that MLC is unsuitable for the scenario that the true speaker population size is unknown and needs to be estimated. On the other hand, in a pioneering work[19] on the use of BIC for speaker clustering, the generation of clusters is performed via the aforementioned GLR-based similarity computation, followed by hierarchical clustering. Each of the resulting clusters is then represented by a uni-Gaussian density estimated by using the feature vectors of the within-cluster utterances; hence, the model, Λ, is a set of Gaussian densities. Since we have characterized every cluster by a GMM, our work differs from Ref. 19 in that the proposed model Λ is optimized during the generation of clusters and directly reflects the overall homogeneity of within-cluster utterances.

VI. SYSTEM CONFIGURATION AND ANALYSIS OF THE COMPUTATIONAL COMPLEXITY

Figure 2 summarizes the implementation of our speaker-clustering system. In the absence of knowing the true speaker population size, the system in turn hypothesizes that N utterances to be clustered can be from one speaker, two speakers, …, or N speakers. For each of the possible speaker population sizes, MDC is run with the number of clusters specified as the hypothesized speaker population size. This yields N partitionings optimized by GA, along with N BIC-motivated scores. The system then outputs the partitioning associated with the largest BIC-motivated score.

In view of the usability, it is worth comparing the computational complexity of the proposed clustering system with that of a GLR-based hierarchical clustering system. We observe that there are two factors which dominate the overall computational time for both systems. The first factor arises from the Gaussian mixture modeling of feature vectors, e.g., the generation of λ_{nk} in Eq. (1), or the generation of cluster model $\lambda^{(m)}$ in Eq. (25). The second factor arises from the computation of Gaussian functions based on the models, e.g., $\Pr(\mathbf{X}_n|\lambda_{nk})$, or $\mathcal{S}(\lambda^{(m)}, \lambda_n)$. However, if the models are generated using the MAP adaptation, the first factor can be ignored, compared to the second factor. Hence, the system complexity depends mainly on how many Gaussian functions need to be performed.

Consider an agglomerative hierarchical clustering system, in which the intercluster similarities are measured on an utterance-concatenation basis. If the system has generated $M+1$ clusters $c_1, c_2, \ldots, c_{M+1}$, and is going to determine which pair of clusters can be merged, it requires us to compute M likelihood probabilities for every "long utterance" formed by concatenating all the utterances within a cluster.

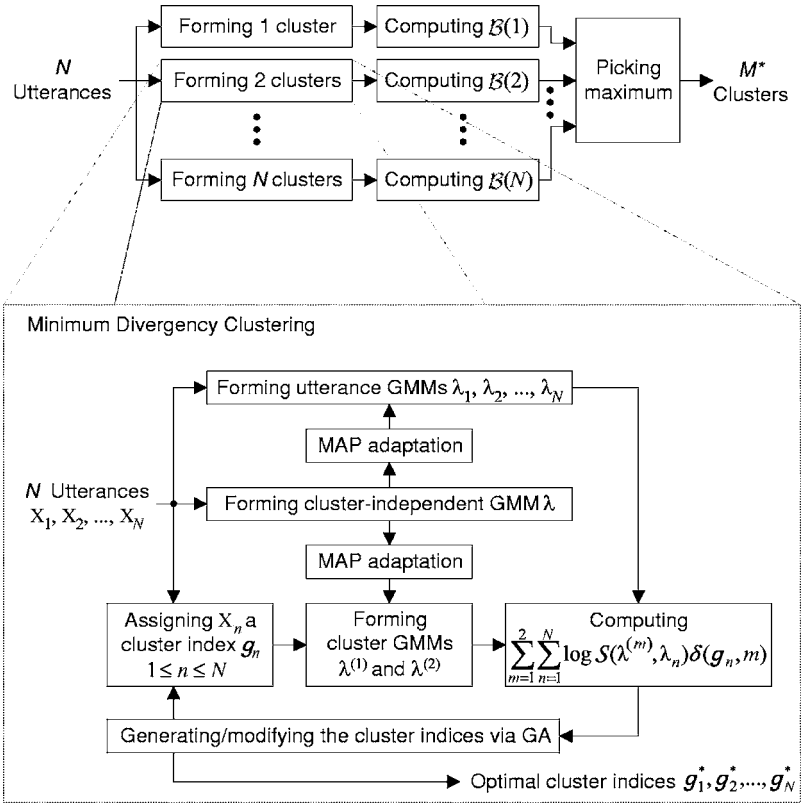


FIG. 2. Block diagram of the proposed speaker-clustering system.

The likelihood probability is concerned with the term $\Pr(\mathbf{X}_n|\lambda_{nk})$ in Eq. (1), where $\mathbf{X}_n$ represents the long utterance for c_n, and λ_{nk} represents the GMM trained using two long utterances $\mathbf{X}_n$ and $\mathbf{X}_k$. For example, to determine which two among the three long utterances $\mathbf{X}_1, \mathbf{X}_2$, and $\mathbf{X}_3$ can be merged, the system requires us to compute two likelihood probabilities $\Pr(\mathbf{X}_1|\lambda_{12})$ and $\Pr(\mathbf{X}_1|\lambda_{13})$ for $\mathbf{X}_1$, two likelihood probabilities $\Pr(\mathbf{X}_2|\lambda_{12})$ and $\Pr(\mathbf{X}_2|\lambda_{23})$ for $\mathbf{X}_2$, and two likelihood probabilities $\Pr(\mathbf{X}_3|\lambda_{13})$ and $\Pr(\mathbf{X}_3|\lambda_{23})$ for $\mathbf{X}_3$. Note that though Eq. (1) contains another likelihood probability $\Pr(\mathbf{X}_n|\lambda_n)$, such a term has been computed already in the last merging stage, i.e., merging $M+2$ clusters into $M+1$ clusters. Therefore, to derive M clusters from $M+1$ clusters, the system needs to perform $M \times J \times T$ Gaussian functions, where J is the number of Gaussian mixtures as denoted earlier, and T is the total number of feature vectors in the utterance collection. To complete a cluster tree with the numbers of clusters ranging from 1 to N, the system requires us to perform $\Sigma_{m=1}^{(N-1)} M \times J \times T = \frac{1}{2}N(N-1)JT$ Gaussian functions. Accordingly, the computational complexity of the GLR-based hierarchical clustering system can be characterized by $\mathcal{O}(\frac{1}{2}N(N-1)JT) \simeq \mathcal{O}(\frac{1}{2}N^2JT)$.

With regard to the proposed speaker-clustering system, whenever a chromosome is generated for assigning each utterance a cluster index, the system needs to compute N divergency-based similarities $\mathcal{S}(\cdot)$. Since $\mathcal{S}(\cdot)$ involves J Gaussian functions, optimizing M clusters via GA requires to perform $N \times J \times Z \times Q$ Gaussian functions, where Z and Q are, as denoted earlier, the number of chromosomes and the maximum number of generations, respectively. Thus, to determine the optimal number of clusters, a scan from $M=1$ to $M=N$ involves computational complexity around $\mathcal{O}(N^2JZQ)$. We can see that the proposed system requires $2ZQ/T$ times the computational complexity of the GLR-based hierarchical clustering system. However, if the value of ZQ is set to $T/2$, the two systems have similar computational complexities.

VII. EXPERIMENTS

A. Speech data

Our speech data were extracted from two corpora released by the Linguistic Data Consortium[41]: *the 1998 HUB-4 Broadcast News Evaluation English Test Material* (Hub4-98), which consists of broadcast news speech recorded at a 16 kHz sampling rate, and *the 2001 NIST Speaker Recognition Evaluation Corpus* (SRE-01), which consists of cellular telephone speech recorded at an 8 kHz sampling rate. The data were divided into three subsets. The first contained 399 speaker-homogeneous utterances obtained by segmenting the episode "h4e-98-l" of Hub4-98, according to the annotation file. This subset involved 79 speakers, in which the number of utterances spoken by each speaker ranged from 1 to 48. The second subset contained 428 speaker-homogeneous utterances obtained by also segmenting the episode "h4e-98-2" of Hub4-98, according to the annotation file. There were 89 speakers in this subset, and the number of utterances spoken by each speaker ranged from 1 to 27. The third subset, which stems from the test set of SRE-01, contained 197 speaker-homogeneous utterances spoken by 15 randomly selected male speakers. The number of utterances spoken by each speaker ranged from 5 to 39.

The speaker-clustering methods used in this study were optimized using the utterances in the first subset, and the methods' performances were then evaluated using the utter-

ances in the second and third subsets. Feature vectors, each consisting of 20 MFCCs, were extracted from these utterances for every 20-ms Hamming-windowed frame with 10-ms frame shifts. Prior to MFCC computation, voice active detection[42] was applied to remove salient nonspeech regions that may be included in an utterance. The total nonsilence numbers of feature vectors for "h4e-98-1," "h4e-98-2," and "SRE-01" were 551 019 frames, 545 700 frames, and 418 625 frames, respectively.

B. Performance evaluation metrics

The performance of speaker clustering was evaluated on the basis of two metrics: cluster purity,[7,43,44] and the Rand index.[7,45,46] Cluster purity is the probability that if we pick any utterance from a cluster twice at random, with replacement, both of the selected utterances are from the same speaker. Specifically, the purity of cluster c_m is computed by

$$\rho_m = \sum_{p=1}^{P} \frac{n_{mp}^2}{n_{m^*}^2}. \tag{29}$$

where n_{m^*} is the total number of utterances in cluster c_m, n_{mp} is the number of utterances in cluster cm produced by the pth speaker, and P is the total number of speakers. From Eq. (29), it follows that $n_{m^*}^{-1} \leq \rho_m \leq 1$, in which the upper bound and lower bound reflect that all the within-cluster utterances were produced by the same speaker and completely different speakers, respectively. To evaluate the overall performance of M-clustering, we compute an average purity

$$\bar{\rho} = \frac{1}{N} \sum_{m=1}^{M} n_{m^*} \rho_m. \tag{30}$$

The Rand index used in this study follows Ref. 7, which indicates the level of disagreement in a partitioning. However, for ease of performance comparison, we represent the disagreement as a probability instead of the number of utterance pairs originally used in Ref. 7. Specifically, the Rand index is defined by the probability that two randomly selected utterances from the same speaker are placed in different clusters, or that two randomly selected utterances placed in the same cluster are from different speakers:

$$\mathcal{R} = \frac{\sum_{m=1}^{M} n_{m^*}^2 + \sum_{p=1}^{P} n_{*p}^2 - 2\sum_{m=1}^{M} \sum_{p=1}^{P} n_{mp}^2}{\sum_{m=1}^{M} n_{m^*}^2 + \sum_{p=1}^{P} n_{*p}^2}, \tag{31}$$

where n_{*p} is the number of utterances from the pth speaker. The lower the value of $\mathcal{R}$, the better the clustering performance. Perfect clustering should produce a Rand index of zero.

Note that the cluster purity and Rand index defined above are calculated without taking the length of utterance into account. However, in many applications, assigning a long utterance into a wrong cluster can be more detrimental than assigning a short utterance into a wrong cluster. To reflect this matter, we further compute the two metrics on the basis of frame correctness. Specifically, a frame-based cluster purity is defined by the probability that if we pick any frame from a cluster twice at random, with replacement, both of the selected frames are from the same speaker. Thus, when computing a frame-based purity by Eq. (29), n_{m^*} represents the total number of frames in cluster c_m, and n_{mp} represents the number of frames in cluster c_m produced by the pth speaker. Likewise, a frame-based Rand index is defined by the probability that two randomly selected frames from the same speaker are placed in different clusters, or that two random selected frames placed in the same cluster are from different speakers. In general, when evaluating a certain clustering result, the value of frame-based purity is larger than that of utterance-based purity, while the value of the frame-based Rand index is smaller than that of the utterance-based Rand index.

C. Experimental results

Our first experiment was conducted to assess the speaker-clustering performance by assuming that the total number of speakers is known; hence, the required number of clusters can be specified as *a priori*. Figures 3 and 4 show the performance of agglomerative hierarchical clustering for subsets "h4e-98-2" and "SRE-01," respectively, in which the numbers of clusters were specified as 89 and 15. We examined different intercluster similarity measures along with GLRs computed with different numbers of component densities in Gaussian mixture modeling. The "concatenation" in Figs. 3 and 4 stands for the intercluster similarity measured by concatenating all the utterances within each cluster into a long utterance, and then computing the GLR between the concatenated utterances. Except for the single-Gaussian models (number of Gaussian mixtures=1), which were full-covariance structures and trained via maximum likelihood estimation, all the GMMs (number of Gaussian mixtures $\geq$2) used in this study comprised diagonal covariance matrices trained via MAP-adaptation from an utterance-independent GMM. We observe from Figs. 3 and 4 that, of the three linkages, complete linkage performs the best, which almost always yields larger values of purity and smaller values of the Rand index than those of the others; single linkage performs the worst, and average linkage is between the two extremes. It can also be seen from the figures that concatenation surpasses complete linkage in terms of the largest value of purity and smallest value of the Rand index that can be produced. However, there were no consistent results that could indicate the optimal number of Gaussian mixtures used in GLR computation.

Figures 5 and 6 show the speaker-clustering results obtained by our proposed methods. Here, GLR-HC concatenation represents concatenation shown in Figs. 3 and 4. In GA optimization, the parameter values used for the maximum number of generations Q, the chromosome population size Z, the crossover probability, and the mutation probability were determined to be 4,000, 200, 0.32, and 0.2, respectively, according to the test on subset h4e-98-1. We can see from Figs. 5 and 6 that both MLC and MDC yield larger values of purity and smaller values of the Rand index than most GLR-HC concatenation cases can attain. Table I summarizes the individual best speaker-clustering performance that MLC, MDC, and GLR-HC concatenation achieved, in which

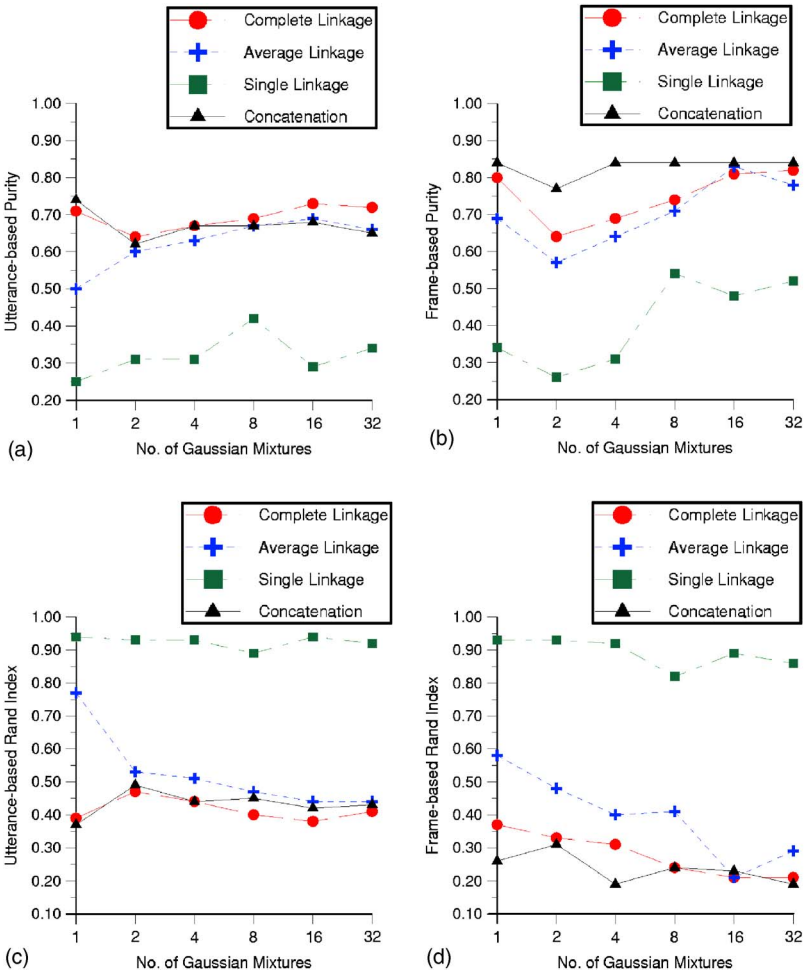


FIG. 3. (Color online) Performance of agglomerative hierarchical clustering for subset "h4e-98-2," in which the number of clusters was specified as the true size of speaker population, i.e., 89. Except for the single-Gaussian models (no. of Gaussian mixtures=1), which were full-covariance structures and trained via maximum likelihood estimation, all the GMMs (no. of Gaussian mixtures ≥ 2) comprised diagonal covariance matrices trained via MAP adaptation. (a) Utterance-based purity. (b) Frame-based purity. (c) Utterance-based Rand index. (d) Frame-based Rand index.

the numbers in parentheses indicate the number of component densities used in Gaussian mixture modeling. It is clear from Table I that the proposed methods are superior to the hierarchical clustering method.

In addition, we observe from Figs. 5 and 6 that the performance of MLC is slightly better than that of MDC. However, as mentioned earlier, MLC is rather computationally extensive, due to the need to compute Gaussian densities frame by frame. Quantitatively, MLC required 2000 times the computational time of MDC for this clustering task, and took weeks to complete a trial on a 3-GHz Pentium PC. This makes it difficult to use MLC to determine how many clusters should be generated if the number of speakers is not known in advance. Therefore, in the following experiments, we concentrated on examining the validity of MDC-based speaker clustering.

Figures 7 and 8 show the speaker-clustering performance as a function of the number of clusters, in which the numbers of Gaussian mixtures used in subsets h4e-98-2 and SRE-01 were 1 and 32, respectively. We can see that the average purity increases as the number of clusters increases. It is also clear from the figures that MDC consistently yields larger values of purity than those obtained with GLR-HC concatenation, regardless of the number of clusters. On the other hand, we observe that the Rand index decreases with the increase in the number of clusters initially, but increases gradually when too many clusters are generated. In general, the smallest value of the Rand index occurs when the number of clusters is close to the speaker population size. It can be seen from the figures that the smallest value of the Rand index obtained with MDC is not only smaller than that obtained with GLR-HC concatenation, but also located at the number of clusters closer to the true size of speaker population.

To investigate if the optimal number of clusters can be determined automatically, we computed the BIC-motivated scores with respect to different numbers of clusters using Eq. (27). Figure 9 shows the resulting scores obtained with the penalty factor γ set to be equal to, slightly greater than, and slightly smaller than 1, respectively. The arrowed peak of each curve in the figures indicates the optimal number of clusters determined by the criterion of Eq. (28). We can see from the figure that most of the peaks appeared near the actual number of speakers, and the scores declined significantly after an excess of clusters was created. In general, the larger the value of the penalty factor, the smaller the estimated optimal number of clusters, and vice versa. The results show that the number of speakers in subset SRE-01 was estimated very well, whereas the number of speakers in subset h4e-98-2 tends to be underestimated, if the penalty factor is simply set to be 1. We speculate that this underestimation is mainly because, among the total 89 speakers in subset h4e-98-2, there were 30 speakers who spoke only one utterance, and many of these speakers' utterances were shorter

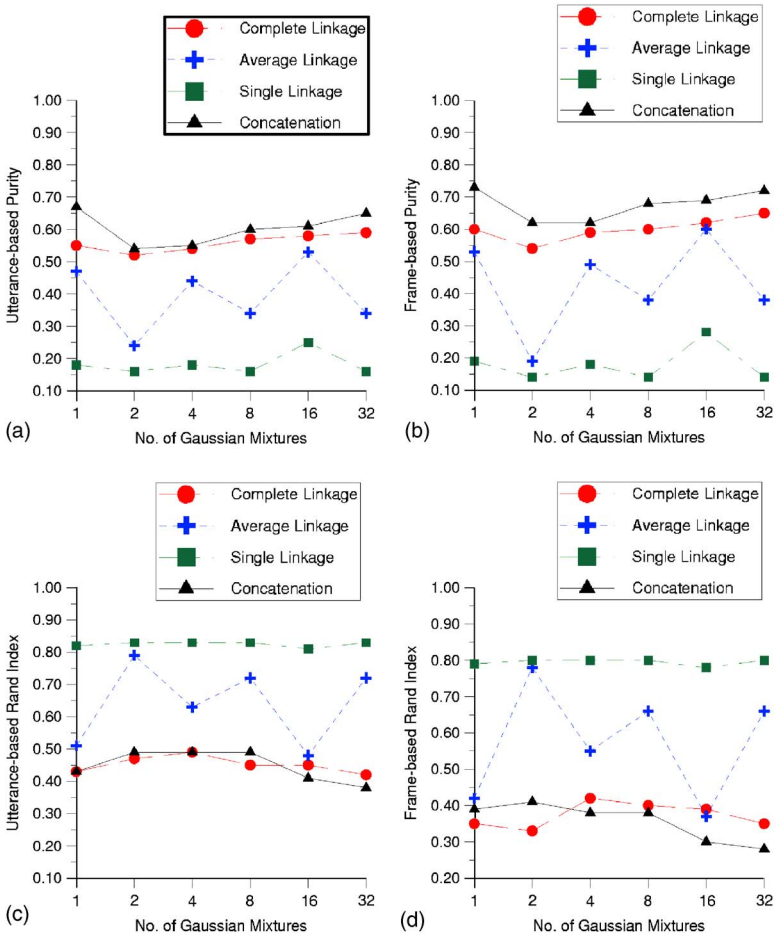


FIG. 4. (Color online) Performance of agglomerative hierarchical clustering for subset "SRE-02" in which the number of clusters was specified as the true size of speaker population, i.e., 15. Except for the single-Gaussian models (no. of Gaussian mixtures=1), which were full-covariance structures and trained via maximum likelihood estimation, all the GMMs (no. of Gaussian mixtures ≥ 2) comprised diagonal covariance matrices trained via MAP adaptation. (a) Utterance-based purity. (b) Frame-based purity. (c) Utterance-based Rand index. (d) Frame-based Rand index.

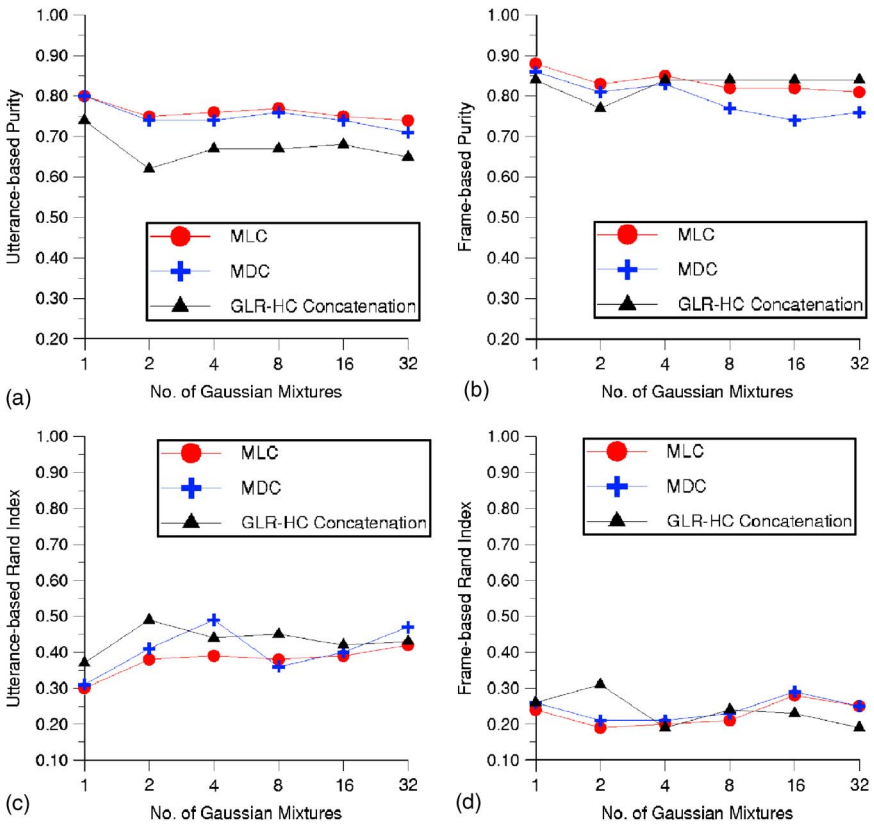


FIG. 5. (Color online) Performance of the proposed maximum likelihood clustering (MLC) and minimum divergence clustering (MDC) for subset "h4e-98-2," in which "GLR-HC concatenation" represents "concatenation" shown in Fig. 2. (a) Utterance-based purity. (b) Frame-based purity. (c) Utterance-based Rand index. (d) Frame-based Rand index.

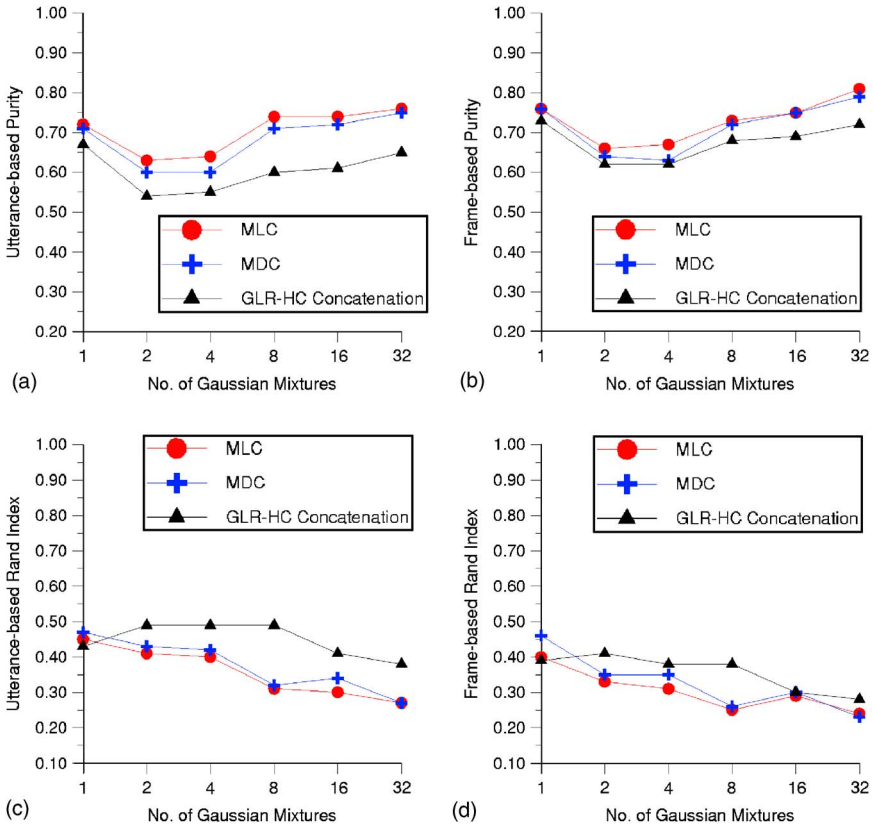


FIG. 6. (Color online) Performance of the proposed maximum likelihood clustering (MLC) and minimum divergence clustering (MDC) for subset "SRE-02," in which "GLR-HC concatenation" represents "concatenation" shown in Fig: 3. (a) Utterance-based purity. (b) Frame-based purity. (c) Utterance-based Rand index. (d) Frame-based Rand index.

than 5 s, which leads to the tendency that these speakers are ignored. Despite this, we observe from the figure that such an underestimation can be mitigated using the penalty factor slightly smaller than 1. This result validates the proposed method for estimating the speaker population size.

We conclude the experiments with a note that, compared to GLR-HC concatenation, the proposed method improves the speaker-clustering performance at the cost of slightly higher computational requirement. Quantitatively, our system took $2ZQ/T \simeq 3$ times the computational time of GLR-HC concatenation, in which $Z=200$, $Q=4000$, and the total number of feature vectors T was roughly 500 000 frames for all the three subsets. However, as the values of Z and Q are adjustable, the proposed system is more flexible than GLR-HC concatenation in terms of computational complexity.

TABLE I. Summary of Figs. 5 and 6, in which the individual best speaker-clustering performance that MLC, MDC, and "GLR-HC concatenation" achieved is listed. The numbers in parentheses indicate the number of component densities used in Gaussian mixture modeling.

	Clustering method		
Evaluation Metric	MLC	MDC	GLR-HC concatenation
(a) Results of clustering subset "h4e-98-2"			
Utterance-based purity	0.80 (1)	0.80 (1)	0.74 (1)
Frame-based purity	0.88 (1)	0.86 (1)	0.84 (1)
Utterance-based Rand index	0.30 (1)	0.31 (1)	0.37 (1)
Frame-based Rand index	0.19 (2)	0.21 (2)	0.19 (4)
(b) Results of clustering subset "SRE-02"			
Utterance-based purity	0.76 (32)	0.75 (32)	0.67 (1)
Frame-based purity	0.81 (32)	0.79 (32)	0.73 (1)
Utterance-based Rand index	0.27 (32)	0.27 (32)	0.38 (32)
Frame-based Rand index	0.24 (32)	0.23 (32)	0.28 (32)

VIII. CONCLUSIONS

This study has investigated methods for partitioning speech utterances into clusters so that the level of homogeneity of within-cluster utterances can be as high as possible. Such homogeneity has been characterized by either the likelihood probability that a cluster model tests for each of the within-cluster utterances, or the divergence-based similarity between a cluster model and each of the within-cluster utterance models. For greater efficiency, both cluster models and utterance models are adapted from a cluster-independent model trained using all the utterances to be clustered. To enable the best partitioning with the maximal amount of within-cluster homogeneity to be found effectively, we have proposed the use of a genetic algorithm to determine the cluster in which each utterance should be placed. Our experimental results show that the proposed method achieves a notable improvement in speaker-clustering performance, compared to the conventional method using GLR-based similarity measurement followed by agglomerative hierarchical clustering. In addition, the proposed clustering method adapts the Bayesian information criterion to determine how many clusters should be generated. The experimental results show that the automatically determined number of clusters approximates the actual speaker population size.

To be of more practical use, our future work will extend the current speaker-clustering methods to deal with speech

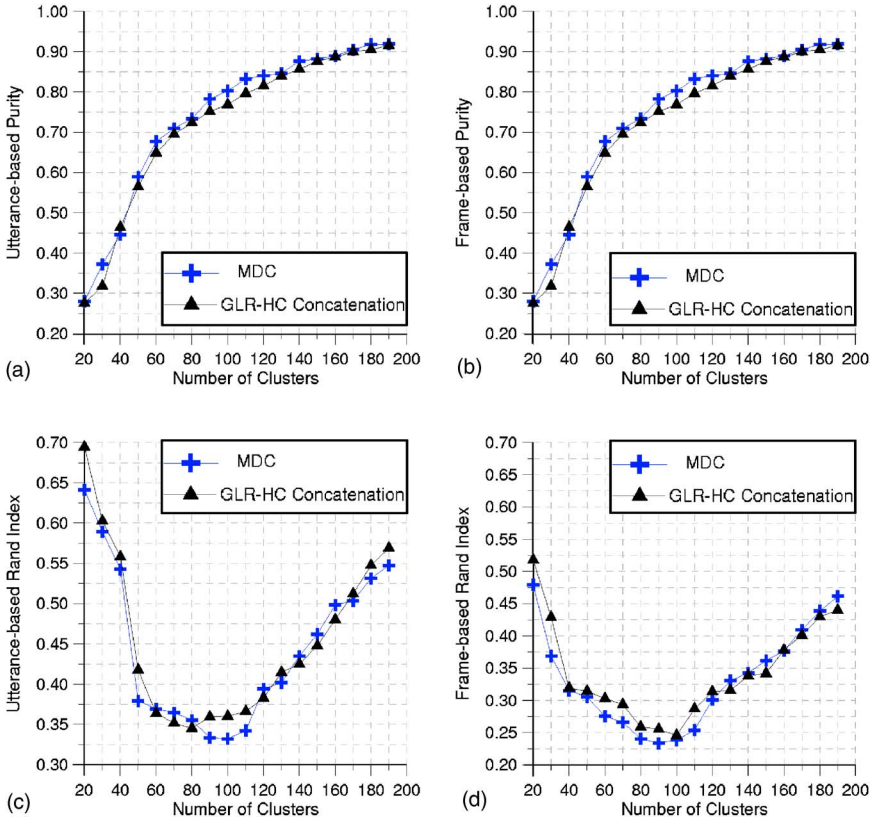


FIG. 7. (Color online) Performance of clustering subset "h4e-98-2" as a function of the number of clusters, in which the number of component densities used in Gaussian mixture modeling was 1 (single-Gaussian model with full covariance matrix). (a) Utterance-based purity. (b) Frame-based purity. (c) Utterance-based Rand index. (d) Frame-based Rand index.

data containing multiple speakers. This could be done by either assigning each utterance to multiple related clusters,[47] or presegmenting utterances into small speaker-homogeneous regions and then clustering those regions. In parallel, speaker segmentation may be improved with the aid of speaker clustering.[12] Specifically, speech segments assigned to each cluster can be used to train a speaker-related model, thereby examining the speaker change boundaries of an audio recording in a manner of frame-by-frame recognition. Speaker clustering can then be performed on the updated speech segments, and the segmentation and clustering procedures repeated iteratively to attain the goal of speaker

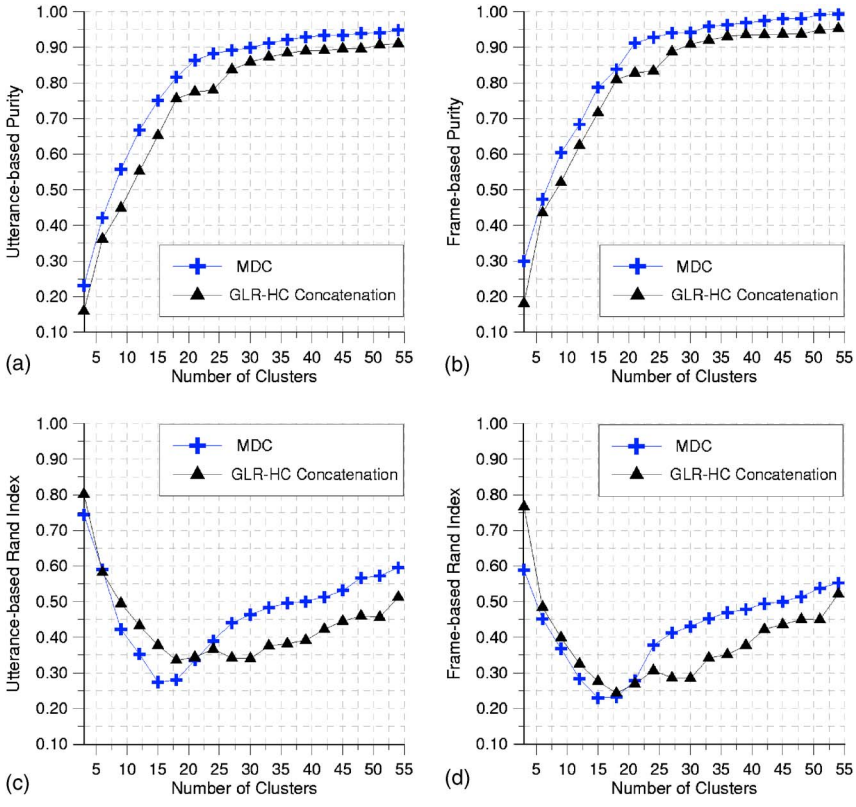


FIG. 8. (Color online) Performance of clustering subset "SRE-02" as a function of the number of clusters, in which the number of component densities used in Gaussian mixture modeling was 32. (a) Utterance-based purity. (b) Frame-based purity. (c) Utterance-based Rand index. (d) Frame-based Rand index.

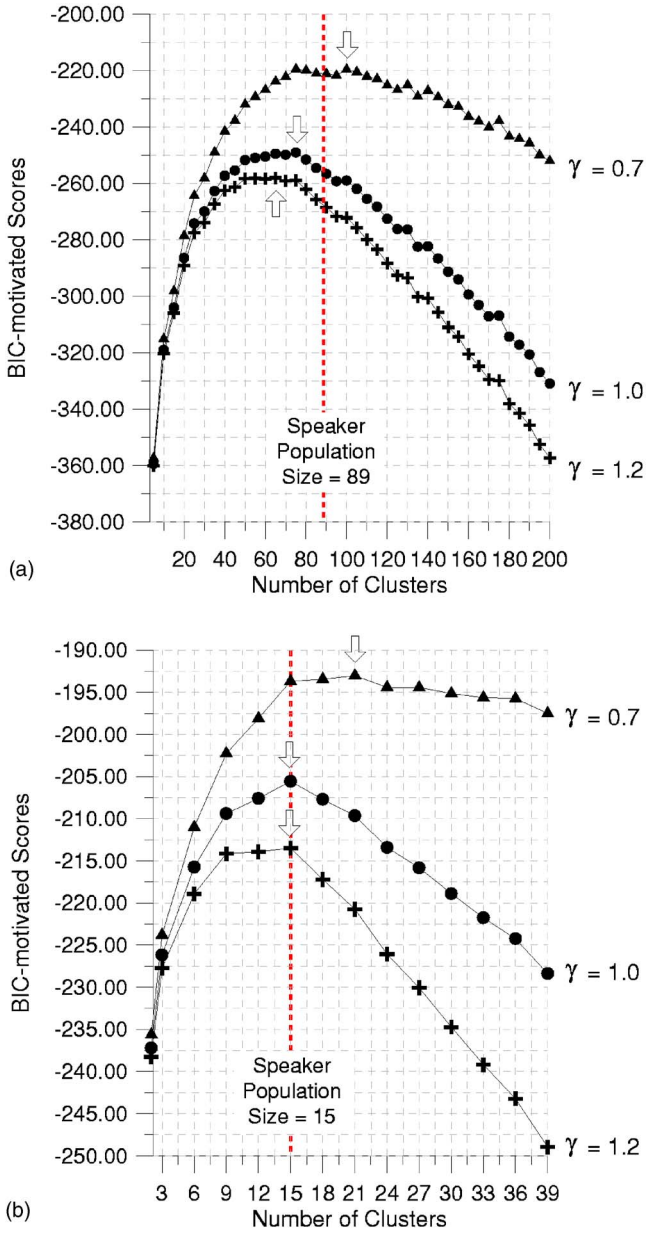


FIG. 9. (Color online) BIC-motivated scores as a function of number of clusters, in which the penalty factor γ was set to be equal to, slightly greater than, and slightly smaller than 1, respectively. The arrowed peak of each curve indicates the optimal number of clusters according to the criterion in Eq. (28). (a) Tests on subset "h4e-98-2". (b) Tests on subset "SRE-02".

diarization. On the other hand, recognizing the high computational complexity of the proposed speaker-clustering system, more work is still needed to study the way to improve the system efficiency.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Council, Taiwan, ROC, under Grant No. NSC93-2213-E-001-017. The authors are very grateful to the anonymous reviewers and the associate editor, Dr. Douglas O'Shaughnessy, for their careful reading of this article and their constructive suggestions.

[1]J. P. Campbell, "Speaker recognition: a tutorial," Proc. IEEE **85**(9), 1437–1462 (1997).
[2]D. O'Shaughnessy, "Speaker Recognition," IEEE ASSP Mag. **3**(4), 4–17 (1986).
[3]D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," Digit. Signal Process. **10**, 19–41 (2000).
[4]S. Kwon and S. Narayanan, "Unsupervised speaker indexing using generic models," IEEE Trans. Speech Audio Process. **13**(5), 72–83 (2005).
[5]I. Lapidot, H. Guterman, and A. Cohen, "Unsupervised speaker recognition based on competition between self-organizing maps," IEEE Trans. Neural Netw. **13**(4), 877–887 (2002).
[6]J. Makhoul, F. Kubala, T. Leek, D. Liu, L. Nguyen, R. Schwartz, and A. Srivastava, "Speech and language technologies for audio indexing and retrieval," Proc. IEEE **88**(8), 1338–1353 (2000).
[7]A. Solomonoff, A. Mielke, M. Schmidt, and H. Gish, "Clustering speakers by their voices," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (1998).
[8]B. Zhou and J. H. L. Hansen, "Unsupervised audio stream segmentation and clustering via the Bayesian information criterion," *Proc. International Conference on Spoken Language Processing (ICSLP)* (2000).
[9]S. E. Johnson, "Who spoke when?—Automatic segmentation and clustering for determining speaker turns," *Proc. European Conference on Speech Communication and Technology (Eurospeech)* (1999).
[10]D. A. Reynolds, E. Singer, B. A. Carson, G. C. O'Leary, J. J. McLaughlin, and M. A. Zissman, "Blind clustering of speech utterances based on speaker and language characteristics," *Proc. International Conference on Spoken Language Processing (ICSLP)* (1998).
[11]M. Ben, M. Betser, F. Bimbot, and G. Gravier, "Speaker diarization using bottom-up clustering based on a parameter-derived distance between adapted GMMs," *Proc. International Conference on Spoken Language Processing (ICSLP)* (2004).
[12]D. Moraru and S. Meignier, "The ELISA consortium approaches in broadcast news speaker segmentation during the NIST 2003 Rich Transcription Evaluation," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (2004).
[13]NIST, "Benchmark Tests: Rich Transcription," http://www.nist.gov/speech/tests/rt/rt2005/spring/index.htm
[14]S. E. Tranter, "Two-way Cluster Voting to Improve Speaker Diarisation Performance," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (2005).
[15]NIST, "Benchmark Tests: Speaker Recognition," http://www.nist.gov/speech/tests/spk/ 2001/index.htm
[16]S. Furui, "Unsupervised speaker adaptation method based on hierarchical spectral clustering," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (1989).
[17]H. Gish, M. H. Siu, and R. Rohlicek, "Segregation of speakers for speech recognition and speaker identification," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (1991).
[18]T. Kosaka and S. Sagayama, "Tree-structured speaker clustering for fast speaker adaptation," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (1994).
[19]S. S. Chen and P. S. Gopalakrishnan, "Clustering via the Bayesian information criterion with applications in speech recognition," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (1998).
[20]R. Faltlhauser and G. Ruske, "Robust speaker clustering in eigenspace," *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding* (2001).
[21]H. Jin, F. Kubala, and R. Schwartz, "Automatic speaker clustering," *Proc. DARPA Speech Recognition Workshop* (1997).
[22]D. Liu and F. Kubala, "Online speaker clustering," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (2003).
[23]Z. Liu, "An efficient algorithm for clustering short spoken utterances," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (2005).
[24]Y. Moh, P. Nguyen, and J. C. Junqua, "Towards domain independent speaker clustering," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (2003).
[25]M. A. Siegler, U. Jain, B. Raj, and R. M. Stern, "Automatic segmentation, classification and clustering of broadcast news audio," *Proc. DARPA Speech Recognition Workshop* (1997).
[26]S. Furui, "Cepstral analysis technique for automatic speaker verification," IEEE Trans. Speech Audio Process. **29**(2), 254–272 (1996).
[27]H. Hermansky, N. Morgan, A. Bayya, and P. Kohn, "Compensation for the

effects of the communication channel in auditory-like analysis of speech," *Proc. European Conference on Speech Communication and Technology (Eurospeech)* (1991).
[28]R. Sinha, S. E. Tranter, M. J. F. Gales, and P. C. Woodland, "The Cambridge University March 2005 Speaker Diarisation System," *Proc. European Conference on Speech Communication and Technology (Eurospeech)* (2005).
[29]X. Zhu, C. Barras, S. Meignier, and J. L. Gauvain, "Combining speaker identification and BIC for speaker diarization," *Proc. European Conference on Speech Communication and Technology (Eurospeech)* (2005).
[30]S. E. Johnson and P. C. Woodland, "Speaker clustering using direct maximisation of the MLLR-adapted likelihood," *Proc. International Conference on Spoken Language Processing (ICSLP)* (1998).
[31]J. Pelecanos and S. Sridharan, "Feature warping for robust speaker verification," *Proc. ISCA Odyssey Workshop* (2001).
[32]D. E. Goldberg, *Genetic Algorithm in Search, Optimization and Machine Learning* (Addison-Wesley, New York, 1989).
[33]L. Kaufman and P. J. Rousseuw, *Finding Groups in Data: An Introduction to Cluster Analysis* (Wiley, New York, 1990).
[34]J. E. Baker, "Adaptive selection methods for genetic algorithm," *Proc. International Conference on Genetic Algorithms and Their Applications* (1985).
[35]A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. R. Stati. Soc. **39**(1), 1–38 (1977).
[36]S. Kullback, *Information Theory and Statistics* (Dover, New York, 1968).
[37]C. S. Huang, H. C. Wang, and C. H. Lee, "A study on model-based error rate estimation for automatic speech recognition," IEEE Trans. Speech Audio Process. **11**(6), 581–589 (2003).
[38]F. Bimbot, I. Magrin-Chagnolleau, and L. Mathan, "Second-order statistical measures for textindependent speaker identification," Speech Commun. **17**, 177–192 (1995).
[39]H. Gish, "Robust discrimination in automatic speaker identification," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (1990).
[40]G. Schwarz, "Estimating the Dimension of a Model," Ann. Stat. **6**, 461–464 (1978).
[41]LDC, http://www.ldc.upenn.edu/.
[42]The VIMAS speech codec, http://www.vimas.com
[43]J. Ajmera, H. Bourlard, I. Lapidot, and I. McCowan, "Unknown-multiple speaker clustering using HMM," *Proc. International Conference on Spoken Language Processing (ICSLP)* (2002).
[44]I. Lapidot, "SOM as likelihood estimator for speaker clustering," *Proc. European Conference on Speech Communication and Technology (Eurospeech)* (2003).
[45]L. Hubert and P. Arabie, "Comparing Partitions," J. Classif. **2**, 193–218 (1985).
[46]W. M. Rand, "Objective criteria for the evaluation of clustering methods," J. Am. Stat. Assoc. **66**, 846–850 (1971).
[47]J. McLaughlin, D. Reynolds, E. Singer, and G. C. O'Leary, "Automatic speaker clustering from multi-speaker utterances," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (1999).

Acoustic microscope lens modeling and its application in determining biological cell properties from single- and multi-layered cell models

Tribikram Kundu[a] and Joon-Pyo Lee
Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, Arizona 85721

Christopher Blase and Jürgen Bereiter-Hahn
Kinematic Cell Research Group, Biocentre, J. W. Goethe University, Marie Curie Str. 9, D 60439 Frankfurt am Main, Germany

(Received 21 February 2006; revised 15 June 2006; accepted 16 June 2006)

The acoustic microscopy technique provides some extraordinary advantages for determining mechanical properties of living cells. It is relatively fast, of excellent spatial resolution, and of minimal invasiveness. Sound velocity is a measure of the cell stiffness. Attenuation of cytoplasm is a measure of supramolecular interactions. These parameters are of crucial interest for studying cell motility and volume regulations and to establish the functional role of the various elements of the cytoskeleton. Using a scanning acoustic microscope, longitudinal wave speed, attenuation and thickness profile of a biological cell were measured earlier by Kundu *et al.* [Biophys. J. **78**, 2270–2279 (2000)]. In that study it was assumed that the cell properties did not change through the cell thickness but could vary in the lateral direction. In that effort the acoustic-microscope-generated signal was modeled as a plane wave striking the cell at normal incidence. Such assumptions ignored the effect of cell inhomogenity and the surface skimming Rayleigh waves. In this paper a rigorous lens model, based on the DPSM (distributed point source method), is adopted. For the first time in the literature the cell is modeled here as a multi-layered material and the effect of some external drug stimuli on a living cell is studied. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2221556]

PACS number(s): 43.80.Cs, 43.80.Ev, 43.80.Jz [CCC] Pages: 1646–1654

I. INTRODUCTION

The use of acoustic waves in microscopy technology provides many benefits compared to its conventional optical microscope counterpart. One such benefit is its ability to inspect a specimen in the dark, which is often needed for biological specimens. Another is its capability to see inside an optically opaque specimen; this property made this technology popular with electronic packaging industries for inspecting the integrated circuit (IC) chips. Today, very high frequency, 1 Giga Hertz (10^9 Hz) and sometimes even higher frequency, ultrasound is being used for biological cell characterization while electronic circuit chips are being inspected with the ultrasound in the frequency range 25 to 400 MHz.

In acoustic microscopy experiments the elastic wave speed and wave attenuation in the specimen are measured by $V(f)$ or $V(z)$ techniques. $V(f)$ and $V(z)$ plots are the ultrasonic signal voltage (V) variation plots as a function of the signal frequency (f) and defocus distance (z), respectively. A specimen's density, Poisson's ratio, and Young's modulus are directly related to its acoustic wave speed. In order to accurately estimate the wave speed and attenuation in the specimen, the acoustic field in absence of the specimen must be correctly modeled first. A semi-analytical method called the distributed point source method (DPSM) is used in this paper to model a 1-GHz acoustic microscope lens. Then the model-predicted pressure field is used to calculate the thickness profile and properties of biological cell specimens from experimental data.

Rigorous transducer modeling at 1 GHz has rarely been attempted earlier because it requires an immense amount of computer time and memory when conventional numerical techniques based on finite element or boundary element methods are followed. In all previous analyses the high frequency acoustic microscope lense has been modeled using geometric theories like the ray theory,[1] similar to optics. Such simplifying assumptions in the microscope lens modeling may result in inaccurate predictions in the cell properties.

Other methods have also been used for the determination of mechanical properties of living cells. For a review of older methods one is referred to Ref. 2. These methods include local aspiration of cytoplasm with a pipette,[3] local poking of cytoplasm,[4] magnetometry,[5] and atomic force microscopy.[6,7] In all these methods the forces needed to evoke a certain deformation are measured and thus allow the determination of elastic and viscous properties from stress-strain relations and the dependence of these relations on the frequency of oscillating stress applications. Among the methods for the determination of mechanical properties of living cells, acoustic microscopy provides some extraordinary advantages: It is relatively fast, of excellent spatial resolution,

[a]Author to whom correspondence should be addressed. Electronic mail: tkundu@email.arizona.edu

and of minimal invasiveness. Relative changes of cytoplasmic forces may be resolved with a time resolution of about 5 s,[8,9] exact determination of mechanical properties requires a series of five to six images requiring an acquisition time between 25 and 30 s at a resolution of 512×256 pixels. Spatial resolution is in the range of 3 μm^2. Viscoelastic properties, however, may not change over such very small areas. The minimal invasiveness results from the very high frequency ultrasound (GHz range) that does not cause any damage or disturbance of the cells.[10] The only effect that has to be considered is the oscillating shear force caused by the scanning movement of the acoustic lens that is coupled with the specimen via the culture medium. The main parameters obtained by an acoustic microscope are sound velocity and sound attenuation as well as dimensions of the cell. Sound velocity is related to the Young's modulus and Poisson's ratio. Attenuation of cytoplasm is a measure of the supramolecular interactions.

Previous approaches on the computation of tissue properties have been based on shear wave propagation[11] or on measurements of the reflected sound as a function of the defocus position or $V(z)$ curves.[12] For the computation of cell properties from acoustic-microscope-generated signals an estimation of the cell thickness profile [13–18] is needed. The cell profile is traditionally computed by counting the interference rings and estimating the value of the longitudinal wave speed in the cell. Kundu *et al.*[17,18] used this information to get a rough estimate of the probable upper and lower bounds of the cell thickness at different pixels or cell positions. A simplex inversion algorithm is then followed to predict the cell thickness and other cell parameters. They used the $V(z)$ curve or voltage (V) versus defocus distance (z) curve, generated by an acoustic microscope operating at 1-GHz frequency to compute the cell properties. A new transducer geometry was proposed by Chubachi and co-workers[19] to be combined with a defocusing series in the range of about one wavelength in water to obtain $V(z)$ curves for each point in a biological sample. Later Kundu *et al.*[20] obtained cell properties from the $V(f)$ curves. The main advantages of the $V(f)$ curve technique in comparison to the $V(z)$ curve technique is that it is less time consuming and more accurate because monitoring the signal frequency "f" is easier than changing the defocus distance z.

The $V(f)$ technique has been developed by Kundu *et al.*[20] for the investigation of cellular behavior on stretchable materials using acoustic microscopes.[21] The method to investigate cells on silicon rubber was introduced by Harris.[22,23] In the present paper two improvements are introduced over Kundu, *et al.*'s earlier work[20]—first, in this study the cell can be modeled as an inhomogeneous material in both lateral and vertical (through-thickness) directions while in the earlier study only the inhomogenity in the lateral direction was considered. Second, in the current study the acoustic-microscope-generated ultrasonic beam is modeled rigorously using the DPSM semi-analytical modeling technique while in the previous study it was modeled as a plane wave striking the substrate at normal incidence. Assumption of normal incidence ignores the contribution of any surface skimming Rayleigh waves that might be generated at the substrate from the critically incident beams. When the substrate is placed at the focal plane then the incident beam has approximately a plane wave front, and strong surface waves are not generated at the substrate. However, if the substrate is defocused (i.e., z in Figure 2 has a nonzero value), then the incident beam strikes the substrate at different angles varying from 0° (normal incidence) to a nonzero θ; in our case θ is equal to 50°, the half angle of the lens. Therefore, the old technique might be used only for the focused position of the substrate while the current technique is applicable for any substrate position. Thus, the analysis presented in this paper is more general and rigorous.

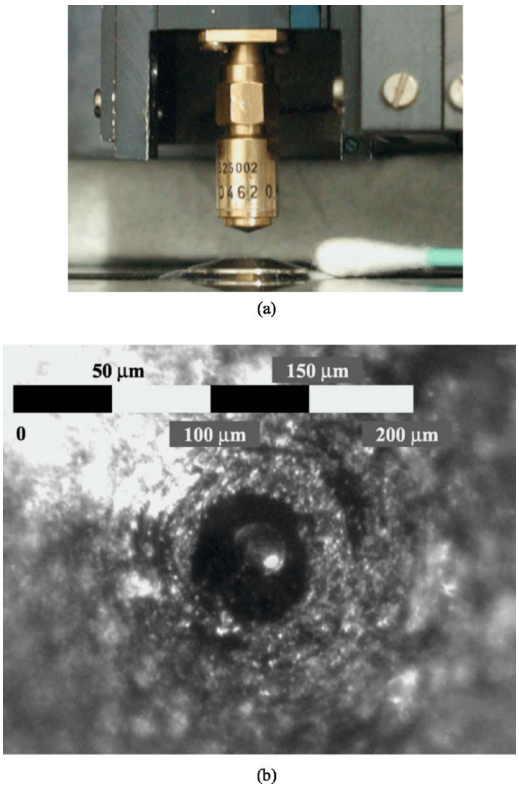


FIG. 1. (Color online) (a) Side view of the acoustic microscope lens. (b) Magnified acoustic image of the 1-GHz lens tip.

II. ACOUSTIC MICROSCOPE LENS MODELING

The side view of the acoustic microscope lens is shown in the photograph presented in Fig. 1(a) and the acoustic image of the lens tip is shown in Fig. 1(b). The dark circular region (of diameter >50 μm) at the center of the acoustic image of Fig. 1(b) corresponds to the concave lens.

Figure 2 shows the position of the biological cell on substrate relative to the lens rod. A transducer is mounted on the top of the lens rod. The point focused lens has a radius of curvature r (=40 μm for our microscope). Its center of curvature is located at point C and the focal point at F. A coupling fluid—typically a saline solution—is present between the lens and the substrate. The cell is placed on the substrate, which is slightly defocused. The defocus distance z is measured from the focal point to the top reflecting surface. The cell generally has a shape like a fried egg with its nucleus in the central region and the lamellar part near the periphery.

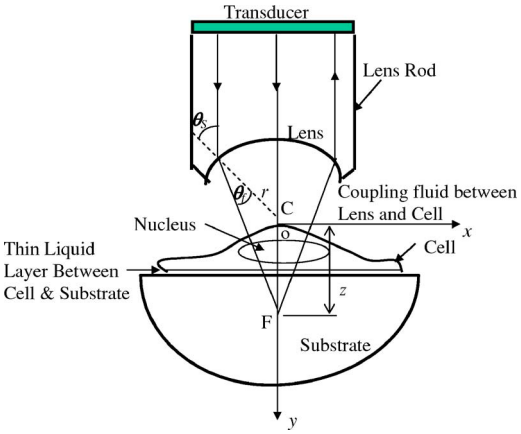


FIG. 2. (Color online) Schematic of the acoustic microscope inspecting a biological cell on a substrate: C is the center of curvature of the lens; F is its focal point. For details see text.

The cells are typically a few microns thick while the substrate thickness is in the order of millimeters. Since the substrate thickness is several orders of magnitude higher than the cell thickness or the acoustic wave length (also in the order of microns), the substrate can be modeled as a semi-infinite half-space. A very thin (10 nm) liquid layer is present between the cell and the substrate as shown in Fig. 2. To accurately measure the cell properties, first of all one needs to correctly model the acoustic field in front of the lens where the cell is placed.

The lens geometry with all its dimensions is shown in Fig. 3. Note that the radius of the curvature of the lens is 40 μm, its half angle is 50°, and the diameter of the lens opening is 56.57 μm. The ultrasonic field generated by this lens will be modeled by the DPSM technique.

The DPSM is a semi-analytical technique that can be used to calculate the ultrasonic fields produced by transducers of finite dimension placed in a homogeneous or nonhomogeneous media. The DPSM technique for ultrasonic field modeling was first developed by Placko and Kundu.[24] They successfully used this technique to model ultrasonic fields in a homogeneous fluid, and in a nonhomogeneous fluid with one interface[25,26] and multiple interfaces.[27] Ultrasonic fields near a fluid-solid interface have been modeled for the first time by the DPSM technique by Banerjee and Kundu.[28,29] In DPSM modeling one layer of point sources is distributed

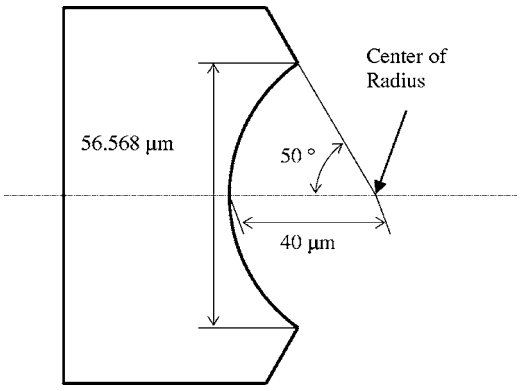


FIG. 3. Acoustic lens geometry.

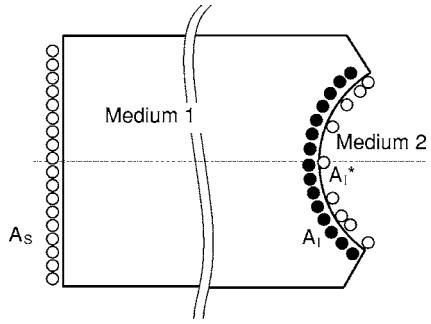


FIG. 4. Point source distribution relative to the lens geometry for the DPSM modeling of the acoustic lens.

near the ultrasonic signal generators or transducers and two layers are distributed near the fluid-solid interface as shown in Fig. 4. On the flat end of the lens rod the ultrasonic transducer (signal generator) is mounted. A layer of point sources denoted by A_S is distributed on this flat surface. The ultrasonic signal that is generated by the transducer propagates through the lens rod and strikes the concave surface of the lens. Part of the signal energy is reflected back into the solid lens material (medium 1) and part is transmitted into the coupling fluid (medium 2) used between the lens and the specimen. Two layers of point sources, A_I and A_I^*, are distributed along the interface. One layer (A_I) models the transmitted field in medium 2 and the other layer (A_I^*) models the reflected field in medium 1. Note that the point sources may be distributed uniformly as shown in A_S and A_I layers or can have nonuniform distribution as shown in the A_I^* layer. The individual source strengths are determined by satisfying the continuity conditions across the interface.

To compute the pressure (or stress) and velocity (or displacement) at a point in medium 2, the contributions of all point sources distributed along A_I will have to be superimposed while for computing the ultrasonic field in medium 1 we will have to superimpose the contributions of all point sources denoted as A_S and A_I^*. Note that the solid circles of Fig. 4 contribute to the ultrasonic fields in medium 2 and open circles contribute to the ultrasonic fields in medium 1. If we denote the pressure field by P_i and the velocity field by V_i in the ith medium ($i=1$ or 2), then we can write [30]

$$
\begin{aligned}
P_1 &= Q_{1S}A_S + Q_{1I}^*A_I^*, \\
V_1 &= M_{1S}A_S + M_{1I}^*A_I^*, \\
P_2 &= Q_{2I}A_I, \\
V_2 &= M_{2I}A_I,
\end{aligned}
\tag{1}
$$

where Q and M are matrices that relate source strength vectors, A_I, A_S, etc. to the pressure and velocity values at any point of interest.

If two points are considered at the common boundary of media 1 and 2, then these two points should have same pressure (or stress) and displacement (or velocity) in the direction normal to the interface because of the displacement and stress continuity conditions across the interface. Thus,

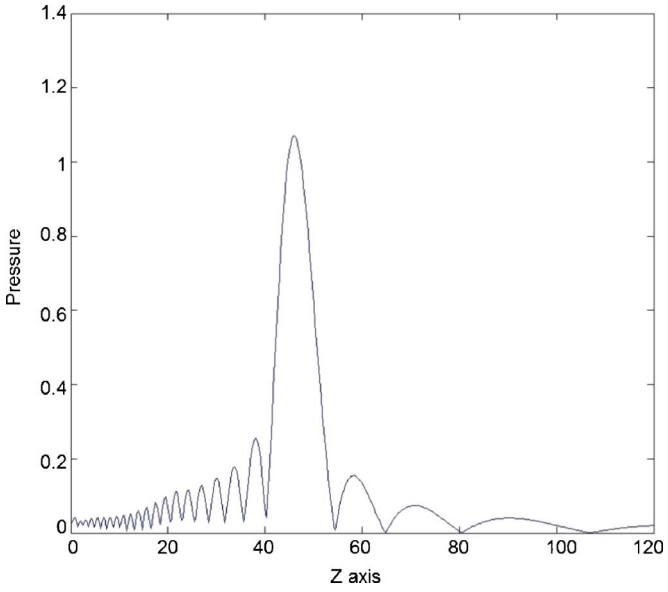


FIG. 5. Acoustic pressure variation along the central axis (axis of symmetry) of the lens rod computed from DPSM. Z is plotted from 0 to 120 μm. In this plot $Z=0$ corresponds to the point of intersection of the central axis and the concave lens.

$$Q_{2I}A_I = Q_{1S}A_S + Q_{1I}^*A_I^*,$$
$$M_{2I}A_I = M_{1S}A_S + M_{1I}^*A_I^* \tag{2}$$

or

$$Q_{2I}A_I - Q_{1I}^*A_I^* = Q_{1S}A_S,$$
$$M_{2I}A_I - M_{1I}^*A_I^* = M_{1S}A_S. \tag{3}$$

Note that the right-hand side of Eq. (3) is known from the strength of the transducer mounted on the flat side of the lens rod. The interface source strengths are obtained from Eq. (3) in the following manner,

$$\begin{Bmatrix} A_I \\ A_I^* \end{Bmatrix} = \begin{bmatrix} Q_{2I} & Q_{1I}^* \\ M_{2I} & M_{1I}^* \end{bmatrix}^{-1} \begin{Bmatrix} Q_{1S}A_S \\ M_{1S}A_S \end{Bmatrix}. \tag{4}$$

After obtaining the source strength vector A_I the pressure and velocity at any point of medium 2 can be calculated from the last two equations of equation set (1).

The pressure field computed in this manner along the central axis (axis of symmetry) of the lens rod is shown in Fig. 5. Note that it is plotted in the coupling fluid medium (medium 2) in front of the lens along the dotted axis of symmetry of Figs. 3 and 4. The value of Z is 0 on the lens surface. It should be noted here that the definition of Z in Fig. 5 is different from the defocus distance z in Fig. 2. The peak value of the pressure appears at a distance close to 46 μm from the center of the lens. It should also be mentioned here that in many simple geometric models for point focus transducers the acoustic pressure becomes infinity at the focal point. Such unrealistic values are avoided here by the DPSM modeling.

The pressure variation on the plane perpendicular to the central axis and going through the focal point is shown in Fig. 6. Note that although the detail pressure variations

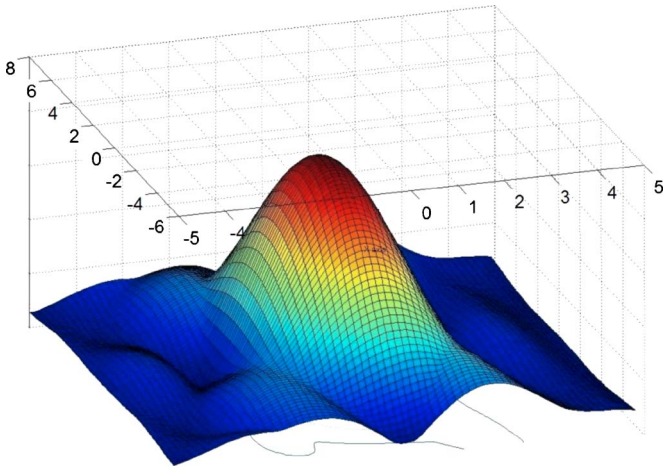


FIG. 6. (Color online) Acoustic pressure variation on the focal plane (perpendicular to the axiz of symmetry) obtained by DPSM.

shown in Figs. 5 and 6 have several peaks and dips, the pressure field close to the focal plane and near the central axis is not very complex and can be modeled by simple curve fitting. The expression of this curve does not change as long as the acoustic microscope lens is not changed. The function that describes the ultrasonic field near the focal point depends on the lens geometry as the pupil function does.

For easier curve fitting a new x-Z coordinate system is introduced with its origin at the focal point. In the new coordinate system Z is measured from the focal point along the central axis of the lens with its positive direction away from the lens. Coordinate x is measured in the direction perpendicular to the central axis of the lens. After several attempts the following expression is found to best fit the DPSM generated results:

$$p(x,Z) = (1.7\times10^{-4}Z^4 - 8.4\times10^{-4}Z^3 - 0.031Z^2 + 4.6 \times 10^{-5}Z + 0.99) \times \exp(-0.296x^2) + 0.01. \tag{5}$$

Equation (5) is plotted in Fig. 7 for different values of x, and when Z is measured from the focal plane. Continuous lines are the plots of Eq. (5) for different values of x and the circles are obtained from the DPSM computation for $x=0$. Note that the matching between the DPSM results and the highest continuous line (that corresponds to $x=0$) is excellent in the region $-6\mu\text{m}<Z<7$ μm. Therefore, as long as the sample is placed within this 13 μm range the pupil function given in Eq. (5) is acceptable. Since cells are only a few microns (1 to 4 μm) thick and the substrate surface is defocused between 0 and 2 μm, the entire cell lies in this 13 μm zone. It should also be noted here that Z in Fig. 7 is measured from the focal point while that in Fig. 5 is measured from the center point of the concave lens. Thus, there is a 46 μm shift in the Z-axis origin between these two figures.

III. CELL MODELING

In Fig. 2 one can see that the lamellar part of the cell is different from the nucleus part both in their their constituents and geometry. The lamellar region can be modeled as a ho-

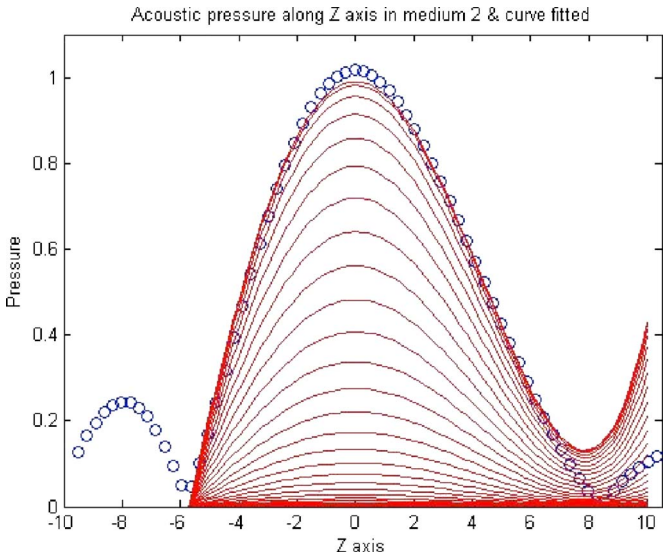


FIG. 7. (Color online) Pressure variations near the focal plane as a function of z for differnt values of x. The focal plane is located at $Z=0$. Various continuous lines are plots of Eq. (5) for different values of x. Outermost continuous line is for $x=0$ that corresponds to the pressure variation along the axis of symmetry of the lens and the open circles are the pressure variations along the same axis obtained from the DPSM analysis. Note that the matching between the DPSM results and the values obtained from Eq. (5) is very good in the region $-6\mu\text{m}<Z<7\ \mu\text{m}$.

mogeneous material. The lamellar region on the substrate with a thin liqud layer between the cell and the substrate (see Fig. 2) should be modeled as a three-layered half-space—the top layer is the lamellar part of the cell, the second layer is the thin liquid layer between the cell and the substrate, and the bottom-most material is the substrate. On the other hand, the central part of the cell itself should be modeled as a three-layered material—the middle layer is the nucleus and the top and bottom layers are cytoplasm. Thus, for analyzing the central region of the cell the reflecting specimen should be modeled as a five-layered material—three layers coming from the cell, the fourth layer is the thin liquid layer between the cell and the subtrate, and the fifth medium is the substrate half space.

Kundu [31] has computed the reflected field generated by a focused ultrasonic beam striking a multi-layered specimen. Part of this theory that is directly relevant to the current problem geometry is briefly reviewed here.

The reflected field at a general point (x,y) in the coupling fluid (see Fig. 2) is given by

$$\Phi_R(x,y)=\left(\frac{1}{2\pi}\right)\int_{-\infty}^{\infty}R(k)\phi_i(k,0)\exp(-ikx-i\delta_f y)dk, \tag{6}$$

where $R(k)$ is the reflection coefficient of the layered specimen that is being inspected by the acoustic microscope, x is the horizontal distance from the central axis, and y is the vertical distance from the reflecting surface. $\phi_i(k,0)$, is the Fourier transform of the incident field $\phi_i(x,y)$ at $y=0$. Figure 2 shows the origin and directions of coordinates x and y. Note that $y=0$ corresponds to the reflecting surface at the top of the specimen. Therefore, as the cell thickness varies so does the position of the origin and the defocus length z. The Fourier transform is taken in the wave number domain or k domain to replace the variable x by the wave number k. The reflected field can be transformed back into the xy space by carrying out the inverse Fourier transform, as done in Eq. (6). Note that the exponential term in the integrand represents the propagating wave from the specimen surface at $y=0$ to the point of interest (x,y).

The incident field on the specimen surface can be expressed in the following form,

$$\phi_i(x,0)=\exp\left\{ik_s r(1-\cos\theta_S)+ik_f\left[\frac{r}{(\cos\theta_f-\mu\cos\theta_S)}-\sqrt{x^2+z^2}\right]\right\}p(x,Z), \tag{7}$$

where r is the radius of curvature of the acoustic microscope lens; for our microscope it is 40 μm. The incident angle θ_S and the transmission angle θ_f at the lens-fluid interface are shown Fig. 2. k_s is the longitudinal wave number (that is defined as the angular frequency of the acoustic signal divided by the longitudinal wave speed) of the lens material, k_f is the longitudinal wave number of the coupling fluid. z is the defocus distance (see Fig. 2) and $\mu\alpha_f/\alpha_s$, where α_f and α_s are longitudinal wave speeds in the coupling fluid and the lens material, respectively. The function $p(x,Z)$ has been defined in Eq. (5).

The incident field in k domain is given by

$$\phi_i(k,0)=\exp(ik_s r)\int_0^{\infty}p(x,Z)\exp\{-ik_s r\cos\theta_S+ik_f[rf_1(\theta_S)-\sqrt{x^2+z^2}]+ikx\}dx, \tag{8}$$

where

$$f_1(\theta_S)=\frac{1}{\cos\theta_f-\mu\cos\theta_S}. \tag{9}$$

δ_f of Eq. (6) is given by

$$\delta_f=\sqrt{(k_f^2-k^2)}\left\{1+\frac{ik_f^2}{2q_f\alpha_f(k_f^2-k^2)}\right\}, \tag{10}$$

where k and k_f are the same as those defined in Eqs. (6) and (7), respectively. q_f is the quality factor of the coupling fluid. Note that the quality factor is related (inversely proportional) to the attenuation [32].

Reflection coefficient $R(k)$ of Eq. (6) has the following form,

$$R(k)=\frac{i\delta_f J_{12}^{12}-\rho_f\omega^2 J_{13}^{12}}{i\delta_f J_{12}^{12}+\rho_f\omega^2 J_{13}^{12}}, \tag{11}$$

where δ_f is defined in Eq. (10), and ρ_f and ω are coupling fluid density and angular frequency of the acoustic signal, respectively. J_{kl}^{ij} has the following form for the n-layered specimen:

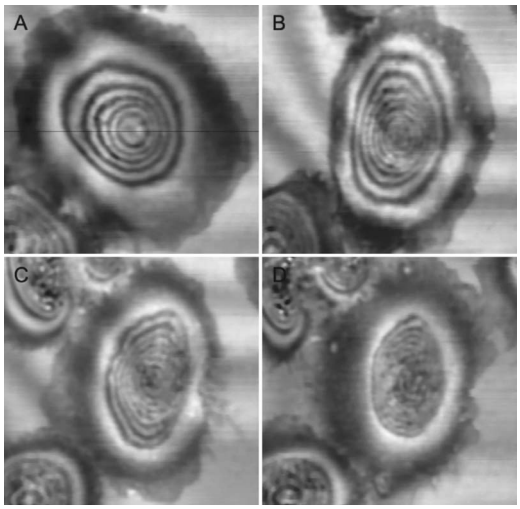


FIG. 8. Acoustic image of a HaCaT cell: four images correspond to four different times T1 (=−15 min., A), T2 (=0 min., B), T3 (=30 min., C) and T4 (=45 min., D). T1 and T2 are before the drug application while T3 and T4 are 30 and 45 minutes after the drug application. Drug is applied right after taking image B at time T2=0, T1=−15 indicates the image is taken 15 minutes before the drug application and positive signs of T3 and T4 indicate images taken after the drug application. The scan line along which the cell profiles and its properties are calculated is shown in A.

$$J_{kl}^{ij} = J_{ik}J_{jl} - J_{il}J_{jk},$$
$$J = E_n^{-1}A_{n-1}A_{n-2}\cdots A_2A_1. \tag{12}$$

A_m is the 4×4 layer matrix of the mth layer; E_n is also a 4×4 matrix derived from the properties of the substrate material. These layer matrices and transmission coefficients are given in Ref. 30 and are omitted here.

$\phi_R(x,y)$ of Eq. (6) gives the ultrasonic field in the coupling fluid after it is reflected by the specimen. Thus, the reflected ultrasonic field at the lens-coupling fluid interface can be computed. Finally, the field reaching the receiver mounted on the top of the lens rod can be obtained by multiplying the transmitted field with the propagation term in the lens rod. Thus, the potential field reaching the receiver is obtained as

$$\phi_T = \phi_R^c T_{fs}(\theta_f)\exp\{ik_s r(1-\cos\theta_S)\}. \tag{13}$$

Superscript c of ϕ_R indicates that only the critically reflected part of ϕ_R is considered in the final field computation at the receiver location. To obtain the voltage produced by the transducer, it is necessary to integrate ϕ_T over the receiving transducer. In this research it is computed by the Gauss-Legendre integration formula.

IV. EXPERIMENTAL RESULTS

The technique described above is used to study living HaCaT cells on plastic substrates over a time period of 60 min. HaCaT cells are studied to investigate the effect of Cytochalasin D drug (1 μg/μL) on these cells. Experimental results are presented for two different cells. The first cell, whose image is shown in Fig. 8, is scanned at four different times by the acoustic microscope. The first scan [(A) in Fig. 8] is carried out at time T1=−15 min. The negative sign indicates that the image is taken 15 min before the drug application. Scan (B) is completed at T2=0, just before the drug is applied. Then two more images (C and D) are generated 30 and 45 minutes after the drug (Cytochalasin D, 1 μg/μL) application. Table I shows properties of the coupling fluid and the substrate medium used for the cell property determination. Several interference fringes generated due to the cell thickness variation can be seen in all four images of Fig. 8. Variations of the cell thickness and its other properties along the black scan line [see Fig. 8(A)] are computed and presented in Fig. 9 by inverting the acoustic-microscope-generated $V(f)$ curves.

Before inverting the $V(f)$ curves along a scan line using a simplex algorithm [20,17,33] to obtain the cell properties, the absolute upper and lower bounds of the P-wave speed, attenuation, and density of the cell material are estimated. From the earlier studies by the authors [20,16,17] and other investigators the range in which the cell property values must lie can be obtained. These values are necessary for the inversion process and are shown in Table II. In Table II bounds for only three unknown quantities are given; bounds for the fourth unknown, the cell thickness (h), are obtained from the phase values; bounds on the P-wave speeds are calculated using the following formula:[20]

$$h = \frac{n}{512f(1/\alpha_f - 1/\alpha_c)}, \tag{14}$$

where n is the gray scale value of the phase. f, α_f, and α_c are signal frequency, P-wave speed in the coupling fluid, and P-wave speed in the cell, respectively. Since the phase value n varies from one pixel to the next, the bounds on h also vary

TABLE I. Properties of the coupling fluid (saline solution), the lens material (sapphire), and the substrate material (plastic).

Material	P-wave speed (km/s)	S-wave speed (km/s)	Density (gm/cc)	Attenuation (normalized with respect to that of water)
Coupling fluid (saline solution)	1.5	0	1.0	1.0
Lens material (sapphire)	11.1	6.25	4.0	0
Substrate material (plastic)	2.37	1.15	1.04	0.06

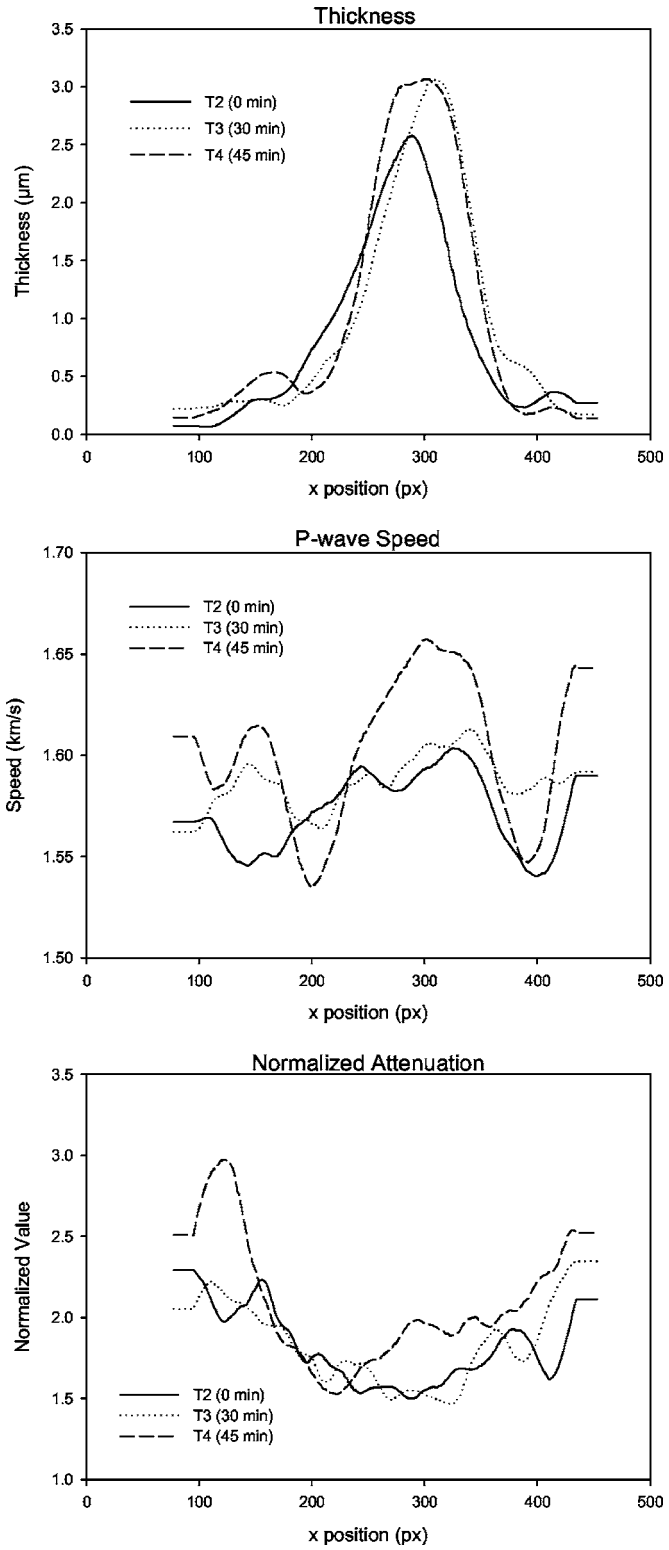


FIG. 9. Cell profile (top figure), P-wave speed (middle figure), and attenuation (bottom figure) variations in the HaCaT cell of Fig. 8 before (T2=0 min) and after (T3=30 min and T4=45 min) the treatment with cytochalasin D. Note an increase in the P-wave speed and swelling of the cell in the central region after the drug application.

from pixel to pixel, although the defined bounds of the other three parameters do not change over the cell.

Absolute bounds are the bounds of the parameters that must never be violated. A second set of bounds called probable bounds are then defined. These bounds must lie within the absolute bounds. Probable bounds are thus narrower than the absolute bounds. True values most probably lie within the probable bounds, but it is not guaranteed. These bounds are naturally not as high or as low as the absolute bounds. Simplex geometries or initial estimates of the cell parameters are obtained from the probable bounds. Iterations are carried out to minimize the least squares error between experimental and theoretical $V(f)$ values at every pixel point.

TABLE II. Absolute bounds of the unknown parameters.

Unknown parameters	Upper bound	Lower bound
P-wave speed (km/s)	1.8	1.5
Attenuation (normalized with respect to water)	5.8	1.0
Density (μm)	1.1	1.02

The computer program that is developed here can simultaneously determine four unknown parameters. The longitudinal wave speed, attenuation, and cell thickness are obtained from the program. In addition to these three main unknowns the cell density is also treated as an additional unknown parameter. A thin (0.01 μm) liquid layer between the cell and the substrate is considered in the model as shown in Fig. 2. Using the bounds of Table II, the cells shown in Fig. 8 are analyzed at three different times T2 (=0 min), T3 (=30 min), and T4 (=45 min).

From Fig. 9 one can see that after the drug application the nuclear region significantly swelled. The P-wave speed is also affected by the drug but its effect on the attenuation is not that significant.

One can see in Fig. 2 that the cell is not a single-layered material in the nucleus region. Now it will be investigated how the predicted cell properties change when the cell is modeled as a multi-layered material. In the multi-layered cell analysis the central part of the cell is modeled as a three-layered material. The central layer is the nucleus and is assumed to occupy 80% of the total cell thickness. Top and bottom layers occupy the remaining 20% of the total thickness and have identical properties and thickness. P-wave speeds and attenuations in the nucleus (middle layer) and in the surrounding cytoplasm (top and bottom layers) are treated as four unknown quantities. Cell thickness and density obtained from the single-layer analysis of the cell is assumed to remain unchanged when the nucleus is modeled as a three-layered material. Since it is difficult to consider more than four unknowns at a time, the density and the individual layer thickness are treated as known quantities for the multi-layered cell modeling.

The central region (pixels 225 to 300) of the second HaCaT cell (cell image not shown) is modeled as a three-layered material. The cell profile in this region is shown in the top plot of Fig. 10. Predicted P-wave speed and attenuation variations in the middle layer (nucleus) and in the top and bottom layers (surrounding cytoplasm) are shown in the middle and bottom plots of Fig. 10.

For comparison purposes the cell properties obtained from the single-layer cell model are also plotted on the same graph. Thus, three curves in each plot correspond to the

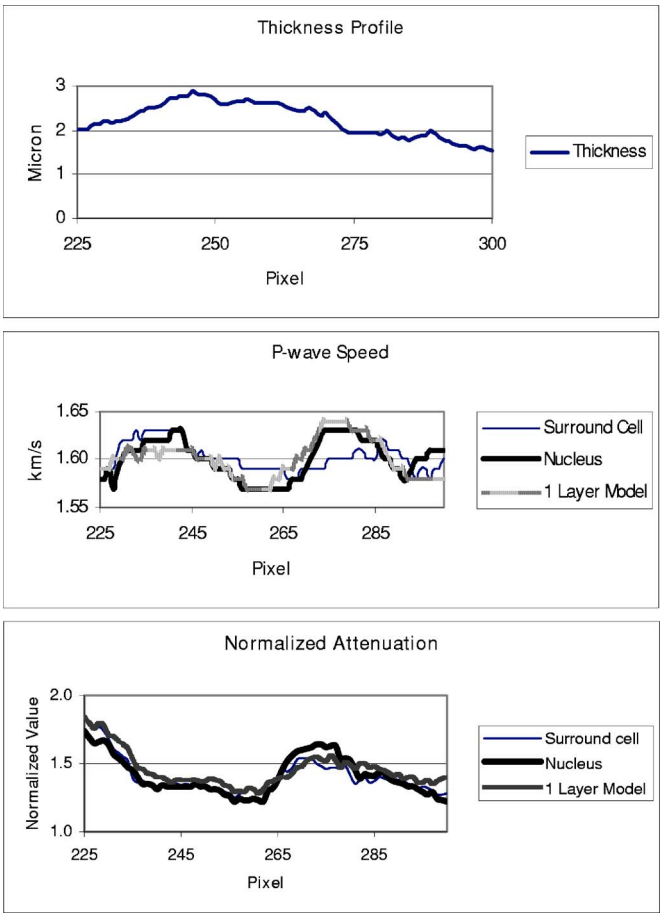


FIG. 10. (Color online) Cell profile (top figure), P-wave speed variation (middle figure), and attenuation variation (bottom figure) in the central region of a HaCaT cell—thinnest and thickest lines in bottom two plots correspond to the nucleus properties and surrounding cytoplasm properties, respectively, obtained from the three-layer cell model, while the third line shows the properties predicted by the single-layer cell model.

nucleus property, cytoplasm property, and the single-layer cell model predicted values. Note that all three curves show similar variations. For the attenuation variation the curves are very close to one another. However, for the P-wave speed variation, the curves are comparatively more separated but they show similar trend. Note that in the P-wave speed variation plot the predicted properties from the single-layer cell model are closer to the nucleus properties. This is expected since the nucleus occupies 80% of the cell thickness.

From Fig. 10 it can be concluded that the trend of the P-wave speed and attenuation variations from single-layer and three-layered cell models are similar; therefore, the accuracy of the single-layer cell model is sufficient at the present time.

V. CONCLUDING REMARKS

It is shown in this paper that the cell properties and the cell thickness profiles can be obtained by an acoustic microscope. The thickness profile, longitudinal wave speed, and attenuation variations in a cell, predicted in this manner, are sensitive enough to detect the effect of a drug (Cytochalasin D, 1 $\mu g/\mu L$) on the HaCaT cell considered here. Based on our investigation with the HaCaT cell it is concluded that the applied drug results in a swelling of the nucleus region of the cell, affects its P-wave speed, but the attenuation is not significantly affected in the nucleus region by the drug.

The cell is modeled both as a single-layer material and a three-layered material (in the nucleus region). A comparison of the predicted results from these two models shows similar trends in the cell property variations from both models. Therefore, if one is interested in obtaining the nucleus properties or the average cell properties, then the single-layer cell model is good enough because the properties obtained from the single layer model are very close to that of the nucleus. However, if an accurate determination of the properties of thin cytoplasm surrounding the nucleus is necessary, then an analysis of the multi-layered model of the cell may be warranted. In this paper, for the first time the DPSM is used to model an acoustic-microscope-generated ultrasonic field in front of the microscope lens. The model-predicted results clearly show that the ultrasonic field is not focused at a point but over a finite region. This realistic ultrasonic field is used to compute the cell properties.

ACKNOWLEDGMENTS

The experimental component of this research was carried out during the first author's visit to Germany as the winner of Humboldt Research Prize (Senior Scientist Award). The financial contribution of the Alexander von Humboldt Foundation to this research is gratefully acknowledged.

[1]T. Kundu, "A Theoretical Analysis of Acoustic Microscopy with Converging Acoustic Beams," J. Appl. Phys. B: Photophys. Laser Chem. **46**, 325–331 (1988).
[2]Y. Hiramoto, "Evaluation of cytomechanical properties," in *Cytomechanics*, edited by J. Bereiter-Hahn, O. R. Anderson, and W.-E. Reif (Springer Verlag, Berlin, 1987), pp. 31–46.
[3]G. W. Schmid-Schönbin, "Leukocyte biophysics," an invited review, Cell Biophys. **17**, 107–135 (1990).
[4]M. Duszyk, B. Schwab III, G. I. Zahalak, H. Qian, and E. L. Elson, "Cell poking: Quantitative analysis of indentation of thick viscoelastic layers," Biophys. J. **55**, 683–690 (1989).
[5]F. Ziemann, J. Radler, and E. Sackmann, "Local measurements of viscoelastic moduli of entangled actin networks using an oscillating magnetic bead micro-rheometer," Biophys. J. **66**, 2210–2216 (1994).
[6]E. Hassan, W. F. Heinz, M. D. Antonik, N. P. DCosta, S. Nageswaran, C. A. Schoenenberger, and J. H. Hoh, "Relative microelastic mapping of living cells by atomic force microscopy," Biophys. J. **74**, 1564–1578 (1998).
[7]M. Radmacher, M. Fritz, C. Kacher, J. Cleveland, and P. Hansma, "Measuring the viscoelastic properties of human platelets with the atomic force microscope," Biophys. J. **70**, 556–567 (1994).
[8]J. Bereiter-Hahn, and H. Lüers, "The role of elasticity in the motile behaviour of cells," NATO ASI Ser H: Cell Biol. **84**, 181–230 (1994).
[9]J. Bereiter-Hahn and H. Lüers, "Subcellular tension fields and mechanical resistance of the lamella front related to the direction of locomotion," Cell Biochem. Biophys. **29**, 243–262 (1998).
[10]J. Bereiter-Hahn, "Probing biological cells and tissues with acoustic microscopy," Adva. Acoustic-Microsc. **1**, 79–115 (1995).
[11]A. Sarvazyan, A. Scovoroda, and D. Vucelic, "Utilisation of surface acoustic waves and shear acoustic properties for imaging and tissue characterization," Acoust. Imaging **19**, 463–468 (1992).
[12]H. Sasaki, M. Tanaka, Y. Saijo, H. Okawai, Y. Terasawa, S. Nitta, and K. Suzuki, "Ultrasonic characterisation of renal cell carcinoma tissue," Nephron **74**, 125–130 (1996).
[13]J. A. Hildebrand, D. Rugar, R. N. Johnston, and C. F. Quate, "Acoustic microscopy of living cells," Proc. Natl. Acad. Sci. U.S.A. **78**, 1656–1660 (1981).

[14]J. A. Hildebrand and D. Rugar, "Measurement of cellular elastic properties by acoustic microscopy," J. Microsc. **134**, 245–260 (1984).
[15]J. Litniewski and J. Bereiter-Hahn, "Measurements of cells in culture by scanning acoustic microscopy," J. Microsc. **158**, 95–107 (1990).
[16]J. Litniewski and J. Bereiter-Hahn, "Acoustic velocity determination in cytoplasm by V(z) shift," Acoust. Imaging **19**, 535–538 (1992).
[17]T. Kundu, J. Bereiter-Hahn, and K. Hillmann, "Measuring elastic properties of cells by evaluation of scanning acoustic microscopy V (Z) values using simplex algorithm," Biophys. J. **59**, 1194–1207 (1991).
[18]T. Kundu, J. Bereiter-Hahn, and K. Hillmann, "Calculating acoustical properties of cells: Influence of surface topography and liquid layer between cell and substrate," J. Acoust. Soc. Am. **91**, 3008–3017 (1992).
[19]N. Chubachi, K. Hiroshi, S. Toshio, and T. Wakahara, "Acoustic microscope for measuring acoustic properties by micro-defocusing method," Acoust. Imaging **19**, 685–689 (1992).
[20]T. Kundu, J. Bereiter-Hahn, and I. Karl, "Cell Property Determination from the Acoustic Microscope Generated Voltage Versus Frequency Curves," Biophys. J., **78**, 2270–2279 (2000).
[21]I. Karl and J. Bereiter-Hahn, "Tension modulates cell surface motility," Cell Motil. Cytoskeleton **29**, 349–359 (1999).
[22]A. K. Harris, "Traction, and its relation to contraction in tissue cell locomotion," in *Cell Behaviour*, edited by R. Bellairs, A. Curtis, and G. Dunn (Cambridge U.P., Cambridge, 1982), 109–134.
[23]A. K. Harris, P. Wild, and D. Stopak, "Silicon rubber-substrata: a new wrinkle in the study of cell locomotion," Science **208**, 177–179 (1980).
[24]D. Placko and T. Kundu, "A Theoretical Study of Magnetic and Ultrasonic Sensors: Dependence of Magnetic Potential and Acoustic Pressure on the Sensor Geometry," in *Advanced NDE for Structural and Biological Health Monitoring*, edited by T. Kundu, SPIE 6th Annual International Symposium on NDE for Health Monitoring and Diagnostics, 4–8 March, Newport Beach, CA, Proc. SPIE **4335**, 52–62 (2001).
[25]D. Placko, T. Kundu, and R. Ahmad, "Theoretical Computation of Acoustic Pressure Generated by Ultrasonic Sensors in Presence of an Interface," Smart NDE and Health Monitoring of Structural and Biological Systems, SPIE's 7th Annual International Symposium on NDE and Health Monitoring and Diagnostics, San Diego, CA, Proc. SPIE **4702**, 157–168 (2002).
[26]J. P. Lee, D. Placko, N. Alnuamaini, and T. Kundu, "Distributed Point Source Method (DPSM) for Modeling Ultrasonic Fields in Homogeneous and Non-Homogeneous Fluid Media in Presence of an Interface," First European Workshop on Structural Health Monitoring, Ecole Normale Superieure de Cachan, France, Edited by D. L. Balageas (DEStech, Lancaster, PA, 2002), pp. 414–421.
[27]S. Banerjee, T. Kundu, and D. Placko, "Ultrasonic Field Modelling in Multilayered Fluid Structures Using DPSM Technique," J. Appl. Mech. (2006) (in press).
[28]S. Banerjee, N. A. Alnuaimi, and T. Kundu, "DPSM Technique for Ultrasonic Field Modelling Near Fluid-Solid Interface," Ultrasonics (2006) (in press).
[29]S. Banerjee and T. Kundu, "Elastic wave propagation in sinusoidally corrugated waveguides," J. Acoust. Soc. Am. **119**, No. 4, 2006–2017 (2006).
[30]D. Placko and T. Kundu, *DPSM—Distributed Point Source Method for Modeling Engineering Problems* (Wiley, New York, 2007).
[31]T. Kundu, "A complete acoustic microscopical analysis of multilayered specimens," J. Appl. Mech. **59**, 54–60 (1992).
[32]T. Kundu, *Ultrasonic Nondestructive Evaluation: Engineering and Biological Material Characterization* (CRC, Boca Raton, FL, 2004).
[33]T. Kundu, "Inversion of acoustic material signature of mulitlayered solids," J. Acoust. Soc. Am. **91**, 591–600 (1991).

Acoustic analysis of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*)[a)]

Ashley L. Pistorio, Brett Vintch, and Xiaoqin Wang[b)]
Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

(Received 11 January 2006; revised 9 June 2006; accepted 21 June 2006)

In contrast to humans and songbirds, there is limited evidence of vocal learning in nonhuman primates. While previous studies suggested that primate vocalizations exhibit developmental changes, detailed analyses of the extent and time course of such changes across a species' vocal repertoire remain limited. In a highly vocal primate, the common marmoset (*Callithrix jacchus*), we studied developmental changes in the acoustic structure of species-specific communication sounds produced in a social setting. We performed detailed acoustic analyses of the spectral and temporal characteristics of marmoset vocalizations during development, comparing differences between genders and twin pairs, as well as with vocalizations from adult marmosets residing in the same colony. Our analyses revealed significant changes in spectral and temporal features as well as variability of particular call types over time. Infant and juvenile vocalizations changed progressively toward the vocalizations produced by adult marmosets. Call types observed early in development that were unique to infants disappeared gradually with age, while vocal exchanges with conspecifics emerged. Our observations clearly indicate that marmoset vocalizations undergo both qualitative and quantitative postnatal changes, establishing the basis for further studies to delineate contributions from maturation of the vocal apparatus and behavioral experience. [DOI: 10.1121/1.2225899]

PACS number(s): 43.80.Ka [JAS] Pages: 1655–1670

I. INTRODUCTION

It is commonly held that relatively little learning is involved in primate vocal communication, particularly with regard to vocal production (Seyfarth and Cheney, 1997). Past studies on the ontogeny of vocalizations in primates showed few significant changes in the acoustic structure of vocalizations (Seyfarth and Cheney, 1986; Seyfarth and Cheney, 1997; Biben and Bernhards, 1995; Fischer, 2003; Inoue, 1988; Kalin and Shelton, 1998; Masataka, 1993; McCowan *et al.*, 2001; Merker and Cox, 1999; Newman, 1995; Newman and Symmes, 1974; Owren *et al.*, 1992, Owren *et al.*, 1993; Snowdon, 1997; Snowdon *et al.*, 1997; Winter, 1969, Winter *et al.*, 1973). One study showed morphological changes in the acoustic parameters of developing squirrel monkey vocalizations, but these changes were attributed strictly to maturation (Hammerschmidt *et al.*, 2001). Many of the studies investigating changes in acoustic morphology during vocal development did not employ extensive quantitative analysis methods due to the limitations of audio recording quality in the field and analytical techniques. A careful characterization of ontogenetic changes in the acoustic structure of vocalizations is a necessary first step for future experiments intended to delineate physical maturation from potential vocal production learning in nonhuman primates (Egnor and Hauser, 2004). Collecting and quantitatively analyzing a large sample set of vocalizations from multiple animals of a single species through the early stages of physical and social development is the best way to achieve this goal.

In the present study, we investigated the acoustic changes that occurred in vocalizations during the ontogenetic development of an arboreal New World Primate species native to the Atlantic coastal forests of Brazil, the common marmoset (*Callithrix jacchus*). Marmosets have a broad vocal repertoire and live in small groups, where only the dominant male and female breed but other group members participate in raising infants (Stevenson and Poole, 1976). Observations in the field suggest that common marmosets use vocalizations to maintain group cohesion, to announce food, to warn of predators, to announce territory to other groups, and to locate their own group members over distances (Epple, 1968; Moody and Menzel, 1976). In captivity, marmosets maintain an extensive repertoire of vocal communication sounds, some of which have been the subject of behavioral studies (Norcross and Newman, 1993, 1997; Schrader and Todt, 1993; Miller and Wang, 2006). Most of these studies investigated the Phee call because they are produced in a clear social context (isolation) and are easily elicited. Behavioral observations show that typically upon hearing a conspecific produce a Phee, marmosets will emit a Phee in response; this type of behavior is known as antiphonal calling. In their study, Miller and Wang (2006) showed that antiphonal calling in the common marmoset is dynamic; changes to the sensory input (i.e., the presence of a conspecific, identity and sex of the caller, and whether the occluded

[a)]Portion of this work presented in "Ontogenetic Changes in Vocal Production and Usage in Common Marmoset Monkeys," Society for Neuroscience Abstract No. 541.1(2001).

[b)]Author to whom correspondence should be addressed. Tel. (410) 614-4547; electronic mail: xiaoqin.wang@jhu.edu

individual participated in reciprocal antiphonal exchanges) result in modulations of vocal output. Other types of marmoset vocalizations that are produced in more ambiguous social contexts than Phee calls have not been systematically examined in any type of study.

The present study represents a comprehensive analysis of vocal development in marmosets. Quantitatively documenting developmental changes in the acoustics of marmoset vocalizations is the first step in creating a model for testing the influence of physical maturation, sensory experience, and learning in vocal development of a nonhuman primate. In this report, we examine the acoustic morphology of marmoset vocalizations produced during development. We accomplished this by taking advantage of the common marmoset's high vocal output rates in captivity to conduct vocal recordings in the social setting of a large breeding colony. Following these recordings, we then used advanced signal acquisition and processing techniques to analyze a large set of vocalization samples collected at multiple time points during development from single infants and twin pairs of different social groups within the colony. Utilizing sensitive acoustic analysis tools, we quantified changes in the acoustics of vocalizations and showed the course of normal vocal development in the common marmoset. Our results showed that marmoset vocalizations exhibit developmental changes in acoustic morphology.

II. METHODS

A. Subjects

This study included data from nine common marmosets (*C. jacchus*), born into a captive colony of approximately 50 of the same species. Eight were part of twin pairs and one was a single infant; subjects were followed from age 3 weeks to age 25 weeks, the period over which the most noticeable vocal production changes take place. The colony, which included family and social groups ranging in ages from infancy to ~12 years, was kept in spacious cages furnished with natural hardwood branches, nesting boxes, hammocks, and various other physical enrichment items.

B. Vocalization recording

We conducted all recordings without an experimenter in the room and video documentation was occasionally made for sessions with one twin pair. Infants were separated from their family group for brief intervals of 20 min to 2 h, depending on age: at 3–4 weeks of age, the time interval was 20 min; at 5–7 weeks, it was 40 min; at 8–12 weeks, it was 1 h; at 11+weeks, it was 2 h. They were placed in a custom-made recording cage surrounded by 3 in. acoustic absorption foam (Sonex, Illbruck). Each infant was placed in the recording cage individually, to ensure clean and classifiable vocalization samples. Recordings were taken using two microphones (AKG C1000S), one facing the infant and the other facing its family group. Microphone output signals were amplified (Symetrix SX202) and then recorded using a two-channel professional digital audio tape (DAT) recorder (Panasonic SV-3700 or TASCAM DA-40) at a sampling rate of 48 kHz.

C. Data analysis

Experimenters screened all vocalizations recorded on DAT tapes via both audio and visual display of real-time spectrogram using a PC equipped with RTS software (Engineering Design, CA). Segments of acoustic data containing vocalizations captured from DAT tapes were transferred onto a PC hard drive and analyzed using custom software written in MATLAB (MathWorks, MA) programming language. We used three levels of analysis with this custom software to process the vocalization data in increasing detail. Statistical tests were then used to determine the significance of changes in the vocalization acoustics of individual animals during development, and contributions of gender and twin pairing in these changes.

1. Digitizing and categorizing vocal signals

The first level of analysis utilized custom software to automatically screen through the saved sound files and extract vocalizations from the animal of interest based on specified parameters for signal and noise levels. Each extracted vocalization was manually examined to assure it is complete and free of noise. Vocalizations which passed quality criteria were then classified by acoustic features into adultlike major call types (Twitter, Trill, Phee, Trillphee) and additional call types (Egg, Ock, Tsik, and Compound Calls) (Epple, 1968; Agamaite, 1997). The three most commonly produced infant marmoset vocalizations that were unclassifiable as adultlike call types were noted in this study. We also observed various vocalizations which did not match the criteria as adultlike or the three major infant-specific calls. These vocalizations were variable in their acoustic structures and were not analyzed further.

2. Computing acoustic parameters

Acoustic features in each of the four major call types were quantified based on those previously analyzed in the adult marmoset in our laboratory (Agamaite, 1997), with modifications to detect novel features only observed in infants. Parameters analyzed included spectral, temporal, and amplitude characteristics (see Table I for a list of parameters).

3. Statistical analysis of vocalization features

Statistical analysis of vocalization features was carried out by call type, age, gender, and individuals. Trends of group data were calculated by taking average parameter values across animals for each week. Two variations of this calculation were made. For the individual mean (IM), the average was calculated by first taking the mean of each animal. Using the individual mean data, slopes of change over time for each animal were computed for each parameter and for each call type using the method of least squares. These individual means were also averaged to obtain the group mean, where each monkey factors equally. For the population mean (PM), the average was calculated from all calls recorded from every animal. No weighing is given to offset the number of calls produced by each animal.

TABLE I. Vocalization parameters. Parameter names are listed in the left column, with the corresponding parameter number for each of the four major call types (Phee: 1–30, Trill: 1–36, Trillphee: 1–36, and Twitter: 1–28), followed by a description of the parameter, in subsequent columns. Parameters with no corresponding number for any of the vocalizations indicate that parameter does not apply to the vocalization type.

Parameter	Phee	Trill	Trillphee	Twitter	Additional Description
Duration	1	1	1	1	Total duration of a call
$\text{Time}_{\text{AmpHi}}$	2	2	2	2	Time of the absolute high amplitude for the entire call
Front Amp_{Mean}	3	3	3	3	Relative mean amplitude, first phrase for the Twitter
Front Freq_{Dom}	4	4	4	4	Dominant frequency, first phrase for the Twitter
Front Freq_{QLo}	5	5	5	5	Quarter-level low frequency, first phrase for the Twitter
Front Freq_{QHi}	6	6	6	6	Quarter-level high frequency, first phrase for the Twitter
Front Freq_{Min}	7	7	7	7	Absolute minimum frequency, first phrase for the Twitter
Front $\text{Freq}_{\text{Mean}}$	8	8	8	8	Absolute mean frequency, first phrase for the Twitter
Front Freq_{Max}	9	9	9	9	Absolute maximum frequency, first phrase for the Twitter
Middle Amp_{Mean}	10	10	10	10	Relative mean amplitude, middle phrase for the Twitter
Middle Freq_{Dom}	11	11	11	11	Dominant frequency, middle phrase for the Twitter
Middle Freq_{QLo}	12	12	12	12	Absolute minimum frequency, middle phrase for the Twitter
Middle Freq_{QHi}	13	13	13	13	Quarter-level high frequency, middle phrase for the Twitter
Middle Freq_{Min}	14	14	14	14	Absolute minimum frequency, middle phrase for the Twitter
Middle $\text{Freq}_{\text{Mean}}$	15	15	15	15	Absolute mean frequency, middle phrase for the Twitter
Middle Freq_{Max}	16	16	16	16	Absolute maximum frequency, middle phrase for the Twitter
End Amp_{Mean}	17	17	17	17	Relative mean amplitude, last phrase for the Twitter
End Freq_{Dom}	18	18	18	18	Dominant frequency, last phrase for the Twitter
End Freq_{QLo}	19	19	19	19	Absolute minimum frequency, last phrase for the Twitter
End Freq_{QHi}	20	20	20	20	Quarter-level high frequency, last phrase for the Twitter
End Freq_{Min}	21	21	21	21	Absolute minimum frequency, last phrase for the Twitter
End $\text{Freq}_{\text{Mean}}$	22	22	22	22	Absolute mean frequency, last phrase for the Twitter
End Freq_{Max}	23	23	23	23	Absolute maximum frequency, last phrase for the Twitter
Absolute Freq_{Max}	24	24	24		Maximum frequency over the entire call
Time Abs Freq_{Max}	25	25	25		Time into the call of the maximum frequency
Absolute Freq_{Min}	26	26	26		Minimum frequency over the entire call
Time Abs Freq_{Min}	27	27	27		Time into the call of the minimum frequency
$\text{Frequency}_{\text{start}}$	28	28	28		First measurable frequency in a call signal
$\text{Frequency}_{\text{End}}$	29	29	29		Last measurable frequency in a call signal
$\text{Time}_{\text{Transition}}$	30	30	30		Time of transition from Trilling to Pheeing (value=0 for Phee)
$\text{Time}_{\text{Period}}$		31	31		Time of a trilling cycle period
$\Delta\text{Freq}_{\text{Max}}$		32	32		Maximum frequency change in one trilling cycle
Time $\Delta\text{Freq}_{\text{Max}}$		33	33		Time it takes for the maximum frequency change in a cycle
$\Delta\text{Freq}_{\text{Min}}$		34	34		Minimum frequency change in one trilling cycle
Time $\Delta\text{Freq}_{\text{Min}}$		35	35		Time it takes for the minimum frequency change in a cycle
$\Delta\text{Freq}_{\text{Mean}}$		36	36		Mean frequency change in one trilling cycle
IPI				24	Average inter-phrase interval for all phrases of a Twitter call
Number Phrases				25	Total number of phrases in a Twitter call
$\text{Time}_{\text{Sweep}}$ $\text{Phr}_{\text{First}}$				26	Time for the first phrase frequency sweep
$\text{Time}_{\text{Sweep}}$ $\text{Phr}_{\text{Middle}}$				27	Time for the middle phrase frequency sweep
$\text{Time}_{\text{Sweep}}$ Phr_{Last}				28	Time for the last phrase frequency sweep

a. Distance measures: A useful comparison of vocalizations between infant and adult groups, as well as between individual animals at a specific age, can be made using distance measurements in a multidimensional space, where each dimension represents a particular vocalization feature. We compute distances between comparison groups using multidimensional Euclidean distance. In a two- or three-dimensional space, the Euclidean distance is the geometric distance between two points. Because the Euclidean distance can be biased by differences in scale of each of the dimensions used to compute it, it is good practice to transform the dimensions so they have similar scales. Here we measured the Euclidean distance between two groups of data using parameter values normalized by the variance along each dimension, as follows:

$$D(x,y) = \sqrt{\sum_{i=1}^{N} \left(\frac{x_i - y_i}{\sigma_i} \right)^2}, \tag{1}$$

where D is the N-dimensional Euclidean distance between data group X and group Y; x_i and y_i are parameter values on the ith dimension for each data group, respectively; σ_i is the variance of the parameter values along the ith dimension.

b. Differences between gender, sibling pair, and age groups: An analysis of nonindividual factors in data trends is accomplished using statistical covariance tests. Animals were either grouped by gender or sibling pairs and their vocalization trends (both by absolute parameter values and by distance from adulthood) were compared with analysis of

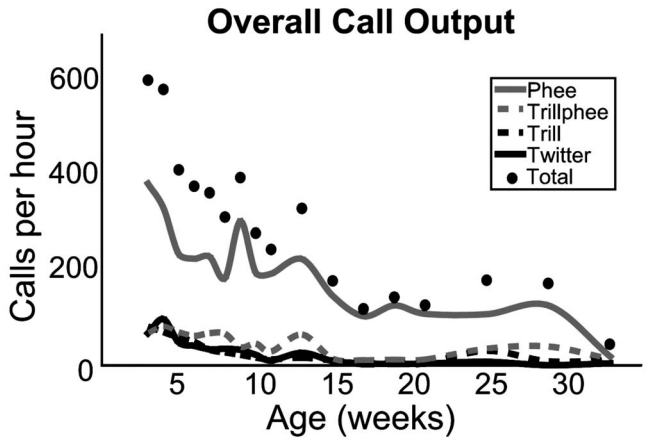


FIG. 1. Population average of call output (in calls per hour) as a function of age (in weeks) for all infant marmosets used in this study. Both the total number of calls (black circles) and the number of calls for each of the four major call types are shown.

covariance (ANCOVA). Using a confidence level of 0.05, the likelihood that vocalization trends were similar between gender groups and sibling pairs was assessed along multiple dimensions.

In evaluating the significance of differences between mean values in call parameters, we also grouped vocalization samples by age and gender using individual parameter distributions. Significance in mean difference between each set of groups was assessed using a two-tailed t-test at both alpha=0.05 and alpha=0.01.

Principal component analysis (PCA) was used to determine the call parameters that most explain variance between age groups. Vocalizations from 5 weeks and 25 weeks were used to simplify the process while maintaining broad age separation. Weighting vectors were found that described a transform to project the data into an orthogonal basis set describing maximum variance. The highest weighted parameters in the principal component vector explained the most amount of variance.

c. Discriminability of vocalizations between individual animals: Discriminability of vocalizations between individual animals was determined using a discrimination algorithm. The vocalizations were first reduced to a two-dimensional representation using the first two PCA components. Then, for all four major call types, the mean and distribution of principal components for each animal's set of vocalizations were calculated. Using this information, each individual vocalization's distance is measured from the animal means, normalized by standard deviation, and a minimum distance is found. A vocalization is considered correctly classified if the smallest distance belongs to the mean of animal from whom the vocalization was recorded.

III. RESULTS

A. Qualitative description of general observations

1. Overall call output

Production of all call types decreased in occurrence with advancing age. The overall call output averaged over the population of infants from this study is plotted in Fig. 1. The Phee call is the most commonly-produced call, while the other three major call types, often used in close (Trill, Trillphee) or distant (Twitter) contact situations, were infrequently produced in older animals physically isolated from their family group but with visual contact maintained.

2. Vocal exchanges between infant subjects and other conspecifics

For vocalizing infants aged 3–8 weeks, there were no discernable vocal responses from adults. Infant subjects often vocalized shortly after (≤1 s) vocalizations by other marmosets with the same call type. The infant subjects rarely decreased or ceased vocal output when other animals in the colony alarm-called or became suddenly silent. Several examples of infant responses to calls made by adult conspecifics are shown in Fig. 2. Infant responses occurred within an average of 600 ms after the offset of the conspecific call in the examples shown in Fig. 2(a) that include all four major call types. Figure 2(b) shows examples of infants vocalizing Phee strings (which sometimes began with a Trillphee) in response to Phees made by adult animals in the colony.

As the infants grew into juveniles, and then adults, there gradually emerged more directed responses based on conspecific vocalizations, and juveniles became silent in response to colony alarms. Responses of conspecifics to the subjects

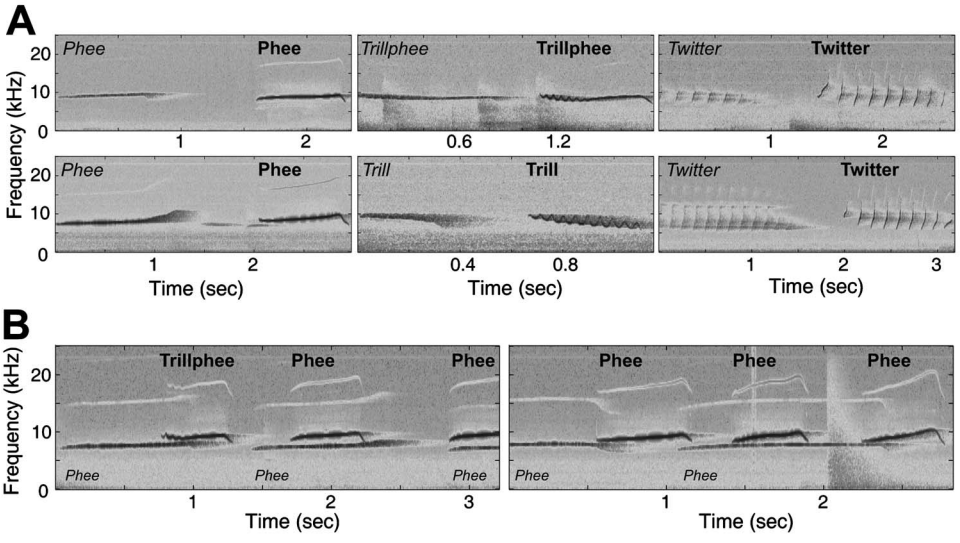


FIG. 2. Examples of 3–4 week old infant vocal responses to adult conspecifics. (A) Spectrograms, labeled for call type (*italic*: adult vocalization; *bold*: infant response) are shown for the four major call types. Each plot shows one instance of adult-infant vocalizations. (B) Two instances of Phee strings made by an infant in response to Phee calls from an adult are shown. Note the higher frequency of infant calls as compared to adult calls in both plots and the timing of infant calls (lagged behind adult calls).

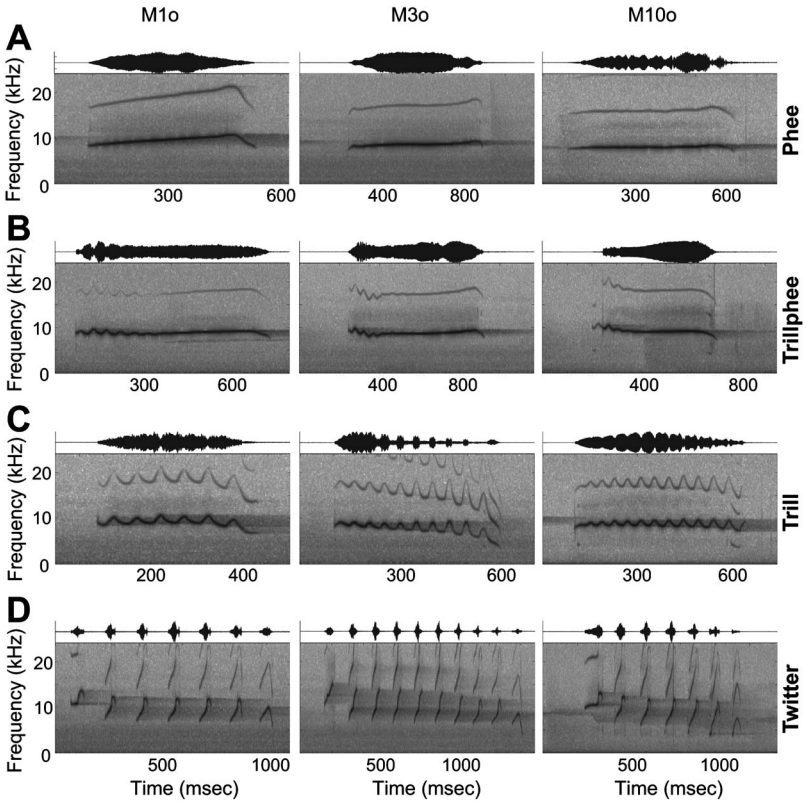


FIG. 3. Spectrogram and amplitude waveform exemplars are shown for the four major call types: (A) Phee; (B) Trillphee; (C) Trill; (D) Twitter, from three different (unrelated) infants (M1o, M3o, M10o) at 3–5 weeks of age.

were difficult to determine in this study, because we did not perform behavioral scoring and the vocal output of juveniles included comparatively fewer calls than that of infants.

B. Call types and acoustic characteristics in marmoset infants

Most common marmoset calls are tonal and can be grouped into call types that remain consistent across individuals. To analyze the vocal development of each call type in infant marmosets, we used parameters defined for the normal adult calls of the same type and quantified changes in each parameter across different ages in the same animals. Additional parameters were used to accommodate acoustic features unique to infant calls.

1. Adultlike call types and novel features in infants

Infant marmosets begin vocalizing shortly after birth and produce some call types which are acoustically similar to the call types produced by adults. Figure 3 shows examples of the four major call types (Phee, Trillphee, Trill, and Twitter), also produced by adult animals, made by three unrelated infant marmosets at 3–5 weeks of age. Both spectrograms and amplitude waveforms are shown. The Phee is a tonelike long call; the Trillphee is similar to a Phee, but with sinusoidal-like amplitude modulation (AM) and frequency modulation (FM) during the first portion of the call. Trill calls are defined by sinusoidal-like AM and FM throughout the entire call. Twitters consist of multiple phrases of upward FM sweeps, some of which ended by brief downward FM sweeps. The first and last phrases of twitters are highly variable in structure.

Progressive changes over time for the four major call types are shown by representative examples in Fig. 4. For example, the Phee and Trillphee calls become longer in duration and lower in frequency over time and the FM depth decreases for the Trillphee and Trill calls. Several vocalization features are seen often in infants, but are apparently not produced in adults. One such feature is the addition of spectral components outside the predominant harmonic structure, presumably related to immaturity of the vocal apparatus [Fig. 4(a), 4-week Phee example, closed arrow]. Another infant-specific feature is the "twitter-hook," a short downward FM sweep at the end of each upward FM sweep in each phrase of a Twitter call [Fig. 4(d), open arrow]; the hooking fades as an animal matures. All infant marmosets in this study exhibited this feature at the earliest time of observation (3–5 weeks) but varied in the extent of the downward FM sweep, its temporal span, and the age at which it faded.

We also observed other, less frequently produced, adultlike call types in infants. Figure 5(a) shows examples of Peeps, similar to a Phee in frequency and structure, but significantly shorter in duration. Figure 5(b) shows a commonly-heard alarm call, the Tsik, and one example of a Tsik followed by an Egg (short-duration harmonic stack, indicated by an arrow), a configuration seen in adults as well. Figure 5(c) shows examples of compound calls, occasionally observed in adults, where the infant uses multiple call types strung together with little or no intersyllable interval.

2. Infant-specific calls

While marmosets produced the four major adultlike call types at all stages of development, three dominant patterns of vocalization emerged in infant marmoset vocalizations: the Cry, Compound Cry, and call strings or "Babbling." The Cry is an extremely broadband, long-duration call [Fig. 6(a)]. In-

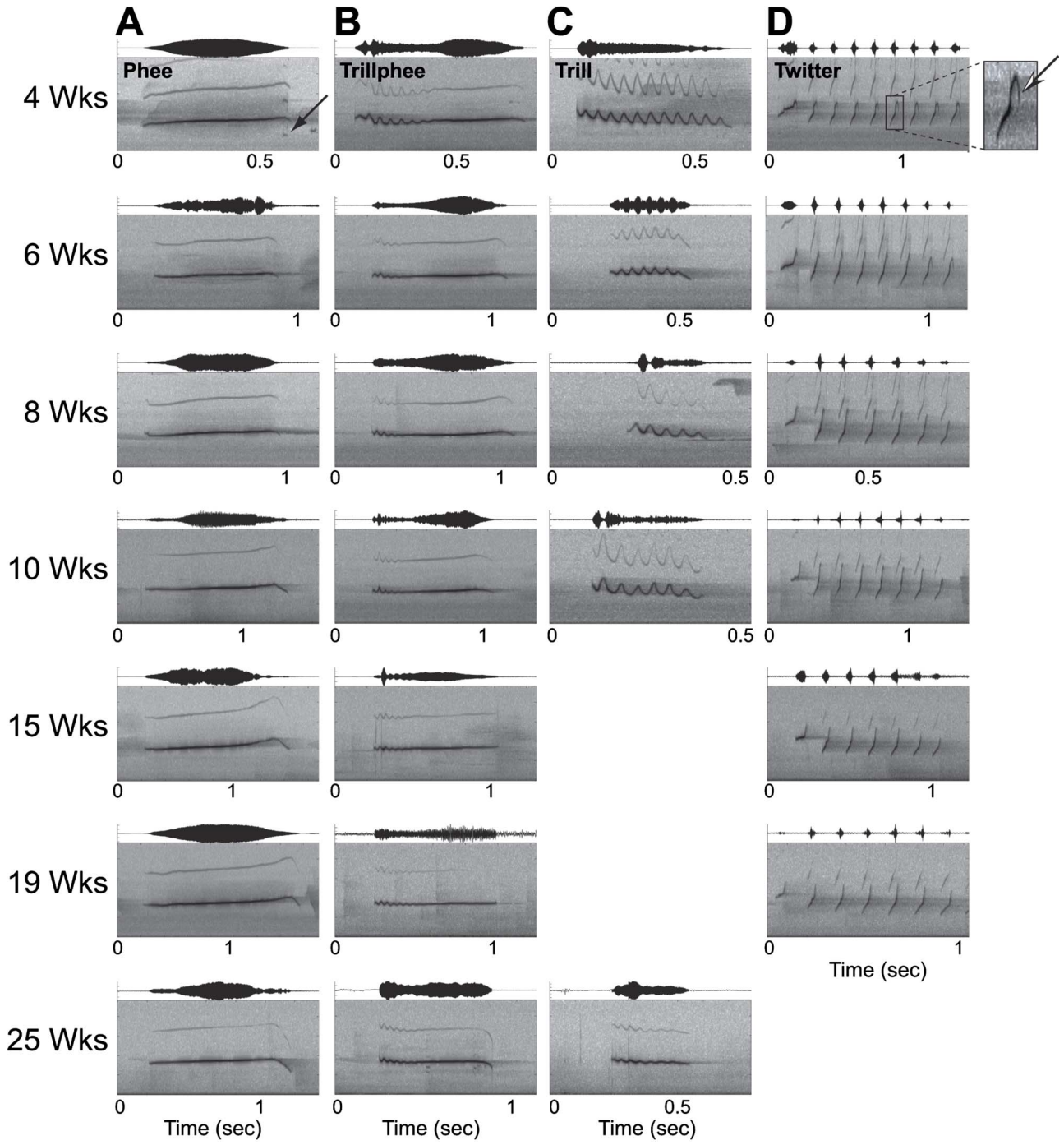


FIG. 4. Exemplars showing the progression of the four major call types as one individual marmoset (m10o) grew (4, 6, 8, 10, 15, 19, and 25 weeks of age). (A) Phee; (B) Trillphee; (C) Trill; (D) Twitter. Closed arrow indicates the appearance of spectral components outside the predominant harmonic structure of the Phee call. Open arrow indicates the "twitter hook."

fant marmosets often use the Cry in compound calls with other call types, as seen in Fig. 6(b). Another infant-specific vocal behavior is the production of continuous strings of multiple call types which can last for minutes. Figure 6(c) shows 38 s of one such string. These call strings are similar in nature to the "babbling" seen in pygmy marmosets by Snowdon and Elowson (2001). At ages up to 6 or 7 weeks, the Cry, Compound Cry, and Babbling were the most common vocalizations uttered by marmosets. However, the Babbling abruptly disappeared, usually by the seventh week, and the use of the Cry and Compound Cry faded gradually and disappeared entirely by 10–11 weeks in all animals in the study. These infant-specific call types or behaviors were not observed in a previous study on classification and acoustic quantification of adult marmoset call types from this same captive colony (Agamaite, 1997).

C. Quantification of changes in vocalization features over time

For each of the four major adult-like call types, we quantified changes in a set of spectral and temporal param-

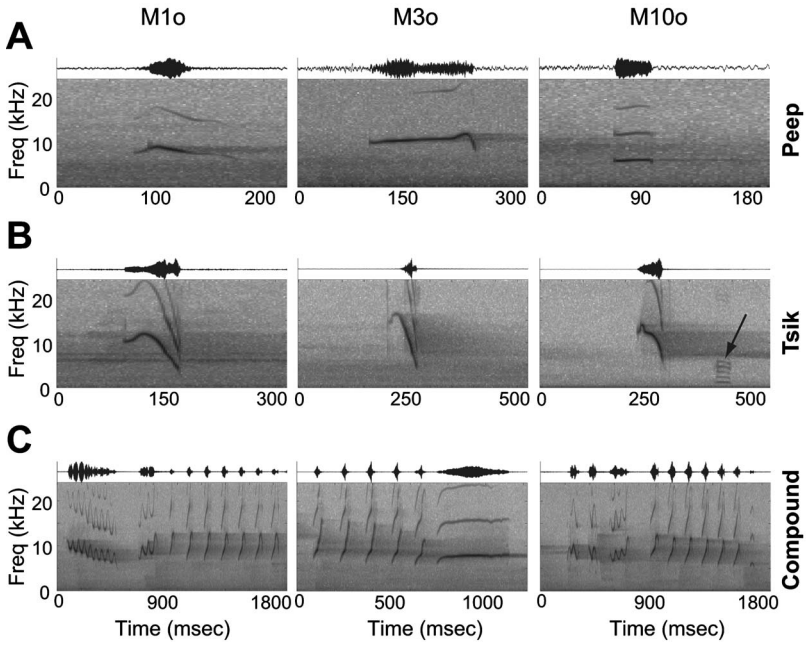


FIG. 5. Spectrogram and amplitude waveforms of other adult-like call types; (A) Peep; (B) Tsik; (C) Compound, produced by three different (unrelated) infants (M1o, M3o, M10o) at 3–5 weeks of age. The arrow [(B) third column] highlights an Egg call, a harmonic stack, occurring after the tsik.

eters (see Table I) as an animal aged. We then compared the differences between individuals, gender, and siblings in their call parameters at different ages, identifying which features become more dissimilar and which features become more stereotyped over time.

1. Individual subjects

Examples of the typical acoustic changes of vocalizations that occurred over time in the four major call types are shown in Fig. 7. Each line represents the least-squares fit of the individual mean (IM) for one vocalization parameter measured at discrete time points from one of the infant marmosets in this study. While some spectral and temporal parameters changed over time at similar rates (i.e., with similar slopes of lines in Fig. 7) for all sampled animals, other parameters showed different changes over time among different animals. We further quantified the rate of acoustic changes over time by computing the slope of least-squares fit for each vocalization parameter measured at different ages.

Figures 8 and 9 show the distribution of the slopes of the least-squares fit for each measured parameter of each call type for all individual animals using the IM method. These boxplots are representative of the rate of change in Hz per week for spectral parameters, seconds per week for temporal parameters, and relative dB per week for amplitude parameters. For each parameter, the area within the box corresponds to the middle 50% of the slopes for all animals, with the median indicated by the black center line. Outliers ($\geqslant$1.5 times the range from the upper and lower quartile) are marked by the short lines outside of the whiskers. The rates of change for all measured spectral parameters of the four major call types are shown in Fig. 8. Slopes of the measured temporal parameters are plotted in Fig. 9(a) and slopes of the measured relative amplitude parameters are plotted in Fig. 9(b).

Phee [Figs. 7(a), 8(a), 9(Aa), and 9(Ba)]: Most spectral parameters for the Phee [Fig. 8(a)] showed trends for each animal in the same general direction, but to varying degrees. Overall, across the population, most frequency parameters for the Phee decreased in their respective values. Temporal parameters for the Phee varied in magnitude of slope for different animals but all were in a positive direction except for one parameter [Fig. 9(Aa)]. A consistent trend across animals for amplitude parameters was that the beginning of the Phee became lower in relative amplitude [param. 3, Fig. 9(Bb)], while the end of the Phee increased in amplitude [param. 17, Fig. 9(Bb)] as the animals aged.

Trillphee [Figs. 7(b), 8(b), 9(Ab), and 9(Bb)]: Spectral parameters of Trillphee [Fig. 8(b)] are less variable than the Trill [Fig. 8(c)] but more so than the Phee [Fig. 8(a)]. Although Trillphee frequency parameters largely showed a downward slope, several animals increased frequency parameters as they aged (beginning of call: param. 4–9; middle of call: param. 11–16; end of call: param 18–23). As in the Phee, the duration (param. 1) and time of maximum amplitude (param. 2) increased as animals grew older [Fig. 9(Ab)]. However, unlike the Phee, the time of high frequency (param. 25) decreased and the time of low frequency (param. 27) increased, on average, as animals matured [Fig. 9(Ab)]. Relative amplitude parameters for the Trillphee showed little change overall across individual animals, with the most consistent change being an increase in amplitude at the beginning of the call (param. 3) relative to the middle (param. 10) and end of the call (param. 17) [Fig. 9(Bb)].

Trill [Figs. 7(c), 8(c), 9(Ac), and 9(Bc)]: Frequency parameter changes in the Trill [Fig. 8(c)] for each section of the call were widely variable across individuals. There were outliers with high positive slope values for all nonmodulation-related spectral measures. On average, however, the slope showed little change or a decrease in these parameters as animals aged. The modulation frequency measures (param. 32, 34, and 36) also decreased, indicating a sinusoidal shape of consistently decreasing bandwidth as animals matured. Temporal parameters for the Trill [Fig. 9(Ac)] showed some

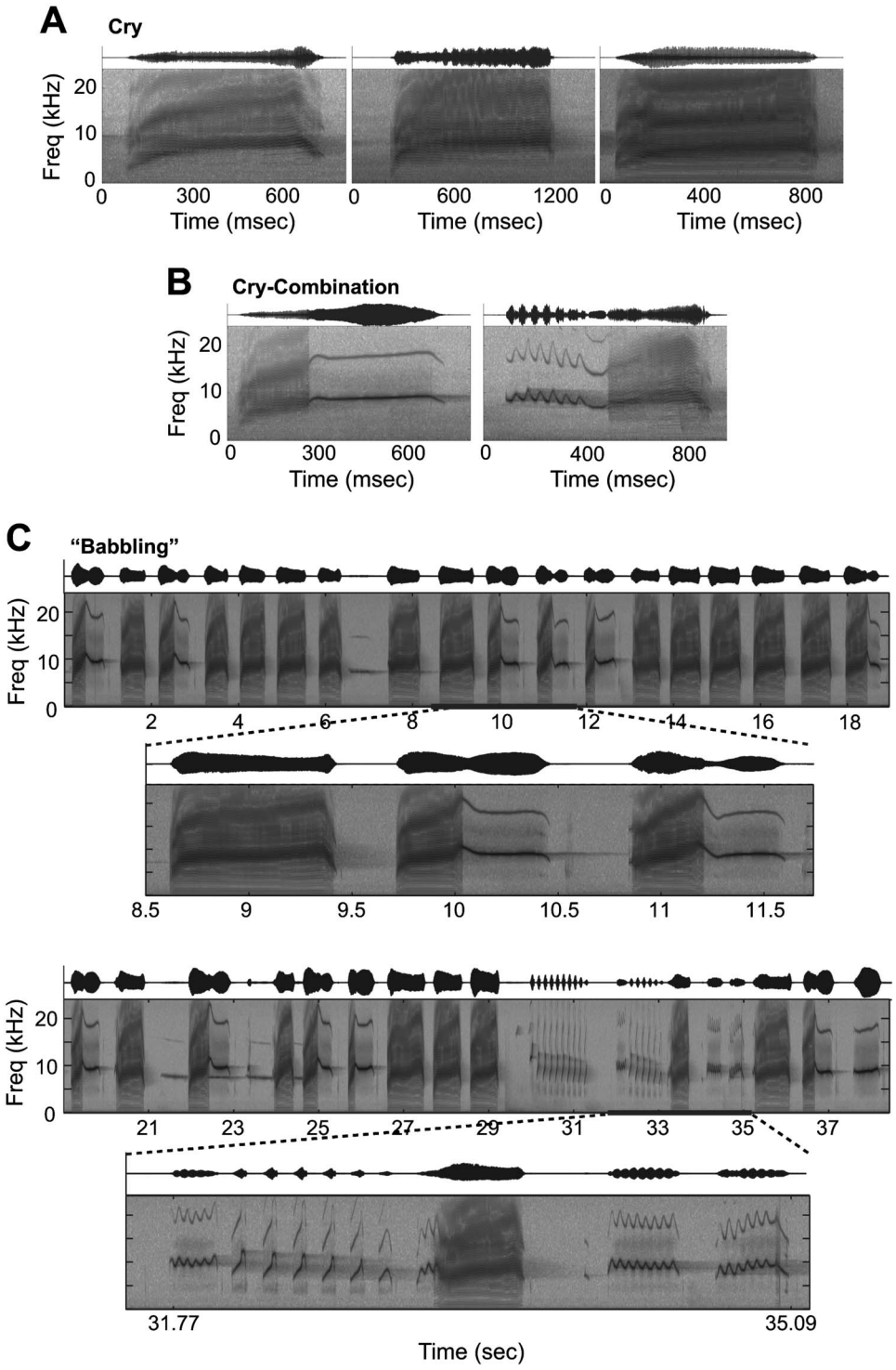


FIG. 6. Example call types produced exclusively by infants. (A) Cry calls; (B) Cry-combination calls; (C) "Babbling:" a 38-s continuous string of vocalizations broken down into two parts for ease of visualization. A small segment of each part is magnified to show detail.

changes similar to those seen in the Trillphee, but to a lesser extent. The range of slope change for two temporal parameters associated with the frequency modulation (param. 33 and 35) was significantly broader than that seen in the Trillphee. The average relative amplitude [Fig. 9(Bc)] of the middle portion of Trill (param. 10) decreased, although its range was highly variable, and the average amplitude of the end portion of the Trill (param. 17) increased.

Twitter [Figs. 7(d), 8(d), 9(Ad), and 9(Bd)]: The spectral structure of the first phrase of each utterance of a Twitter call is extremely variable, contributing to the wider variation in slopes for parameters 3–9 [Fig. 8(d)]. The middle phrase, however, is more stable in comparison and frequency parameter slopes (param. 11–16) indicate an overall decrease in frequency over time for most. The Twitter duration increased for all but one animal, which showed a very small decrease, and the time of maximum amplitude had a positive slope for all animals [Fig. 9(Ad)]. For the Twitter call, most animals

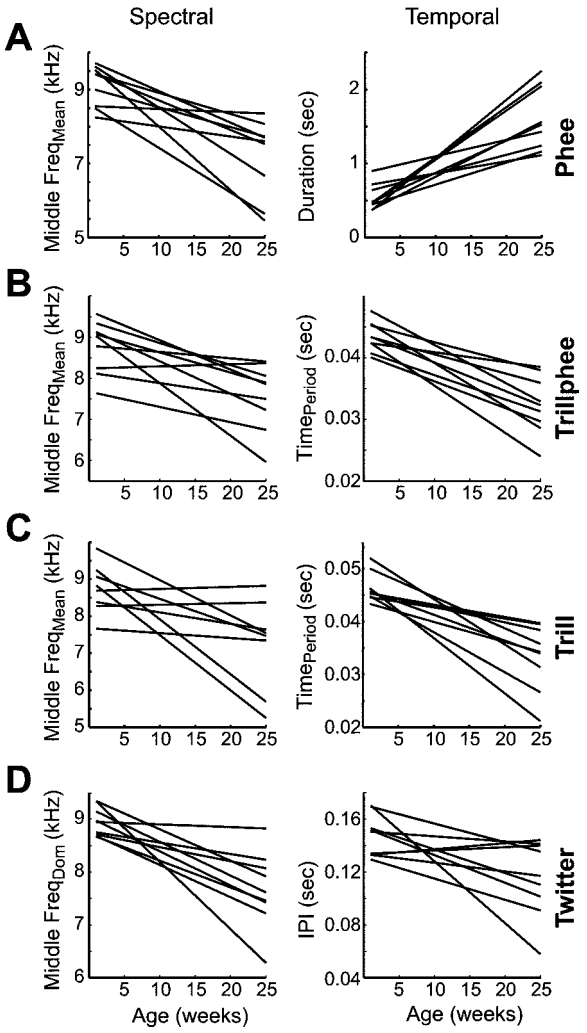


FIG. 7. Least-squares regression analysis of representative spectral (left) and temporal (right) parameters for the four major call types: (A) Phee; (B) Trillphee; (C) Trill; (D) Twitter, for all studied individuals. Parameter value is on the y axis and age (in weeks) is on the x axis. Each best-fit line corresponds to IM data from an individual animal as it aged.

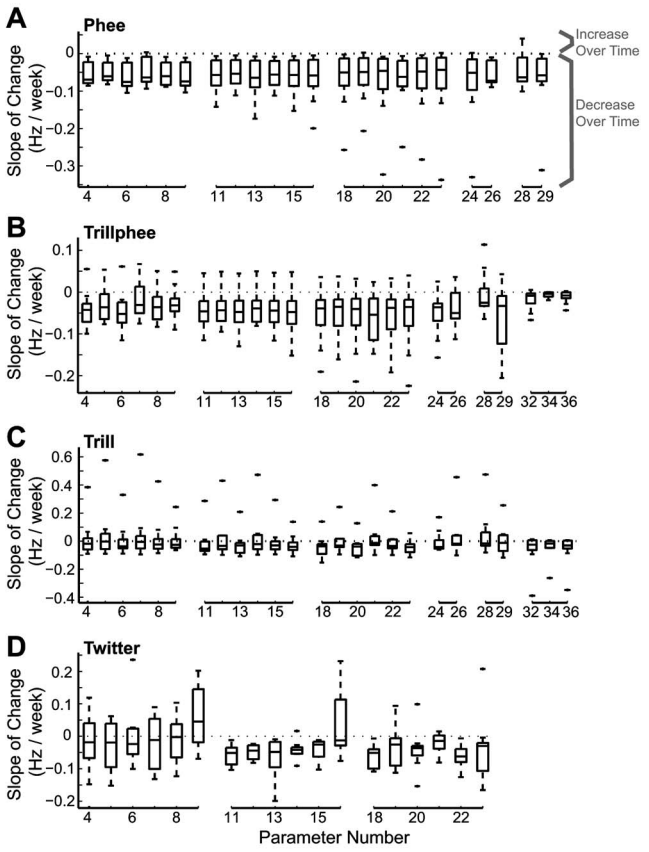


FIG. 8. Overall developmental trends of all measured spectral parameters for the four major call types: (A) Phee; (B) Trillphee; (C) Trill; (D) Twitter is analyzed by the slope of the least-squares regression line (see Fig. 7) of each parameter for each individual animal. The slope (indicating the magnitude and direction of the change in a parameter over time, in Hz/week) is on the y axis and the parameter number (see Table I) is on the x axis. The horizontal dotted line at $y=0$ indicates the dividing point where data above the line represent a positive slope and data points below the line indicate a negative slope, as shown on the right of plot (A). For each parameter, mean slope values from all studied individual animals are shown as boxplot. The median of all data points for each parameter is indicated by the black centerline within the box, and the first and third quartiles are the top and bottom edges of the box. Extreme values ($\leqslant$1.5 times the range from the upper and lower quartile) are the ends of the lines extending from the box. Points $\geqslant$1.5 times the interquartile range (outliers) are plotted as individual points. Boxplots are subdivided into groups (as indicated by a break in the x axis) representing beginning, middle, and end of the call, as well as overall parameters (such as start/end frequency and min/mean/max frequency).

showed a decrease in relative amplitude of the first and last phrases as they aged, while the middle phrase relative amplitude increased in all but one animal, [Fig. 9(Bd)].

2. Population data

For the entire population of tested animals, we show several example parameters of each of the four major call types in Fig. 10 to illustrate the general trends of individual acoustic features as an animal develops. The individual (IM) and population (PM) means are shown for comparison in each plot. Note that for many parameters shown, at early time points, the population and individual means have similar values (and individual mean error is small) but they become more different (and individual mean error becomes larger) as the animals get older. Population measures of parameter changes for the Trillphee [Fig. 10(b)] and Trill [Fig. 10(c)] calls show greater variations among individuals, especially as they became older, as compared to the Phee. The population parameters for the Twitter [Fig. 10(d)] became more variable as animals aged, as indicated by the increased error bars.

3. Comparison of major call types between infants and adults

Infant data at 5 weeks was compared with data previously obtained from the same captive colony in adult marmosets (Agamaite, 1997) for the same parameters across the four major call types. Examples of individual parameter distributions are shown in Fig. 11. In the call parameters where distributions differed between infants and adults, the individual mean and distributions of the measured parameters for infants always trended towards the adult mean and distribution as they aged.

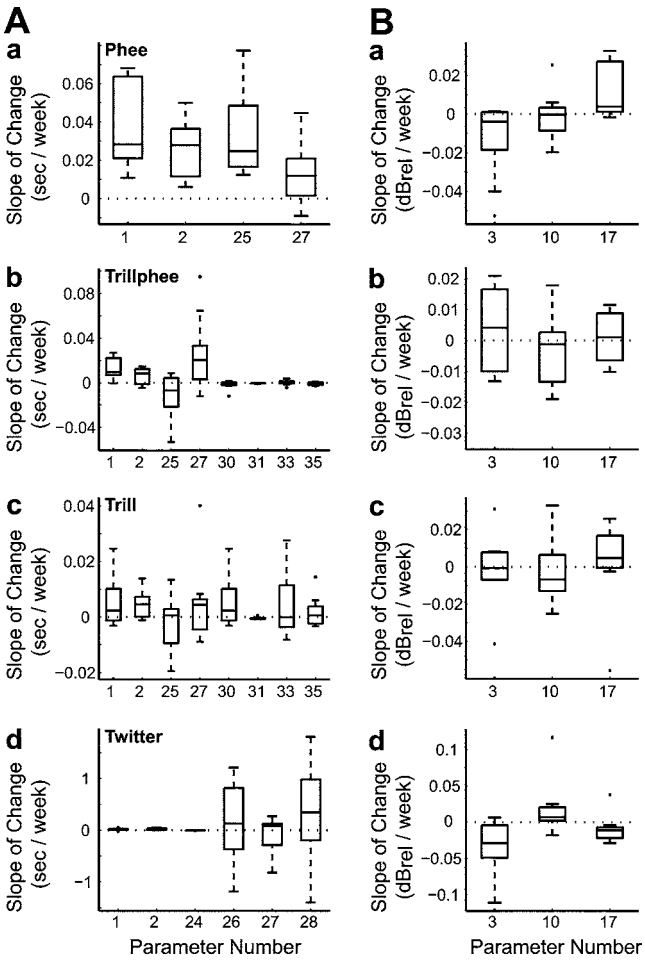


FIG. 9. Overall developmental trends of all measured temporal (left column) and all measured amplitude (right column) parameters for the four major call types (A) Phee; (B) Trillphee; (C) Trill; (D) Twitter. The format is the same as in Fig. 8, except there are no breaks in the x axis because each parameter number is labeled.

Euclidean distance [as defined in Eq. (1)] was calculated using the 18 spectral parameters common to all four major call types between the infants at different ages versus themselves at 25 weeks and between the same infants and adults from an earlier study (Agamaite, 1997). The results show that parameter changes are very similar between the group comparisons [Fig. 12(a)]. Generally, the Phee showed the largest and most rapid change in parameter measurements over time, while the Trillphee and Trill showed intermediate changes. The Twitter showed only a small, gradual change. The percent change for these comparisons, shown in Fig. 12(b), is also not significantly different between groups. These data indicate that the measured acoustic parameters of marmoset vocalizations stabilize by the time animals reach the age of about 25 weeks.

The standard deviation (SD) of call parameters between all studied infants is plotted against those of adults studied earlier (Agamaite, 1997), highlighting a few important trends (Fig. 13). First, most temporal and amplitude parameters for the four major call types show SD values near or below 0.5, while spectral parameters often show higher SD values (>0.5). Secondly, the temporal and amplitude parameters of the Phee show a higher SD for adults than for infants, as do most of the spectral parameters. The standard deviations for spectral parameters of the Trill and Trillphee cluster closely, with infant SD being higher in Trillphee parameters and adult SD being higher for most parameters of the Trill. Spectral parameters for the Twitter show a highly variable SD, with the majority being higher for the infants than for adults.

4. Individual differences

We performed PCA for all measured parameters of each call type, comparing infant vocalizations (5 weeks of age) with late juvenile vocalizations (25 weeks of age) from the same group of animals (Fig. 14). For all call types, the first principal component (PC1) consists largely of frequency pa-

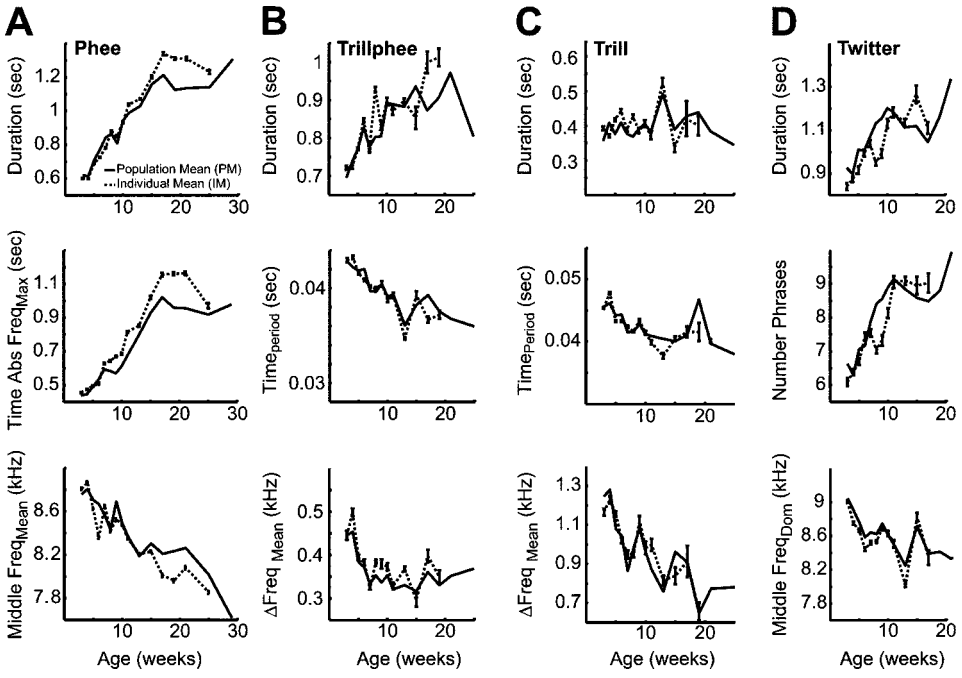


FIG. 10. Mean data from discrete age time points are plotted for select parameters of the four major call types: (A) Phee; (B) Trillphee; (C) Trill; (D) Twitter. Parameter value is on the y axis and age (in weeks) is on the x axis. The solid black line is the PM, and the dotted black line is the IM with standard error.

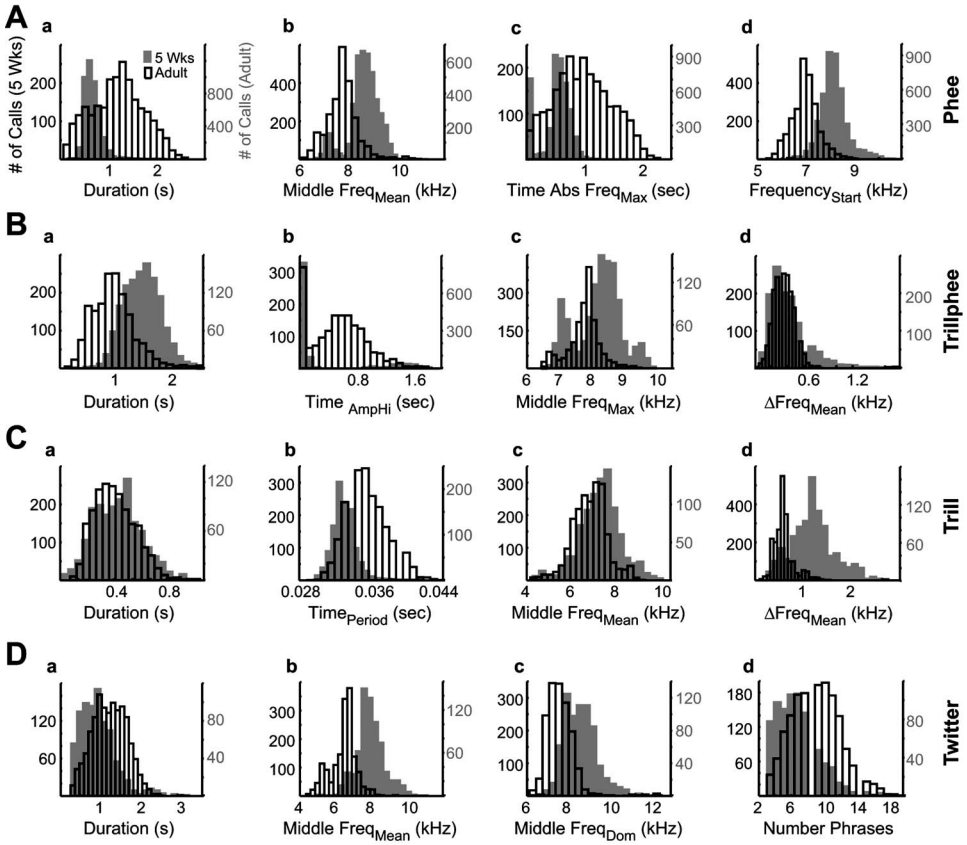
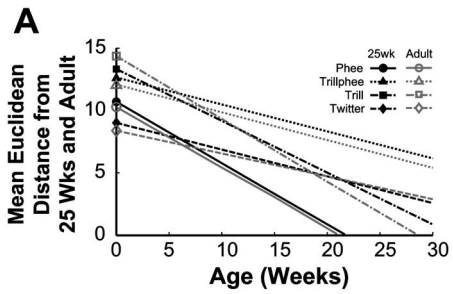


FIG. 11. Comparison between vocalizations of infant (5 weeks) and adult marmosets. Representative parameters of the four major call types: (A) Phee; (B) Trillphee; (C) Trill; (D) Twitter, from all vocalization samples are shown as histograms. Filled grey bars are the 5-week data and open bars are the adult data. The parameter value is on the *x* axis, the number of calls for 5-week-old infants is on the left *y* axis, and the number of calls for the adults is on the right *y* axis.



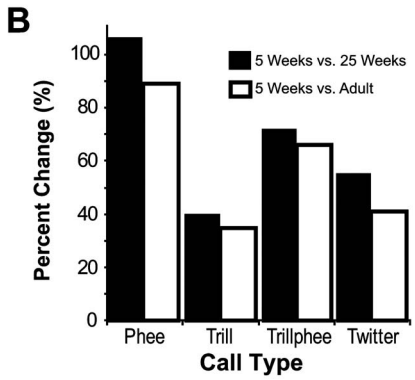


FIG. 12. Measures of parameter stabilization of animals in this study. (A) Mean Euclidean distance [as defined in Eq. (1)] (based on 18 frequency parameters) between infant marmosets at different ages and themselves at 25 weeks (solid lines and black symbols), and between infant marmosets and other adult animals (dashed line and open symbols), respectively, are plotted for the four major call types. (B) Percent change of the Euclidean distance measure between 5 weeks and 25 weeks (filled bar) or 5 weeks and adults (open bar).

rameters (black bars) and temporal parameters (grey bars) became more significant in the second principal component (PC2). In the Phee, PC1 accounts for nearly 70% of the variance. PC1 accounts for nearly 50% of the variance in the Trillphee, and, for the Trill, it accounts for nearly 40%. For the Twitter, the most acoustically complex of the four major call types, the PCA shows that PC1 and PC2 account for less variance than their counterparts for other three call types. Both PC1 (~28% of variance) and PC2 (~22% of variance) were dominated by frequency parameters, though to a lesser degree than those of other call types.

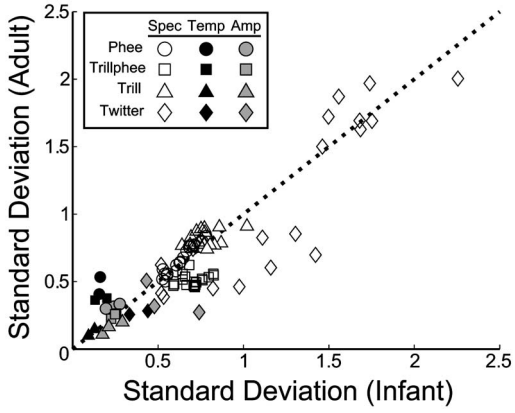


FIG. 13. Standard deviations of all measured parameters are plotted for 5-week-old infant marmosets (*x* axis) vs adults (*y* axis) for the four major call types. The diagonal line has a slope of 1. Symbol shape indicates the call type (circle: Phee; square: Trillphee; triangle: Trill; diamond: Twitter) and shading indicates the class of parameter (open: spectral; black: temporal; grey: amplitude).

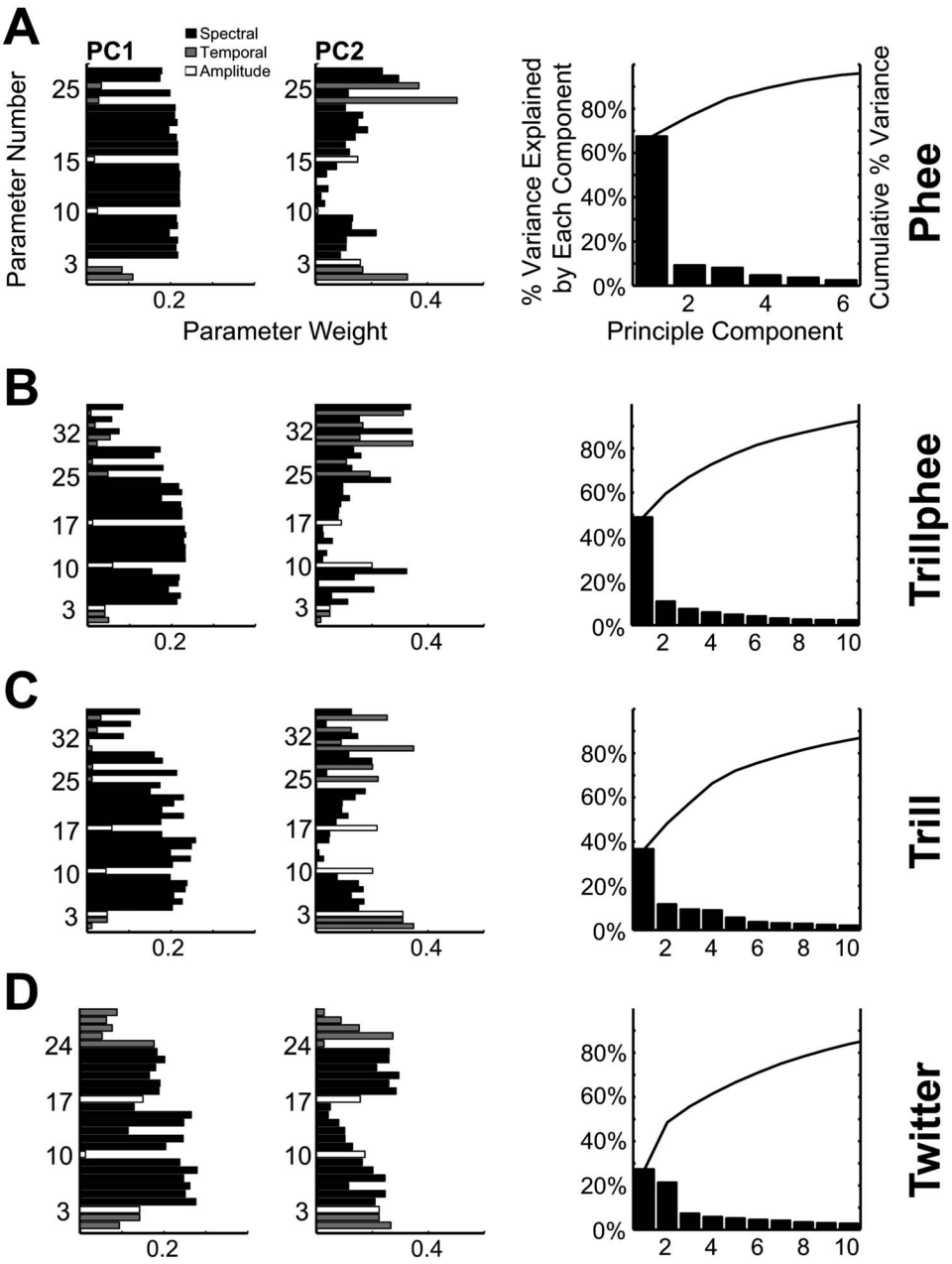


FIG. 14. Principal component analysis comparing vocalization parameters of infant (3–6 weeks) with late juvenile (19–25 weeks) for the four major call types: (A) Phee; (B) Trillphee; (C) Trill; (D) Twitter. The weights of the first (left) and second (center) principal components (PC1, PC2) are plotted in x axis number is shown on the y axis. Black bars indicate spectral parameters, grey bars indicate temporal parameters, and open bars indicate amplitude parameters. The percent of variance explained by each principal component, for principal components 1–10, is plotted on the right, with the line indicating the cumulative percent variance.

A PCA-based Discriminability Index was computed for each call type using "young" (all calls from age 3–6 weeks) versus "old" (all calls from 19 to 25 weeks) and is plotted in Fig. 15. The percent correct discrimination attributed to chance falls at 20% or 25% (dotted line on each plot), depending on the number of animals included. While all call types show above-chance discriminability between individuals at both designated age clusters, there is little change in the discrimination index for the Trillphee [Fig. 15(b), $p=0.978$] and Trill [Fig. 15(c) $p=0.872$], while it decreases significantly for the Phee [Fig. 15(a), $p<0.05$] and increases only slightly for the Twitter [Fig. 15(d), $p=0.435$] when animals became older.

5. Twin-pair comparison

The average slopes of parameter change over time for each set of twins were compared. An analysis of covariance (ANCOVA) for all parameters for each call type was performed and twins were considered significantly different for a particular parameter if $p<0.05$. In Table II, we show only significant parameters and an 18-parameter frequency analysis, which takes into account all 18 spectral parameters common across the four major call types. For the Phee call, parameters showing significant difference included duration, the time of maximum amplitude, the ending frequency, and the 18-parameter frequency distance measure. For the Trill-

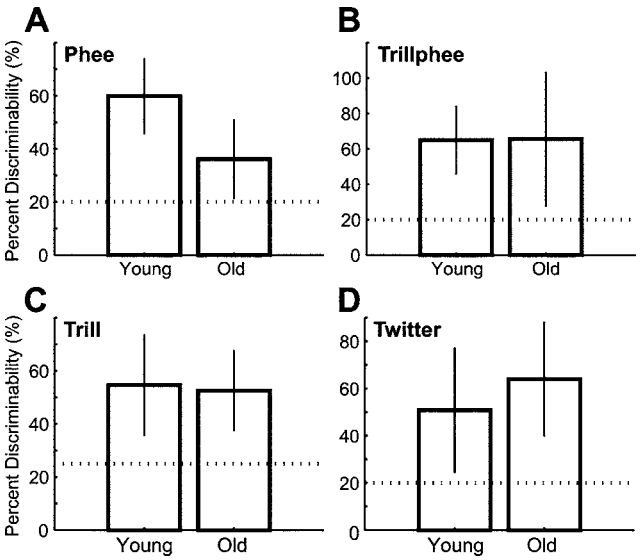


FIG. 15. Discriminability analysis based on the first two principal components (PC1 and PC2) to discriminate between individual marmosets in "young" (3–6 weeks) or "old" (19–25 weeks) groups, respectively, for the four major call types: (A) Phee; (B) Trillphee; (C) Trill; (D) Twitter. The dotted line on each plot indicates the percent correct attributed to chance (equal to one over the number of animals used in each analysis).

phee, only duration was significantly different. For the Trill, the starting frequency and frequency modulation depth differed significantly among twin pairs, and no significant difference was found in selected parameters for the Twitter.

6. Gender comparison

All collected vocalizations of the four major call types from male and female subjects were grouped accordingly and compared at two time points, at 5-weeks and at an old age. For the Phee and Trillphee, we were able to perform the analysis at 25 weeks and 21 weeks, respectively. Due to drastically decreased vocal output by females at older ages, gender differences could only be analyzed with sufficient sample size up to 13 weeks for the Trill and 19 weeks for the Twitter. Figure 16 shows some example results of this analysis. Example Phee parameters are shown from 5 and 25 weeks, Trillphee from 5 and 21 weeks, Trill from 5 and 13 weeks, and Twitter from 5 and 19 weeks to illustrate the shift in distribution over age, as well as the presence or absence of separation by gender.

TABLE II. ANCOVA data sample for twin pairs. Statistically significant and 18-frequency parameters (center column) are shown for each of the four call types (left column) with results from ANCOVA tests. p-values are in the right column and numbers in bold indicate when twin pairs were found to be significantly different ($p<0.05$).

Call type	Parameter	p
Phee	Duration	**0.0004**
	$\text{Time}_{\text{AmpHi}}$	**0.00001**
	$\text{Frequency}_{\text{End}}$	**0.0078**
	18-Parameter Frequency Distance	**0.0001**
Trillphee	Duration	**0.0394**
	18-Parameter Frequency Distance	0.1314
Trill	$\text{Frequency}_{\text{Start}}$	**0.0323**
	$\Delta\text{Freq}_{\text{Mean}}$	**0.0117**
	18-Parameter Frequency Distance	0.0759
Twitter	18-Parameter Frequency Distance	0.0932

Unexpectedly, for the Trillphee [Fig. 16(b)], the mean duration increased in males, while it decreased in females. Another surprising result was the change in the mean of the middle mean frequency, which rose for females while dropping in males. For the Trill [Fig. 16(c)], the mean modulation depth decreased significantly, and the difference between means for males and females increased. Figure 16(d) shows age-related trends among males and females for the Twitter. In general, frequency parameters do not show large changes over time but do show differences between males and females.

In a gender-based ANCOVA analysis (Table III) of the magnitude and direction of change for select parameters for each call type, males and females show little difference in the Phee calls. Some significant differences found were in the absolute maximum frequency in the Trill, the start frequency in the Trill and Trillphee, and the duration and number of phrases in the Twitter. We also computed the 18 frequency-parameter distance measure between males and females for each call type (Table III). The only significant difference in this measure was found for the Twitter call.

IV. DISCUSSION

A. Summary of findings

The present study showed that infant marmosets exhibited a high rate of calling, the use of many call types in the absence of context, as well as possessing infant-specific call types and features such as the Cry and Twitter-hook. Based on our detailed acoustic analysis, the vocalizations of infant marmosets differed from those of adults in both the spectral and temporal domains. In addition, these features changed gradually through development in the direction of the feature distributions of the adult calls. One could argue that a gradual change in vocalizations is more likely to (but not necessarily) result from experience-based plasticity or learning. Had we observed no changes or an abrupt change shortly after birth, such a possibility is much lower. One would have to argue for a genetic or innate explanation in that case.

Because individual recognition based on vocalizations has been shown across many species of animal, we compared changes in our defined parameters for each call type across individuals, twin-pairs, and genders to determine which parameter(s) may be used by marmosets to identify conspecifics as being from the same family group, of a particular gender, or as individuals. Certain features in each call type developed differently depending on the gender and family of origin. When taken as a sum of differences in one animal, based on gender and family group, it is possible to see how individual characteristics could be developing as well. Additionally, the results of the ANCOVA analysis of twin pairs and gender suggest that many of the measured parameters for each call type develop similarly across all infant marmosets. While the starting and ending points of acoustic parameters we measured may be different for *individual* infants, trends

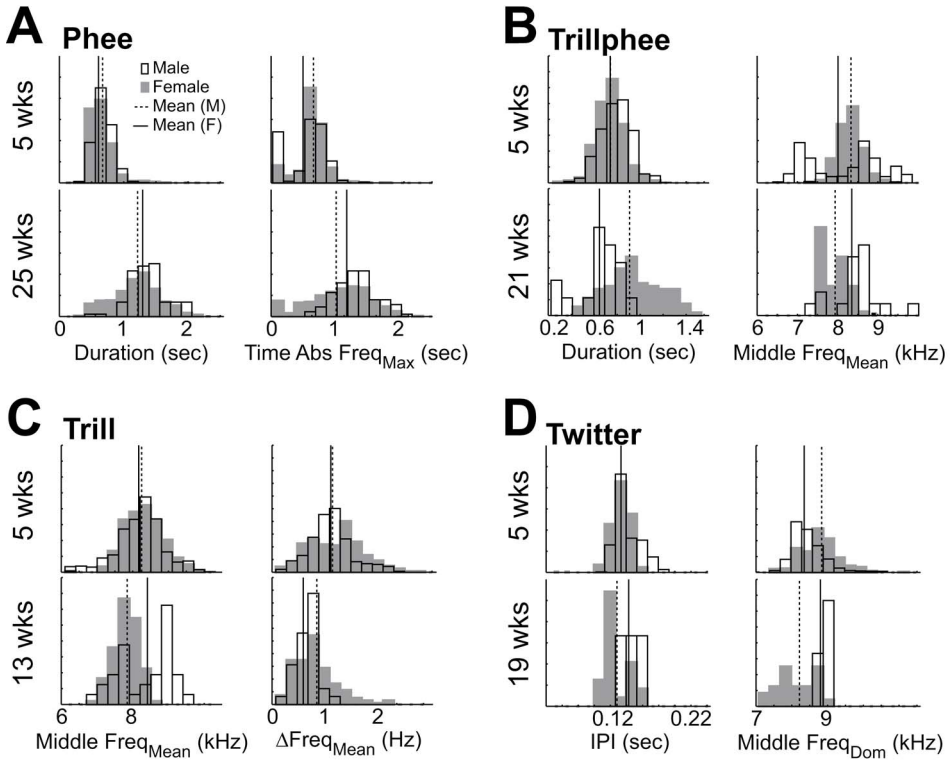


FIG. 16. Gender comparison for two example parameters from the four major call types: (A) Phee; (B) Trillphee; (C) Trill; (D) Twitter, at 5 weeks and an older age (25 weeks for Phee, 21 weeks for Trillphee, 13 weeks for Trill, and 19 weeks for Twitter) is plotted in histogram format. Parameter value is on the *x* axis. Filled grey bars are females and open bars are males. The mean of each distribution is plotted as a vertical solid line for females and a vertical dashed line for males.

in infant marmoset vocalization parameters follow the same approximate slope during development.

B. Limitations of the present study

Definitions of parameters for each call type were based on the acoustic characterization and statistics of adult marmoset calls. Because it is difficult to determine *a priori* which parameters are involved in the neural coding or behavior of vocalization perception and production in common marmosets, we analyzed as many as possible and determined which were contributing most to variability between genders, twin pairs, and individuals. Additionally, there may be parametric characteristics in infant calls associated with a certain behavior, but are less meaningful in adult vocalizations. This study focused on the development of infant marmoset vocalizations and quantified the progress of each infant in fitting adult characteristics, and presumably in developing individuality. For determining the roles of infant-specific behaviors in the social dynamic groups, a different type of study is warranted.

TABLE III. ANCOVA data sample for gender. Same format as Table II.

Call type	Parameter	*p*
Phee	18-Frequency Parameter Distance	0.3752
Trillphee	Frequency$_{Start}$	**0.0316**
	18-Frequency Parameter Distance	0.6547
Trill	Absolute Freq$_{Max}$	**0.0051**
	Frequency$_{Start}$	**0.003**
	18-Frequency Parameter Distance	0.5047
Twitter	Duration	**0.0026**
	Number Phrases	**0.0023**
	18-Frequency Parameter Distance	**0.0438**

In this study, no attempt was made to classify calls within one call type into subtypes. Previous studies of primate species have shown that there is a continuum along vocalization acoustics, making calls less stereotyped as one expands to finer detail in parameters. For example, marmosets use calls fitting the characteristics of a Phee in different contexts and vary the parameters depending on behavioral context (e.g., multipart Phees and low-amplitude Phees). For the purpose of the present study, which emphasized acoustic analysis as the first step towards a comprehensive description of vocal development in this species, we did not analyze behavioral context. This allowed for a large sample size of vocalizations to be collected and analyzed in each of the four major call types studied, and their statistics to be obtained. There are likely to be subtle acoustic differences within call types depending on the behavioral state of the animal during vocal production. Future studies incorporating proper behavioral analyses would be required to clarify these differences.

Although the recording setup used in the present study was vital to obtain high quality samples for detailed acoustic analyses, the physical isolation of infants during recording sessions constitutes an interruption of their normal social environment, even though they maintained visual, auditory, and olfactory contact with their family group. We suspect that such interruptions may have affected the contextual informa-

tion of produced calls more so than their acoustic structures, which were the focus of the analyses in this paper. Presumably, vocal behavior of a normal adult marmoset is different in isolation than it is in a social setting such as group housing, and vocal output would skew towards calls typically used in that context. This is apparent in the older animals in this study, when there is a substantial drop in the overall vocal output and the Phee dominates the recorded samples. This caveat is also an advantage; it is very telling about contextual awareness in infant marmosets. Whether spontaneous vocalization in infants is for the purpose of practice, as seen in birds or human infants, remains to be proven, but there is a clear time point when infant marmosets decrease the number of spontaneous vocalizations and it coincides with an increasing number of vocal exchanges with conspecifics.

C. Can we establish a non-human primate model of vocal production learning?

The nonhuman primate is, perhaps, the most desirable model for human speech and language development. It would be an incredible advancement of the field if it could be proven that they display similar capacities, because of the extremely close ties both genetically and anatomically. While the specific question of vocal learning was not addressed in this study, our results will reopen the debate of whether nonhuman primates exhibit ontogenetic development in the production of their vocalizations.

Recently, Egnor and Hauser (2004) addressed an apparent paradox in recent studies showing auditory feedback-dependent plasticity of adult primate vocalizations, despite older studies showing no plasticity in the development of vocal production in juvenile nonhuman primates, and they appropriately called for a more careful investigation of these contradictory findings. This study aims to lay groundwork for future studies into the vocal development of common marmosets, as well as other nonhuman primate species. We show that with strong acoustic recording and analysis, there are discrete changes in the vocalizations produced by developing marmosets.

An acoustic analysis such as the one presented here is only the first step in a study of whether vocal learning occurs, satisfying the primary requirement that significant ontogenetic changes do, in fact, occur. Determining what changes are due to maturation of the vocal tract and general growth requires experiments that perturb normal development. Behavioral and neural studies should be conducted in conjunction with detailed acoustic analysis techniques, such as those presented here, in animals with abnormal auditory feedback, such as those deafened early in life or congenitally, ear plugged during developmental stages, noise-raised, or with implanted earphone filters which could filter certain frequencies or shift feedback in pitch and/or time. Other methods of perturbation could include normally-hearing infants born to mute or vocally abnormal parents and raised isolated from normal conspecifics, infants with an altered vocal apparatus, or infants cross-fostered with a compatible species that has very different vocalizations. Behavioral studies, acoustic analysis, and other technology such as electrophysiology, functional imaging, and immediate early gene induction could contribute to answering the final question of whether there is any auditory feedback-dependent plasticity in vocal production development of nonhuman primates.

ACKNOWLEDGMENTS

We gratefully acknowledge Dr. Cory Miller for his comments and suggestions on this manuscript. This work was supported by NIH Grant No. DC005808 (X.W.).

Agamaite, J. A. (1997). "A quantitative characterization of the vocal repertoire of the common marmoset," Master's thesis, Johns Hopkins University, Baltimore, MD.

Biben, M. and Bernhards, D. (**1995**). "Vocal ontogeny of the Squirrel Monkey, Saimiri boliviensis peruviensis," in *Current Topics in Primate Vocal Communication*, edited by Zimmerman, E., Newman, J. D., and Jurgens, U., (Plenum, New York), pp. 99–120.

Egnor, S. E. R. and Hauser, M. D. (**2004**). "A paradox in the evolution of primate vocal learning," Trends Neurosci. **27**, 649–654.

Epple, G. (**1968**). "Comparative studies on vocalization in marmoset monkeys (Hapalidae)," Folia Primatol. **8**, 1–40.

Fischer, J., Cheney, D. L., and Seyfarth, R. M. (**2000**). "Development of infant baboons' responses to graded bark variants," Proc. R. Soc. London, Ser. B **267**, 2317–2321.

Fischer, J. (**2003**). "Developmental modification in the vocal behavior of non-human primates," in *Primate Audition: Ethology and Neurobiology*, edited by A. A. Ghazanfar (CRC Press, Boca Raton), pp. 109–125.

Hammerschmidt, K., Newman, J. D., Champoux, M., and Suomi, S. J. (**2000**). "Changes in rhesus macaque 'coo' vocalizations during early development," Ethology **106**, 873–886.

Hammerschmidt, K., Freudenstein, T., and Jürgens, U. (**2001**). "Vocal development in squirrel monkeys," Behaviour **138**, 1179–1204.

Inoue, M. (**1988**). "Age gradations in vocalization and body weight in Japanese monkeys (Macaca fuscata)," Folia Primatol. **51**, 76–86.

Kalin, N. H. and Shelton, S. E. (**1998**). "Ontogeny and stability of separation and threat-induced defensive behaviors in rhesus monkeys during the first year of life," Am. J. Primatol **44**, 125–135.

Masataka, N. (**1993**). "Effects of contingent and noncontingent maternal stimulation on the vocal behaviour of three—to four-month-old Japanese infants," J. Child Lang **20**, 303–312.

McCowan, B., Franceschini, N. V., and Vicino, G. A. (**2001**). "Age differences and developmental trends in alarm peep responses by squirrel monkeys (Saimiri sciureus)," Am. J. Primatol **53**, 19–31.

Merker, B. and Cox, C. (**1999**). "Development of the female great call in Hylobates gabriellae: A case study," Folia Primatol. **70**, 97–106.

Miller, C. T. and Wang, X. (**2006**). "Sensory-motor interactions modulate a primate vocal behavior: Antiphonal calling in common marmosets," J. Comp. Physiol., A **192**, 27–38.

Moody, M. I. and Menzel, E. W., Jr. (**1976**). "Vocalizations and their behavioral contexts in the tamarin Saguinus fuscicollis," Folia Primatol. **25**, 73–94.

Newman, J. D. and Symmes, D. (**1974**). "Vocal pathology in socially deprived monkeys," Dev. Psychobiol. **7**, 351–358.

Newman, J. D. (**1995**). "Vocal Ontogeny in Macaques and Marmosets: Convergent and Divergent Lines of Development," in *Current Topics in Primate Vocal Communication*, edited by E. Zimmerman, J. D. Newman, and U. Jurgens (Plenum, New York), pp. 73–97.

Norcross, J. L. and Newman, J. D. (**1993**). "Context and gender-specific differences in the acoustic structure of common Marmoset (Callithrix jacchus) Phee calls," Am. J. Primatol **30**, 37–54.

Norcross, J. L. and Newman, J. D. (**1997**), "Social context affects phee call production by nonreproductive common marmosets (Callithrix jacchus)," Am. J. Primatol **43**, 135–145.

Owren, M. J., Dieter, J. A., Seyfarth, R. M., and Cheney, D. L. (**1992**). "Food' calls produced by adult female rhesus (Macaca mulatta) and Japanese (M. fuscata) macaques, their normally-raised offspring, and offspring cross-fostered between species," Behaviour **120**, 218–231.

Owren, M. J., Dieter, J. A., Seyfarth, R. M., and Cheney, D. L. (**1993**). "Vocalizations of rhesus (Macaca mulatta) and Japanese (M. fuscata) macaques cross-fostered between species show evidence of only limited modification," Dev. Psychobiol. **26**, 389–406.

Schrader, L. and Todt, D. (**1993**). "Contact call parameters covary with social context in common marmosets, Callithrix j. jacchus," Anim. Behav. **46**, 1026–1028.
Seyfarth, R. M. and Cheney, D. L. (**1986**). "Vocal development in vervet monkeys," Anim. Behav. **34**, 1640–1658.
Seyfarth, R. M. and Cheney, D. L. (**1997**). "Some general features of vocal development in nonhuman primates," in *Social Influences on Vocal Development*, edited by C. T. Snowdon and M. Hausberger (Cambridge University Press, U.K.), pp. 249–273.
Snowdon, C. T. (**1997**). "Affiliative processes and vocal development," Ann. N.Y. Acad. Sci. **807**, 340–351.
Snowdon, C. T., Elowson, A. M., and Roush, R. S. (**1997**). "Social influences on vocal development in New World Primates," in *Social Influences on Vocal Development*, edited by C. T. Snowdon and M. Hausberger (Cambridge University Press, U.K.), pp. 234–248.
Snowdon, C. T. and Elowson, A. M. (**2001**). "'Babbling' in pygmy marmosets: Development after infancy," Behaviour **138**, 1235–1248.
Stevenson, M. F. and Poole, T. B. (**1976**). "An ethogram of the common Marmoset (Callithrix jacchus jacchus): General behavioural repertoire," Anim. Behav. **24**, 428–451.
Winter, P. (**1969**). "The variability of peep and twit calls in captive squirrel monkeys (Saimiri sciureus)," Folia Primatol. **10**, 204–215.
Winter, P., Handley, P., Ploog, D., and Schott, D. (**1973**). "Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation," Behaviour **47**, 230–239.

Estimated communication range of social sounds used by bottlenose dolphins (*Tursiops truncatus*)

Ester Quintana-Rizzo and David A. Mann
College of Marine Science, University of South Florida, 140 7th Avenue South, St. Petersburg, Florida 33701

Randall S. Wells
Chicago Zoological Society, c/o Mote Marine Laboratory, 1600 Ken Thompson Parkway, Sarasota, Florida 34236

(Received 19 December 2005; revised 9 June 2006; accepted 21 June 2006)

Bottlenose dolphins, *Tursiops truncatus*, exhibit flexible associations in which the compositions of groups change frequently. We investigated the potential distances over which female dolphins and their dependent calves could remain in acoustic contact. We quantified the propagation of sounds in the frequency range of typical dolphin whistles in shallow water areas and channels of Sarasota Bay, Florida. Our results indicated that detection range was noise limited as opposed to being limited by hearing sensitivity. Sounds were attenuated to a greater extent in areas with seagrass than any other habitat. Estimates of active space of whistles showed that in seagrass shallow water areas, low-frequency whistles (7–13 kHz) with a 165 dB source level could be heard by dolphins at 487 m. In shallow areas with a mud bottom, all whistle frequency components of the same whistle could be heard by dolphins travel up to 2 km. In channels, high-frequency whistles (13–19 kHz) could be detectable potentially over a much longer distance (>20 km). Our findings indicate that the communication range of social sounds likely exceeds the mean separation distances between females and their calves. Ecological pressures might play an important role in determining the separation distances within communication range. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2226559]

PACS number(s): 43.80.Ka [WWA] Pages: 1671–1683

I. INTRODUCTION

Understanding how acoustic signals are used by animals to communicate is basic to describing how relationships are formed and maintained. This is particularly important in turbid aquatic environments such as those inhabited by coastal bottlenose dolphins (*Tursiops truncatus*). Bottlenose dolphins leave and rejoin their conspecific associates frequently and acoustic communication might be used to find and locate distant conspecifics. The maximum distance that an acoustic signal can travel is likely the maximum distance over which associates can remain in contact with one another (Brenowitz, 1982; Klump 1996). However, the hearing capabilities of a species must be taken into account along with environmental features affecting sound transmission to understand how far a signal can travel before it drops below the masked auditory threshold or noise floor limiting communication. Therefore, knowledge of the maximum propagation distance of an acoustic signal, the characteristics of ambient noise, and the hearing capabilities of bottlenose dolphins is important for understanding what constitutes a group; if individuals are within communication range they may be part of the same social unit despite being temporarily separated. The approach is valuable to understand if individuals considered as different groups based on their distance of temporary separation (Wells and Scott, 1990; Smolker *et al.*, 1993) could be part of a single group maintaining acoustic contact. This paper estimates the communication range of social sounds produced by bottlenose dolphins that are within the hearing threshold of the species.

Bottlenose dolphins use sounds known as whistles to contact conspecifics over long distances (Janik and Slater, 1998; Caldwell *et al.*, 1990). Whistles are narrow-band, frequency-modulated sounds ranging from 4 to 20 kHz (Caldwell *et al.*, 1990). Janik (2000a) revealed that wild, unrestrained dolphins located at distances up to 580 m apart could mimic each other's whistles. He proposed this as evidence that dolphins use whistles to communicate over long distances. The active space of a signaler is the distance that a signal can be detected and recognized by a receiver (Brenowitz, 1982; Klump, 1996; Janik, 2000b). In the only study on communication ranges in bottlenose dolphins, Janik (2000b) examined propagation of natural dolphin whistles in a 10 m deep channel by measuring source levels and then estimating propagation and the active space using a model. He found that the active space where dolphins could perceive unmodulated whistles between 3.5 kHz and 10 kHz was between 20 km and 25 km at sea state zero.

Sound propagation can be dramatically affected by the habitat through which sound travels (Rogers and Cox, 1988; Forrest, 1994; Tyack, 2000; D. P. Nowacek *et al.*, 2001). Large reflecting surfaces or vegetation attenuate some frequencies and amplify others (Michelsen and Larsen, 1983). In the aquatic environment, habitat features such as bottom type, bathymetry, temperature, salinity, and vegetation affect the transmission and reception of sounds (D. P. Nowacek *et*

al., 2001). The effect of vegetation is not surprising since it acts as a discontinuous barrier to the transmission of sound. Researchers have also found that in shallow waters, low frequency sounds do not propagate as far as high frequency sounds (Forrest *et al.*, 1993; Forrest, 1994). This suggests that dolphin whistles may be affected by environmental variables. Hence, as in other species, the structure of dolphin signals might represent an acoustic compromise balancing an ensemble of ecological and perceptual factors (Wiley and Richards, 1978; Brown *et al.*, 1979).

Other factors determining whether a sound is detected and identified by an individual are the animal's hearing threshold, critical ratio, and the spectrum level of background noise. In bottlenose dolphins, the lowest hearing thresholds are in the frequencies near 50 kHz (Johnson, 1967), but whistles have much lower frequencies. Johnson (1967) found that below 50 kHz the threshold increases continuously with decreasing frequency to a maximum of about 137 dB at 75 Hz. Information on the background noise levels is also necessary to estimate the active space of whistles since high-noise levels can significantly mask a sound. The critical ratio is defined as the difference between the level of a just-detectable tone and the spectrum level background noise spanning the same frequency (Johnson, 1968; Janik, 2000b). Like the hearing threshold, critical ratios are also frequency-dependent and they have been calculated for frequencies within the whistle range (Johnson, 1968).

We conducted a series of sound transmission experiments to quantify the propagation of sounds in shallow water areas and channels in Sarasota Bay, FL. This habitat is quite different from the Moray Firth studied by Janik (2000b), in that it is very shallow and many areas contain seagrass. In contrast, the Moray Firth is an unusual habitat for coastal dolphins because the inner waters have depths of up to about 50 m. The outer waters resemble the open sea more with the deepest areas being up to 235 m (Wilson, 1995). The Moray Firth is the northern extreme of the species range. The shallow water of Sarasota Bay provides an excellent opportunity to estimate the active space for typical coastal bottlenose dolphins. We examined the effects of habitat characteristics such as depth, bottom type, vegetation, and bottom sediment on sound propagation. We used regression models to estimate maximum distance of detection taking into account the hearing capabilities of bottlenose dolphins, the background noise levels, and the critical ratios for masking sounds. We also examined the active space of different types of whistles in the same habitats where experiments were conducted. This allowed us to compare estimates of maximum communication range with the distances of separation observed during observations of wild dolphins.

II. MATERIALS AND METHODS

This study consisted of three basic components: (1) behavioral observations of mother/calf pairs to identify the habitats that they used and where they temporarily separated, (2) sound propagation experiments at the locations where mothers and calves temporarily separated, and (3) modeling of sound propagation data and information on hearing sensitivity, background noise levels, and critical ratios to estimate the active space of whistles.

A. Behavioral observations

Behavioral observations of 7 resident female dolphins and their dependent calves (3 years old) were conducted from June to September 2003 from a 7-m-long boat equipped with a 115 hp 4-stroke engine. We recorded acoustic and behavioral data continuously using focal animal observations (Altmann, 1974). Focal animal observations were conducted on the mother of mother-calf pairs. The research vessel was kept at a distance of approximately 20 m from the mother. Separation distances between mother and calf were estimated each time the mother surfaced and the observation zone included an area of approximately 200 m from the mother. Behavioral observations allowed identification of the habitats where sound transmission experiments were to be conducted and acoustic data allowed us to determine the frequency range of whistles used in different habitats to quantify their active space.

The set up of the acoustic recording was as follows. At the bow of the observation boat, two 1.5 m sections of PVC pipe were joined in a T joint and secured across the gunwales (Sayigh *et al.*, 1993). On each side of the boat approximately 2 m of hydrophone cable were extended from the end of the pipe into the water; each calibrated HTI-96-MIN hydrophone was approximately 1 m below the surface when the boat was not moving. We used two hydrophones to ensure that we would have a backup recording. To prevent the hydrophone from bouncing at the surface while the boat was moving, the cable of the hydrophone was weighted with a chain attached to the end of the PVC pipe by a carabiner. Each hydrophone was connected to a 2-kHz high-pass filter to reduce engine noise.

Signals from each hydrophone were digitized at 48.8 kHz with a Tucker-Davis Technologies RP2 module, and stored to a computer hard drive. Signals from each channel were sampled at precisely the same time on each channel. Data were recorded with a 24-bit A-to-D converter and were stored as 32-bit floating point values. Data were analyzed with MATLAB® 6.5.

Behavioral observations allowed us to identify the areas in which to conduct sound transmission experiments. Such areas were chosen based on the fact that (1) dolphins were observed there temporarily separated and thus the active space of whistles used during separations could be examined and (2) the general areas have been identified as areas of high use by dolphins during the long-term studies of the Sarasota Dolphin Research Program (Wells, 2003). Specific locations of temporary separations were recorded as latitude and longitude from a Garmin GPS 12 Personal Navigator.

B. Sound transmission experiments

Nine sound transmission experiments were conducted in Sarasota Bay, Florida, from September to October 2003 (Fig. 1; Table I): five in shallow water areas (<3 m) and four in channels (>3 m, up to 5.3 m). Each experiment was con-

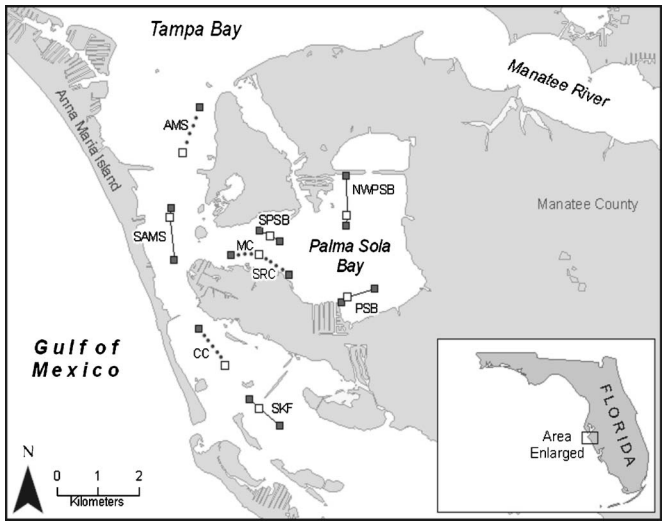


FIG. 1. Study area showing the experimental transects in Sarasota Bay, Florida. White circles represent the location of the hydrophone and black circles represent the last location in which the speaker was placed in each transect. Shallow areas are represented by solid lines: Palma Sola Bay (PSB), North West of Palma Sola Bay (NWPSB), South West of Palma Sola Bay (SWPSB), Sister Key flats (SKF), and South East of Anna Maria Sound (SAMS). Channels are represented by dashed lines: Main Channel (MC), San Remo Channel (SRC), Anna Maria Sound (AMS), and Cortez Channel (CC).

ducted in an area where dolphins were observed to engage in temporary separations. In each experiment, we used computer-generated tones that spanned the same frequencies as dolphin whistles (5, 7, 9, 11, 13, 15, 17, and 19 kHz). Tones were played simultaneously for 10 s and then repeated with a period of silence of 0.03 s separating them. In each transect, the tone-silence loop was broadcast for $\geqslant$1 min from an underwater transducer (source) located in a 7-m-long boat. Sounds were played from a laptop computer through a power amplifier (Hafler P1000) connected to an underwater speaker (Aqua Synthesis). The computer was connected to a RP2-system (Tucker-Davis Technologies) with a HTI-96-MIN hydrophone (sensitivity -169.8 dBV/μPa; 1–24 kHz).

The source and the receiving hydrophone (receiver) were located 1 m below the water surface. This depth was chosen because it was the depth of the hydrophones used to record dolphin whistles during animal observations. Propagated signals and environmental noise were recorded on a NOMAD Jukebox 3 (Creative Labs, Inc.) kept in a stationary kayak. The kayak and the boat were kept in placed during the experiments by using two anchors in each vessel.

The source was moved at a constant heading from the receiver (transect line) to simulate the movement of one dolphin relative to another. The start point of a given transect was the position of the mother at the time of maximum separation from the calf. The position was recorded as a geographical coordinate during behavioral observations. When possible two transect lines were done for every experiment. In such cases, the direction of each transect was different so that each one followed the general direction of the movement of dolphins before and after they were temporarily separated. For example, if dolphins traveled from point A to B to C with

TABLE I. Overview of transects per habitat where sound transmission experiments were conducted in Sarasota Bay, Florida.

Habitat	Site	No. of Transects	Start Latitude/ Longitude	Bottom type	Mean depth±SD
Shallow water	SE Palma Sola Bay (PSB)	2	N 27°28′14″ W 82°39′24″	Seagrass	PSB1: 1.6±0.1 m (n=9) PSB2: 1.6±0.3 m (n=18)
	NW Sister Key (SKF)	1	N 27°26′51″ W 82°40′34″	Sand	SKF: 2.0±0.2 m (n=17)
	NW Palma Sola Bay (NWPSB)	2	N 27°29′14″ W 82°39′26″	Sandy mud	NWPSB1: 2.1±0.5 m (n=12) NWPSB2: 2.6±0.2 m (n=19)
	SW Palma Sola Bay (SPBS)	2	N 27°28′59″ W 82°40′26″	SPSB1: mud seagrass SPSB2: mud	SPSB1: 1.7±0.2 m (n=10) SPSB2: 1.4±0.3 m (n=11)
	SE Anna Maria Sound (SAMS)	2	N 27°29′13″ W 82°41′45″	Mud	AMS1: 2.0±0.3 m (n=18) AMS2: 1.3±0.1 m (n=9)
Channel	Main Channel (MC)	1	N 27°29′57″ W 82°40′31″	N/A	3.5±1.0 m (n=19)
	San Remo Channel (SRC)	1	N 27°28′45″ W 82°40′31″	N/A	4.1±0.7 m (n=14)
	Anna Maria Sound (AMS)	1	N 27°30′09″ W 82°41′34″	N/A	3.1±0.4 m (n=22)
	Cortez Channel (CC)	1	N 27°27′23″ W 82°41′01″	N/A	3.4±0.3 m (n=16)

B being the point of maximum distance of separation, one transect was done from point B to A, and another transect was done from point B to C. The transect size varied based on water depth and size of the sampled area. When possible, the source was located at the following experimental distances from the receiver: 1 m, 5 m, 10 m, 20 m, 30 m, 40 m, 50 m, 75 m, 100 m, 150 m, 200 m, 250 m, 300 m, 350 m, 400 m, 450 m, 500 m, 600 m, 700 m, 800 m, 900 m, 1000 m, and 1100 m. The experimental distances exceeded the maximum separation distances because we were interested in determining sound propagation beyond that range. Experimental distances up to 800 m were measured with a Leica LRF 800 laser range finder. Distances greater than 800 m were measured with a Garmin GPS 12 Personal Navigator (accuracy: 15 m RMS). At each experimental distance, water depth was recorded and in shallow water transects, vegetation type (presence or absence of seagrass), and sediment type (sand or mud) were also noted. A transect was defined as unvegetated if no seagrass was found in more than 75% of the sampling locations. Surface sediment samples were grabbed at each distance and they were classified as sand (granular matter of a few millimeters in size), mud (semiliquid mixture of water and soil), and sandy-mud. The sandy-mud sample was taken to the Geological Oceanography Program of the College of Marine Science at the University of Florida for its classification. All experiments were conducted in sea state zero. Channel width at the narrowest and widest points was measured using digitized bathymetry data from ESRI®ArcGIS™9.0. Transmission experiments were conducted around 2 h before, and 2 h after, the same tidal state recorded during the behavioral observations of the maximum separation event.

For every experiment, the spectrum level sound of tones at each distance was calculated using a 48000-point FFT with a Hanning window, which resulted in a frequency resolution of 1 Hz. In the calculation, we corrected for the analysis bandwidth, the hydrophone sensitivity (−169.8 dB re: 1 V/μPa), the calibration of the NOMAD Jukebox recorder (18.9 dB re: 1 V), and the Hanning window (6 dB).

Noise level was measured using the recorded signal where no sound was being broadcast up to 7.5 ms before each tonal frequency. The signal-to-noise ratio (SNR) of each frequency was calculated by subtracting the noise level from the corresponding received level.

C. Modeling of active space of whistles

The transmission loss data were used to calculate regressions that model sound propagation and the ability of a dolphin to detect that signal in each experimental transect. Regression equations were calculated for each channel transect and for one of the two transects of each shallow water experiment for a total of nine models of sound propagation. A logarithmic curve was fitted to the received levels of sound propagation and the resulting equation was used to estimate the active space of hypothetical whistles taking into account the spectrum level background noise, dolphin critical ratios (Johnson, 1968), and dolphin hearing thresholds (Ljungblad *et al.*, 1982). This is because a sound can be heard by a dolphin only if its received level is above the spectrum level background noise and the animal's critical ratio. Thus, sound detection is limited by the combined effects of the dolphin hearing threshold and the spectrum level background noise plus critical ratio. Since different frequencies propagate different distances, the propagating distance of the first whistle frequency reaching the threshold was defined as the frequency limiting the active space of whistles. For example, if the attenuation of 5 kHz, 9 kHz, and 11 kHz frequencies of a 5–11 kHz whistle are examined, it is found that the 9 kHz signal reaches the noise floor plus critical ratio before the 5 and 11 kHz frequencies. In this example, the hearing range is noise limited.

The results of the regression models were used to calculate the active space of two hypothetical whistles. The first whistle had frequencies from 7 to 13 kHz, which corresponded to the mean minimum and mean maximum frequencies of whistles recorded during separations of dolphins in shallow water areas. We referred to this whistle as a low-frequency whistle. The second whistle had frequencies from 13 to 19 kHz and we referred to it as a high-frequency whistle. The minimum frequency of the high-frequency whistle corresponded to the closest frequency of the tone played to the mean minimum frequency (12 kHz) of whistles recorded during separations in channels. Similarly the maximum frequency of the high-frequency whistle corresponded to the closest frequency of the tones played to the mean maximum frequency (20 kHz) of whistles recorded during separations in channels.

The hearing thresholds of the frequencies played were calculated from the hearing thresholds reported by Ljungblad *et al.* (1982). We fitted a regression line to the hearing thresholds of frequencies 5 kHz, 10 kHz, 15 kHz, and 20 kHz to calculate the hearing threshold of the tone frequencies not included in their study: 7 kHz, 9 kHz, 11 kHz, 17 kHz, and 19 kHz. Since Ljungblad *et al.* (1982) calculated lower-frequency hearing thresholds using two projector systems, we used the calculated regression values from their project LC-10 projector data because they had the highest coefficient (R^2=0.95). Hearing thresholds calculated were 81.5 dB SPL for 5 kHz, 79.8 dB SPL for 7 kHz, 78.4 dB SPL for 9 kHz, 77.5 dB SPL for 11 kHz, 76.7 dB SPL for 13 kHz, 75.9 dB SPL for 15 kHz, 75.3 dB SPL for 17 kHz, and 74.8 dB SPL for 19 kHz.

The critical ratio of each frequency was added to the average background noise of that frequency. The critical ratios were calculated by fitting a regression line to the critical ratios of frequencies that Johnson (1968) did not examine but that were included in our experiments. These critical ratios corresponded to 23.6 dB for 5 kHz, 25.2 dB for 7 kHz, 26.8 dB for 9 kHz, 28.4 dB for 11 kHz, 29.9 dB for 13 kHz, 31.5 dB for 15 kHz, 33.2 dB for 17 kHz, and 34.8 dB for 19 kHz.

The development of a model of whistle propagation required data on the source levels of whistles produced under natural circumstances. No information exists on the source levels of the whistles produced by the Sarasota dolphins. Thus, we constructed a table of possible source levels of

whistles (155 dB, 160 dB, and 165 dB) using as a reference the maximum source level reported for other wild bottlenose dolphins (169 dB re 1 μPa; Janik, 2000b).

We compared the means of the slopes of the regression models using a two way analysis of variance (two-tailed, alpha=0.05) to test for significant differences between habitats, frequencies, and interactions between frequencies habitats. The slope of the regression indicates how sound levels fall off with distance. Statistical analysis was performed with SPPS v. 14.0 (SPSS, Chicago, IL).

III. RESULTS

A. Behavioral observations

A total of 224 separations of females and their dependent calves were observed. Of these, 161 occurred in shallow water and 63 occurred in channels. Mean separation distance in shallow water was 115±48 m and in channels was 99±48 m. Eight separation events were used to identify the areas where sound transmission experiments were conducted. They corresponded to the most recent events and they occurred on different dates. Five separations of three different mother/calf pairs were recorded in shallow water and their separation distances were 90 m, 95 m, 100 m, 120 m, and 200+ m. In channels, three separations events of three different mother/calf pairs were recorded. One separation event occurred in the intersection between two channels and this point was thus used as the start point of each experimental transect in the two channels. Separation distances in the channels were 50 m, 200 m (n=2), and 200+ m.

During the separations, we recorded a total of 204 whistles. In the shallow water areas, 199 whistles were recorded in three of the five separations and they corresponded to three different mother/calf pairs. The three separations lasted a total of 38.0 min. In channels, 5 whistles were recorded in two of the four separations and they corresponded to two different mother/calf pairs. The two separations lasted 17.1 min. In shallow water areas, whistles had a minimum frequency with a mean equal to 7.5±2.5 kHz and a maximum frequency with a mean equal to 13±3.2 kHz. In channels, whistles had a minimum frequency with a mean equal to 12±3.6 kHz and a maximum frequency with a mean equal to 20±7.4 kHz.

B. Sound transmission experiments

Shallow water transects had mean depths varying from 1.3 m (SAMS2) to 2.6 m (NWPSB2; Table I). The overall mean depth of three transects was 1.9±0.5 m. Channel transects had mean depths varying from 3.1 m (North Anna Maria Sound) to 4.1 m (San Remo Channel) and their overall mean depth was 3.47±0.7 m.

Most frequencies either followed the spherical spreading attenuation model or had transmission loss values that were intermediate between the predicted values of the spherical and cylindrical spreading attenuation models (Figs. 2 and 3). Low frequencies traveled much farther than high frequencies and sound propagation varied within and between channels and shallow water areas (Figs. 2 and 3). In shallow water, the attenuation was highly variable at distances up to 50 m when the average water depth was equal to 1.3 m (n=19, SD =0.1). In contrast, when the mean water depth was greater than 2 m, attenuation was highly variable at shorter distances (5 and 20 m; Fig. 2).

To compare sound propagation among shallow-water transects, we examined transmission loss at the 100-m point, which is the maximum distance of the shortest transect. At the 100-m point, three transects (SAMS2, SPSB1, and SPSB2) had a depth of 1.1 m, and mean transmission loss over all frequencies was −27.4 dB. In the other six transects (PSB1, PSB2, SAMS1, SKF, NWPSB1, and NWPSB2) water depth was greater (2.0±4 m), and mean transmission loss over all frequencies was greater than or equal to −30.2 dB over 100 m.

We examined the effect of vegetation and sediment type on sound transmission loss in shallow water areas. At the 100-m point, the mean transmission loss was greater in transects with seagrass and lower in transects with mud or sand bottom sediments (Fig. 2). In seagrass transects, mean transmission loss was approximately −36 dB for low-frequency whistles, and −47 dB for high-frequency whistles. In nonseagrass transects, mean transmission loss was similar between the two type of hypothetical whistles (low-frequency whistles −29.0±6.8 dB, high-frequency whistle −29.8±4.8 dB). Frequencies greater than 17 kHz were less attenuated in sand-areas than in the sand-mud and mud areas. Mean transmission loss was −41.8 dB in transects with seagrass (PSB1 and PSB2), −29.5 dB in transects with sandy-mud sediment (NWPSB 1 and NWPSB 2), −27.7 dB in transects with mud sediment (AMS1, AMS2, SPSB1, and SPSB2), and −25.6 dB in transects with sand sediment (SKF).

We examined transmission loss in channels at the 100 -m point also. At the 100-m point, three of the four channels had relatively similar water depths (Anna Maria Sound =3.4 m, Cortez Channel=3.4 m, and Main Channel=3.2 m), but their mean transmission losses were very different (Fig. 3). The mean transmission loss over all frequencies was −14.3 dB in the Cortez Channel, −26.7 dB in the Main Channel, and −40.1 dB in Anna Maria Sound. Mean transmission loss of low- and high-frequency whistles was, respectively, −15.3 dB and −13.5 dB in the Cortez Channel, −27.0 dB and −30.40 dB in the Main Channel, and −40.8 dB and −38.1 dB in Anna Maria Sound. The fourth channel (San Remo Channel) was deeper at the 100-m point (4.7 m) and its mean transmission loss was equal to −35.2 dB. Mean transmission loss of low- and high-frequency whistles was −35.6 dB and −36.4 dB, respectively.

Mean transmission loss at the 100-m point was more variable in channels than in shallow water areas. Mean transmission loss varied from −14.3 dB to −40.1 dB in channels and from −25.6 dB to −43.4 dB in shallow water areas. However, at the same point, the range between the minimum and maximum transmission losses was relatively similar between the two habitats with shallow water areas being wider by 4 dB (−12 dB to −57 dB) than channels (−4 dB to −45 dB).

Noise levels were variable among transects both in shallow water and channels (Table II). Noise levels plus critical

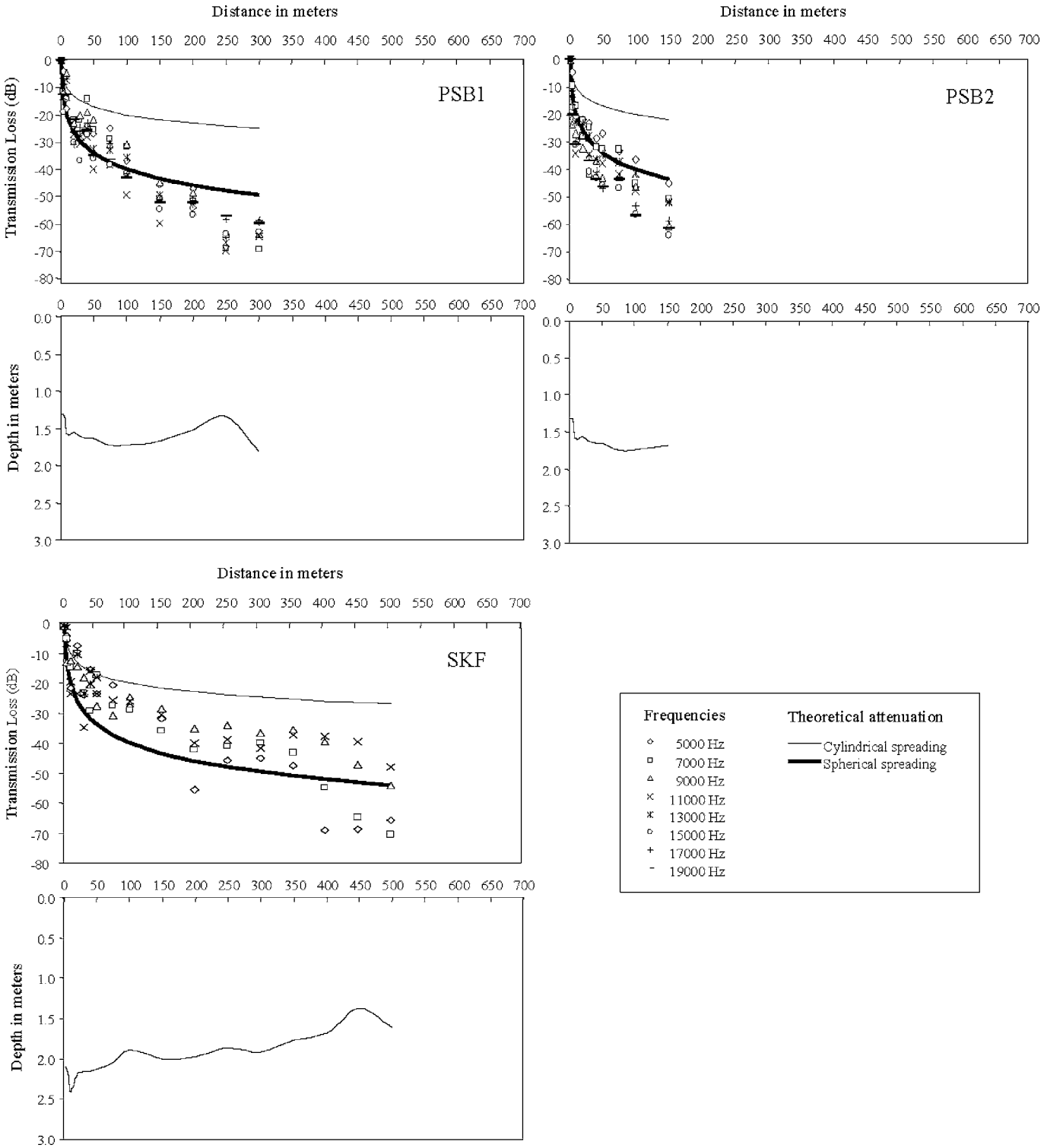


FIG. 2. Transmission loss data with distance for eight tones used during sound transmission experiments conducted in four shallow areas (PSB, SKF, NWPSB, and SPSB). Except for SKF, two transect lines were done in each experiment and each transect line is represented with a number next to the code of the corresponding area. Theoretical attenuation based on cylindrical and spherical spreading is also shown. A profile of the depth contour of each transect is also included.

ratios ranged from 94.5 dB to 106.9 dB re 1 μPa^2/Hz in shallow water areas and from 100.6 dB re 1 μPa^2/Hz to 110.0 dB re 1 μPa^2/Hz in channels. For most frequencies, the noise level plus critical ratio was lower in shallow water transects than in channels. In some shallow water locations (SKF, PSB) and all channels, the noise level plus critical ratio increased with increasing frequency, especially at frequencies greater than 11 kHz. However, in other shallow water locations (SAMS, SPSB) the noise level plus critical ratio was approximately the same among frequencies.

C. Modeling of active space

The theoretical detection range was noise-limited, as opposed to hearing-sensitivity-limited, both in shallow water

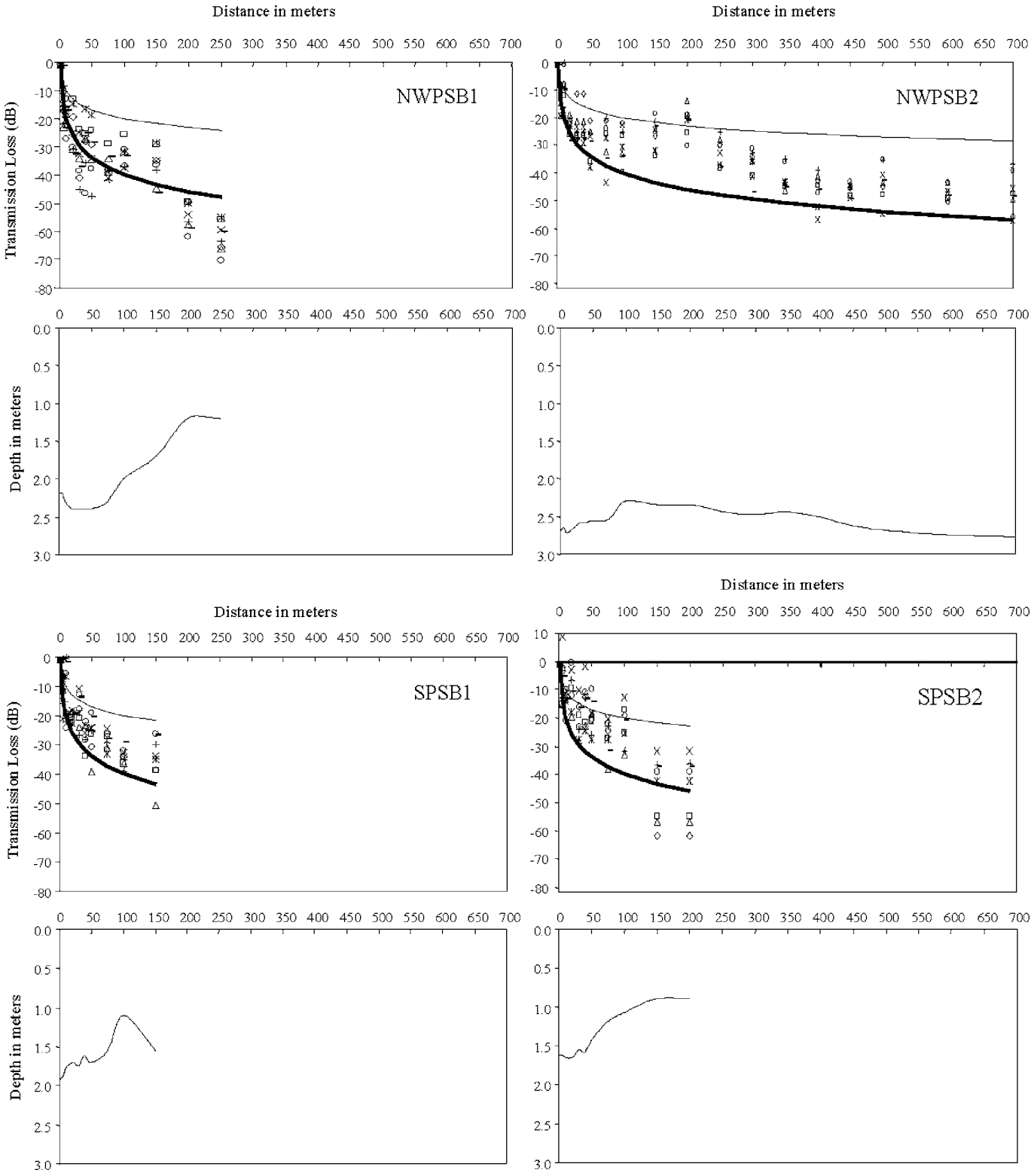


FIG. 2. (*Continued).*

areas and in channels. This was evident in the fact that all noise levels plus critical ratio measurements were greater that the hearing thresholds obtained by Ljungblad *et al.* (1982).There were no significant differences in the mean regression slopes between habitats ($F=2.41$, $p=0.13$, df=1) or between frequencies ($F=0.91$, $p=0.50$, df=7). There was a significant interaction between habitat and frequency ($F=2.42$, $p=0.03$, df=7), which was mainly due to differences between 5 kHz and 9 kHz in channel and shallow water.

Active space was shorter in shallow water areas than in channels. In shallow water, active space of whistles was greater in unvegetated habitats than in seagrass habitats (Fig. 4). For example, the active space of a low-frequency whistle with a source level equal to 155 dB was estimated to be

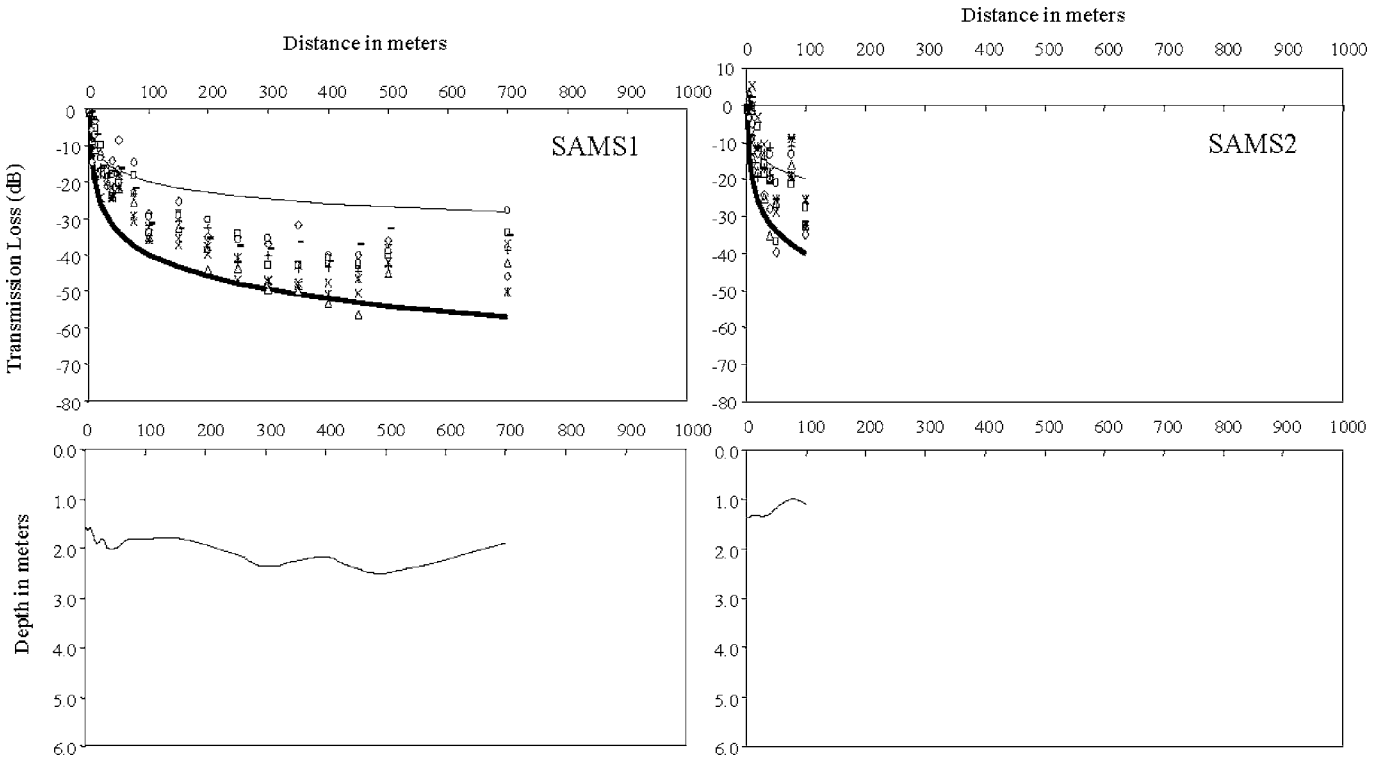


FIG. 2. (*Continued*).

approximately 662 m in an unvegetated habitat (mud bottom: SAMS) and 186 m in a habitat with seagrass (PBS). In channels, the same whistle had an estimated active space approximately between 230 m to 1 km depending on the channel (SRC=230 m, AMS=345 m, MC=750 m, CC =6070 m). Active space was also different between whistles with different frequency components. While a low-frequency whistle with a source level of 155 dB can travel up to a 6 km, a high-frequency whistle with the same source level can travel approximately up to 4 km depending on the channel.

Active space almost doubled with a 5 dB increase in whistle source level (Fig. 4). In shallow water, a low-frequency whistle with a source level equal to 160 dB had an estimated active space of approximately 1260 m in an unvegetated habitat (mud bottom: SAMS) and of 301 m in a seagrass habitat (PBS). The estimated active space of a low-frequency whistle with a source level equal to 165 dB was over 2 km in the same unvegetated habitat and close to 500 m in the same seagrass habitat. In channels, a similar pattern was observed. Low-frequencies with a 160 dB source level can travel between 400 m (SRC) and 13 km (CC). If the source level of the same whistle increased by 5 dB, the estimated active space increased to 1 km (SRC) and 28.5 km (CC). It is important to notice that the estimates assume that a habitat is homogenous in its propagation characteristics.

IV. DISCUSSION

Separation distances of females and their dependent calves were shorter than the estimated active space of whistles. Since whistles are thought to be used by dolphins to maintain group cohesion (Janik and Slater, 1998; Norris *et al.*, 1994; Smolker *et al.*, 1993), the results suggest that dolphins can communicate over the distances that temporary separations occurred. The results also suggest that separation distances are not necessarily determined by the maximum communication range. Other factors such as predation pressure or food distribution may be important. A calf may not wander far from its mother if the cost of predation risk is high. Furthermore, factors like ambient noise can affect communication range by dramatically reducing the active space (Urick, 1975; Forrest, 1994; Janik, 2000b; Slabbekoorn, 2004). In this respect, our results showed the best-case scenario of sound propagation and estimates of active space, because experiments were conducted when no boats were present within a radius of approximately 1 km of the recordings. In fact, the theoretical detection range was noise-limited, as opposed to hearing-sensitivity-limited, both in shallow water areas and in channels. However, background noise can vary widely depending on the number of power boats present, fish choruses, snapping shrimp, and wave action. In Sarasota Bay, dolphins are frequently exposed to boat noise as boats pass within 100 m of them on an average of every 6 min during daylight hours (S. M. Nowacek *et al.*, 2001).

Estimates of active space were based on the propagating distance of the first frequency reaching the noise threshold since different frequencies propagate different distances. Yet, dolphins may be able to discriminate a whistle even if one of its frequency components is lost with distance. However, it is currently unknown how dolphins discriminate and identify sounds. It is one thing to detect a sound and quite another to

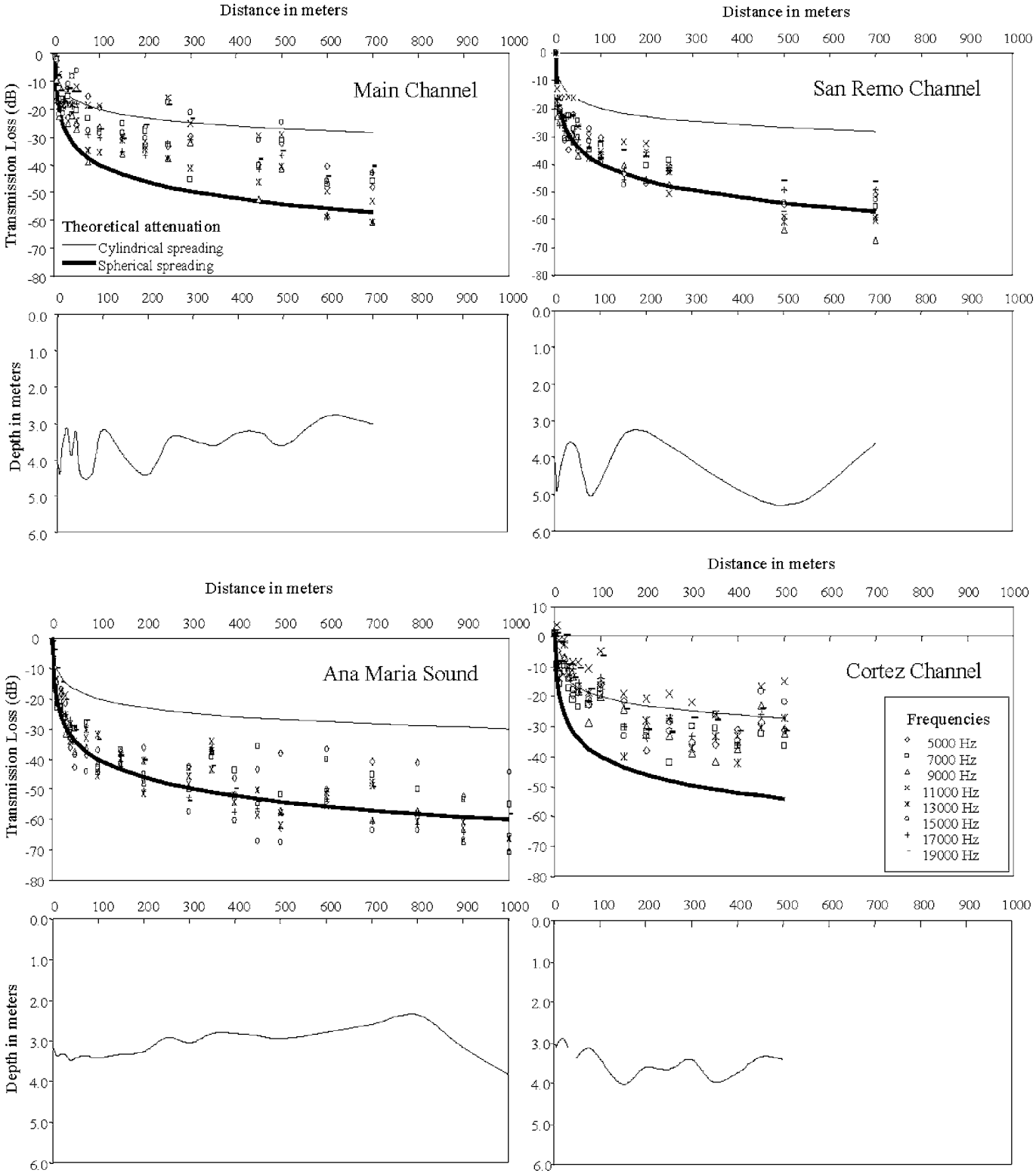


FIG. 3. Transmission loss data with distance for eight tones used during sound transmission experiments conducted in four channels (Main Channel, San Remo Channel, Anna Maria Sound, and Cortez Channel). Theoretical attenuation based on cylindrical (lines) and spherical spreading (dashed lines) are also shown. A profile of the depth contour of each transect is also included.

comprehend its significance. Thus, our estimates may be conservative if the distance over which the meaning of a whistle is transmitted is greater than the distance over which the first whistle frequency component is lost.

Estimated active space of whistles was highly variable according to habitat characteristics. For example, in shallow seagrasss areas, the active space of a 7–13 kHz whistle with a source level equal to 160 dB was estimated to be 301 m. Yet, the same whistle had an estimated active space over 13 km in channels. The regressions in Tables III and IV can be interpreted to understand how sound propagates on average in each transect. The slope of the regression indicates

TABLE II. Spectrum level background noise plus critical ratios in dB re 1 $\mu Pa^2/Hz$ of each frequency in channels (MC, SRC, AMS, and CC) and shallow-water transects (PSB1, PSB2, SKF, NWPSB1, NWPSB2, SPSB1, SPSB2, SAMS1, and SAMS2) in Sarasota Bay, Florida.

	Channels				Shallow-water transects								
Frequency	MC	SRC	AMS	CC	PBS1	PBS2	SKF	NWPSB1	NWPSB2	SPSB1	SPSB2	SAMS1	SAMS2
5 kHz	102.6	106.5	100.3	103.8	99.2	100.3	99.5	101.6	101.6	95.6	94.5	101.7	96.4
7 kHz	101.8	103.7	95.0	105.3	95.8	96.9	95.1	99.0	98.6	95.1	96.0	101.9	99.0
9 kHz	104.9	105.7	103.5	104.0	98.1	98.5	98.9	100.8	100.3	96.9	95.5	101.6	98.3
11 kHz	103.4	105.5	96.4	106.1	98.7	99.4	99.3	99.8	100.4	96.7	98.6	106.9	97.9
13 kHz	105.1	107.2	96.9	107.4	100.0	100.3	100.1	100.9	102.0	96.4	97.1	101.9	96.3
15 kHz	106.3	108.4	101.1	107.7	100.0	105.4	103.4	105.4	105.3	100.4	101.5	106.5	97.2
17 kHz	108.2	110.0	100.6	108.6	103.1	103.7	103.8	105.3	104.5	99.4	101.0	106.5	99.7
19 kHz	108.9	109.5	102.8	107.6	102.6	102.8	106.4	102.9	106.6	101.3	102.5	105.6	102.4

how sound levels fall off with distance. Although we did not find significant differences in the means of the distributions of regression slopes between habitats and between frequencies, it is important to note that individual transects showed a lot of variation, and the transmission loss over each transect is what is biologically relevant, not the mean for each habitat type. A slope of 20 indicates that sound follows a spherical spreading model. If the slope is 10, then it follows a cylindrical spreading model. The smallest slope was −13.8 dB, which was in a channel. The steepest slope (i.e., greatest propagation loss) was −28.0 dB, which was in shallow seagrass. The assumption of the propagation models is that a habitat is homogenous in its propagation characteristics. However, propagation characteristics are likely to be variable and significant changes in habitat features like water depth, substrate, seagrass cover, channel shape (horizontal and vertical), and channel width can alter active space greatly. Figures 2 and 3 show how propagation is affected by changes in water depth and how some frequencies are more attenuated than others with depth. In the Main Channel (MC), received levels of an 11 kHz signal fluctuated by more than 15 dB with water depth changes of up to 1 m. A source level change of even 5 dB can decrease or increase the active space of whistles significantly as dolphins navigate throughout the heterogeneous environment. Although the active space of some whistles could be more than 13 km in channels, in reality most channels sampled are shorter than 3–4 km before their course changes direction. Yet, an active space of even a few kilometers is still be a significant range of communication between dolphins in Sarasota Bay where channels are narrow and shallow areas extend hundred of meters.

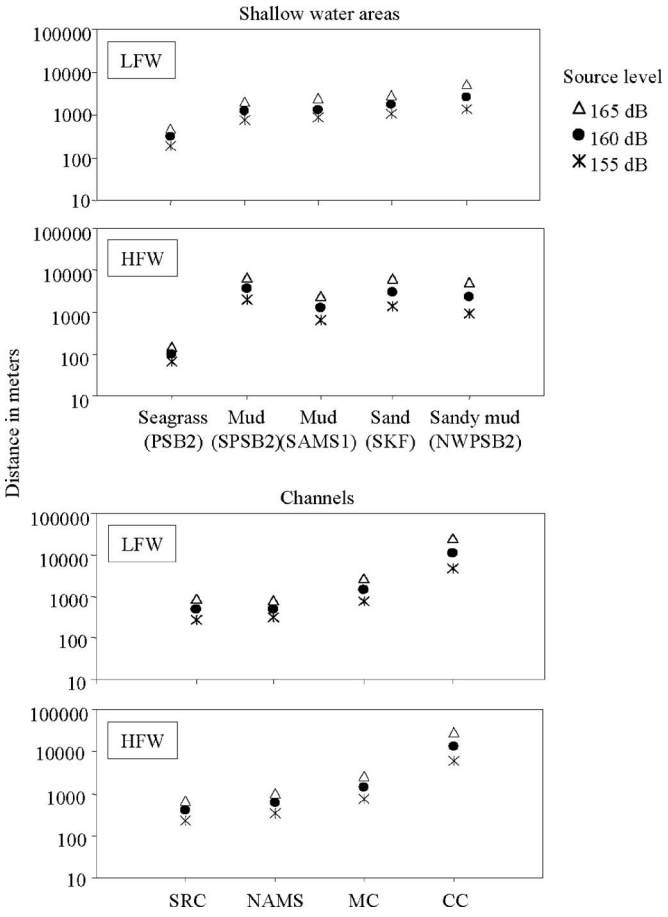


FIG. 4. Estimated active space of low-frequency whistles (LFW = 7–13 kHz) and high-frequency whistles (HFW = 13–19 kHz) with different source levels in channels and shallow water areas of Sarasota Bay, Florida. Distance is presented in logarithmic scale. Labels are defined in Table I.

Variation in estimated active space was observed in shallow water areas. A low-frequency whistle with a source level of 155 dB attenuated up to seven times more in seagrass areas than in areas with other bottom type with respect to distance (Fig. 4). In seagrass, the active space of the same whistle was estimated to be approximately 186 m. In contrast, the active space of the whistle was much greater, approximately 1319 m, in a sandy mud habitat of comparable depth. The scale of the variation in active space is relevant to dolphins because in Sarasota Bay dolphins use seagrass areas extensively to feed (Waples, 1995). If during feeding events dolphins use whistles to maintain contact, their ability to communicate over long distances (several hundreds of meters) is greatly reduced in seagrass areas.

Seagrasses act as a complicated three dimensional diffraction system which causes dramatic filtering of the sound (Willey and Richards, 1978). D. P. Nowacek *et al.* (2001) reported that the transmission loss of frequencies between 4 kHz and 8 kHz was up to 6 dB greater over 50 m in shallow seagrass areas than in shallow areas with mud or sand

TABLE III. Regression equations representing the mathematical model of sound propagation for different frequencies in shallow water habitats. Note: To calculate the propagation distance, source levels are added to the intercept of the equation. Frequencies from 7 to 13 kHz are referred to as a low-frequency whistle and from 13 to 19 kHz are referred to as a high-frequency whistle. R^2 is the proportion of variability explained by the propagation model.

Experimental transect/habitat	Frequency	Propagation distance (x=distance in meters, y=whistle source level)	R^2
PSB2 Seagrass	5 kHz	$y=-19.9\log(x)+3.5$	0.9
	7 kHz	$y=-22.6\log(x)+4.7$	0.9
	9 kHz	$y=-23.9\log(x)-2.2$	0.9
	11 kHz	$y=-22.9\log(x)-4.4$	0.9
	13 kHz	$y=-22.1\log(x)+1.9$	1.0
	15 kHz	$y=-28.0\log(x)+1.3$	1.0
	17 kHz	$y=-26.0\log(x)+0.9$	0.9
	19 kHz	$y=-26.6\log(x)+2.0$	0.9
SKF Sand	5 kHz	$y=-25.8\log(x)+13.9$	0.8
	7 kHz	$y=-23.5\log(x)+11.1$	0.8
	9 kHz	$y=-16.8\log(x)+3.2$	0.9
	11 kHz	$y=-15.9\log(x)+0.6$	0.8
	13 kHz	$y=-16.4\log(x)+2.9$	0.9
	15 kHz	$y=-15.2\log(x)+2.1$	0.8
	17 kHz	$y=-14.5\log(x)+2.5$	0.8
	19 kHz	$y=-26.6\log(x)-1.0$	0.9
NWPSB2 Sandy mud	5 kHz	$y=-25.8\log(x)+8.4$	0.8
	7 kHz	$y=-23.5\log(x)+0.9$	0.9
	9 kHz	$y=-16.8\log(x)+0.7$	0.7
	11 kHz	$y=-15.9\log(x)+0.4$	0.7
	13 kHz	$y=-16.4\log(x)-0.5$	0.8
	15 kHz	$y=-15.2\log(x)+1.9$	0.8
	17 kHz	$y=-14.5\log(x)+2.8$	0.7
	19 kHz	$y=-26.6\log(x)-9.3$	0.7
SPSB2 Mud	5 kHz	$y=-22.9\log(x)+10.6$	0.6
	7 kHz	$y=-20.2\log(x)+6.8$	0.6
	9 kHz	$y=-22.9\log(x)+6.0$	0.8
	11 kHz	$y=-14.9\log(x)+9.8$	0.6
	13 kHz	$y=-19.2\log(x)+5.4$	0.9
	15 kHz	$y=-16.8\log(x)+7.8$	0.7
	17 kHz	$y=-15.5\log(x)+2.6$	0.8
	19 kHz	$y=-15.9\log(x)+5.1$	0.8
AMS1 Mud	5 kHz	$y=-16.1\log(x)+5.9$	0.8
	7 kHz	$y=-17.2\log(x)+5.8$	0.9
	9 kHz	$y=-20.9\log(x)+8.0$	0.9
	11 kHz	$y=-18.0\log(x)+2.5$	0.9
	13 kHz	$y=-18.5\log(x)+3.6$	0.9
	15 kHz	$y=-16.0\log(x)+5.1$	0.8
	17 kHz	$y=-16.2\log(x)+4.7$	0.9
	19 kHz	$y=-26.6\log(x)+5.4$	0.9

TABLE IV. Regression equations representing the mathematical model of sound propagation for different frequencies in channels. Note: To calculate the propagation distance, source levels are added to the intercept of the equation. Frequencies from 7 to 13 kHz are referred to as a low-frequency whistle and from 13 to 19 kHz are referred to as a high-frequency whistle. R^2 is the proportion of variability explained by the propagation model.

Experimental transect/habitat	Frequency	Logarithmic equation (x=distance in meters, y=whistle source level)	R^2
Main Channel	5 kHz	$y=-15.4\log(x)+4.8$	0.9
	7 kHz	$y=-14.9\log(x)+1.5$	0.8
	9 kHz	$y=-18.4\log(x)+2.7$	0.8
	11 kHz	$y=-15.5\log(x)+7.6$	0.7
	13 kHz	$y=-18.3\log(x)+2.8$	0.8
	15 kHz	$y=-12.8\log(x)+2.4$	0.6
	17 kHz	$y=-15.9\log(x)+3.4$	0.9
	19 kHz	$y=-13.8\log(x)+3.5$	0.8
San Remo Channel	5 kHz	$y=-17.0\log(x)-3.9$	0.8
	7 kHz	$y=-18.0\log(x)+0.3$	0.9
	9 kHz	$y=-20.9\log(x)-0.0$	0.9
	11 kHz	$y=-21.0\log(x)+7.3$	0.8
	13 kHz	$y=-19.9\log(x)+0.9$	0.9
	15 kHz	$y=-18.1\log(x)-3.2$	0.9
	17 kHz	$y=-14.5\log(x)+2.5$	0.9
	19 kHz	$y=-16.7\log(x)-3.2$	0.9
Anna Maria Sound	5 kHz	$y=-13.9\log(x)-6.4$	0.8
	7 kHz	$y=-16.5\log(x)-3.8$	0.9
	9 kHz	$y=-20.9\log(x)+1.5$	0.9
	11 kHz	$y=-20.7\log(x)+2.3$	0.9
	13 kHz	$y=-21.8\log(x)+7.1$	0.9
	15 kHz	$y=-24.5\log(x)+7.1$	0.9
	17 kHz	$y=-21.8\log(x)+3.4$	0.9
	19 kHz	$y=-22.4\log(x)+8.7$	0.9
Cortez Channel	5 kHz	$y=-13.9\log(x)+3.4$	0.8
	7 kHz	$y=-16.5\log(x)+1.2$	0.9
	9 kHz	$y=-20.9\log(x)+5.5$	0.8
	11 kHz	$y=-20.7\log(x)+8.1$	0.7
	13 kHz	$y=-21.8\log(x)+4.2$	0.7
	15 kHz	$y=-24.5\log(x)-0.2$	0.7
	17 kHz	$y=-21.8\log(x)+3.6$	0.8
	19 kHz	$y=-22.4\log(x)+7.3$	0.9

bottoms. Sound propagated farther in habitats with sparse grass than habitats with dense grass (D. P. Nowacek *et al.*, 2001). Other factors that influence sound propagation in shallow waters include surface conditions, bottom contour variability, water column sound speed properties, bathymetry, vegetation, and bottom type (Urick, 1975; Forrest *et al.*, 1993; Forrest, 1994; Jensen, 2001).

Among nonseagrass areas, we found that whistles were more attenuated in areas with mud bottoms followed by sand and sandy mud bottoms. Mud, clay, and silt cause energy to dissipate more than sand and gravel (Urick, 1975; Jensen, 2001). Although the number of experiments conducted in areas of each sediment type was low, similar results in the sound transmission of sand and mud areas were found by Marsh and Schulkin (1962). Our estimates of whistle active space in those habitats were greater than 500 m.

The active space of whistles was also variable among channels, but the variation was not always directly related to channel depth as expected. In two cases, active space was related to channel width. In the widest channel, the estimated active space of high-frequency whistles was over 4 km. The mean separation distance of females and their dependent calves in channels (99 m±48 m) was much shorter than the estimated active space. A large whistle active space could result in high masking noise for whistles if other dolphins use the same whistle frequency range (Janik, 2000b) and are whistling at the same time. A large communication range may also result in animals being able to eavesdrop on acoustic interactions (Janik, 2000b). This could be a benefit if dolphins use whistles to look for specific associates since

individual dolphins produce distinctive signature whistles (Caldwell *et al.*, 1990). For example, a large active space could help male dolphins to find each other. In *Tursiops sp.* male dolphins are known to form coalitions with particular males during the mating season. Male coalitions are formed to control and sometimes steal receptive females from other males (Connor *et al.*, 1992). For receptive females, a large active space of whistles could be costly if they are avoiding harassing males.

Janik (2000b) found that the active space of whistles in a channel can decrease by several kilometers when whistle frequency is higher than 10 kHz. In our study, the decrease in active space by several kilometers occurred at 13 or 15 kHz in all four channels. Differences between studies could be due to environmental differences. The channels of our study were much shallower (3.1–4.1 m) and narrower (139–390 m) than the channel studied by Janik (2000b; depth=10 m, width=500 m). Such environmental variability makes it difficult to provide general conclusions about the behavior of specific frequencies. Sound propagation may change as water depth changes with tidal events, temperature gradients, freshwater inputs, and obstacles in the sound path.

Changes in whistle source level or frequency could help the transmission of whistles over long distances when associates are temporarily separated. The active space of whistles almost doubles when there is a 5 dB increase in source level (almost a doubling in energy; Fig. 4). This suggests that there is an advantage if dolphins produce louder whistles in habitats where propagation is poor than in other habitats. There could also be an advantage in changes in frequency. For example, the active space of low-frequency whistles was larger in 75% of the shallow water areas than the active space of high-frequency whistles in the same habitat. Studies examining the characteristics of whistles used in different habitats will help in understanding how dolphins communicate over long distances.

It is important to note that our estimates of active space assume that both the whistling and receiving dolphins are 1 m below the water surface. However, dolphins move vertically within the water column as they surface to breath, search for food, or socialize with other dolphins. It is possible that the active space changes with the position of the whistling and receiving dolphins because the transmission loss of sound varies within the water column. The influence of the sender location on signals has been suggested for birds (Lohr *et al.*, 2003) and other aquatic animals (Forest *et al.*, 1993). Active space also varies with hearing thresholds, which can vary greatly among dolphins of different age groups (Houser and Finneran, 2005). Thus, each dolphin might have a different communication range. Another aspect to take into account is that studies have found that whistles of some delphinds are somewhat directional at higher frequencies, especially in respect to harmonics (Lammers and Au, 2003). Thus, even in a homogeneous environment, the active space may not be radially symmetric around the dolphin.

The results of this study suggest that whistle active space is greater than the distances commonly used to identify dolphins as members of a group: 10-m chain rule (Smolker *et al.*, 1993) and a radius of 100 m (Wells and Scott, 1990). Although such definitions make data collection manageable and replicable in the field, they may not be biologically meaningful if dolphins communicate over much larger distances. Since whistles are thought to be used by dolphins to maintain group cohesion and to communicate over long distances (Janik and Slater, 1998; Norris *et al.*, 1994; Smolker *et al.* 1993), understanding the communication range to define dolphin groups is extremely important.

ACKNOWLEDGMENTS

Many thanks to Peter Tyack, Joseph Torres, and Colin Chapman who contributed significantly to the development of this work. We thank the field assistants helping with the sound transmission experiments: Alison Boler, Meagan Dunphy-Daly, Ida Eskerson, Lindsey Fenderson, Lu Lu, Marde McHenry, Minelia Miravete, and Kathleen Mohning. The assistance of the Sarasota Dolphin Research Program staff in the lab was much appreciated and was integral part of this work. Janet Gannon did the figure of the study area and Jay Sprinkel provided statistical advice. Damon Gannon and Mandy Hill provided helpful comments on the manuscript. Funding for the research was provided by a grant from the NOAA Fisheries to RSW and graduate student fellowships from the College of Marine Science, University of South Florida, to EQR. Research was conducted under NMFS Scientific Research Permit No. 522-1569 issued to RSW by the National Marine Fisheries Service.

Altman, J. (**1974**). "Observational study of behavior: Sampling methods," Behaviour **49**, 227–267.

Brenowitz, E. A. (**1982**). "The active space of Red-winged Blackbird song," J. Comp. Physiol. **142**, 511–522.

Brown, C. H., Beecher, M. D., Moody, D. B., and Stebbins, W. C. (**1979**). "Locatability of vocal signals of Old World monkeys: design features for the communication of position," J. Comp. Physiol. **93**, 806–819.

Caldwell, M. C., Caldwell, D. K., and Tyack, P. L. (**1990**). "Review of the signature-whistles for the Atlantic bottlenose dolphin," in *The Bottlenose Dolphin*, edited by S. Leatherwood and R. R. Reeves (Academic, California), pp. 199–234.

Connor, R. C., Smolker, R. A., and Richards, A. F. (**1992**). "Dolphin alliances and coalitions," in *Coalitions and Alliances in Humans and Other Animals*, edited by A. H. Harcourt and F. B. M. de Waal (Oxford University Press, Oxford), pp. 415–443.

Forrest, T. G. (**1994**). "From sender to receiver: propagation and environmental effects on acoustic signals," Am. Zool. **34**, 644–654.

Forrest, T. F., Miller, G. L., and Zagar, J. R. (**1993**). "Sound propagation in shallow water: Implications for acoustic communication by aquatic animals," Bioacoustics 4, 259–270.

Houser, D., and Finneran, J. (**2005**). "Auditory evoked potentials (AEP) methods for population-level assessment of hearing sensitivity in bottlenose dolphins," J. Acoust. Soc. Am. **117**, 2408.

Janik, V. M. (**2000a**). "Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*)," Science **289**, 1355–1357.

Janik, V. M. (**2000b**). "Source levels and the estimated active space of bottlenose dolphin (*Tursiops truncates*) whistles in the Moray Firth, Scotland," J. Comp. Physiol. [A] **186**, 673–680.

Janik, V. M., and Slater, P. J. B. (**1998**). "Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls," Anim. Behav. **56**, 829–838.

Jensen, F. B. (**2001**). "Acoustics, shallow water," *Encyclopedia of Ocean Sciences*, edited by J. H. Steele (Academic, California), pp. 89–96.

Johnson, C. S. (**1967**). "Sound detection thresholds in marine mammals," in *Marine Bioacoustics Vol. II*, edited by W. N. Tavolga (Pergamon, New York), pp. 247–260.

Johnson, C. S. (**1968**). "Masked tonal thresholds in the bottle-nosed por-

poise," J. Acoust. Soc. Am. **44**, 965–967.
Klump, G. M. (**1996**). "Bird communication in the noisy world," in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell University Press, New York), pp. 321–338.
Lammers, M. O., and Au, W. W. L. (**2003**). "Directionality in the whistles of Hawaiian spinner dolphins (*Stenella longirostris*): A signal feature to cue direction of movement?," Marine Mammal Sci. **9**, 213–220.
Ljungblad, D. K., Scoggins, P. D., and Gilmartin, W. G. (**1982**). "Auditory threshold of captive Eastern Pacific bottle-nosed dolphin, *Tursiops* spp.," J. Acoust. Soc. Am. **72**, 1726–1729.
Lohr, B., Wright, T. F., and Dooling, R. J. (**2003**). "Detection and discrimination of natural calls in masking noise by birds: Estimating the active space of a signal," Anim. Behav. **65**, 763–777.
Marsh, H. W., and Schulkin, M. (**1962**). "Shallow-water transmission," J. Acoust. Soc. Am. **34**, 863–864.
Michelsen, A., and Larsen, O. N. (**1983**). "Strategies for acoustic communication in complex environments," in *Neuroethology and Behavioral Physiology Roots and Growing Points*, edited by F. Huber and H. Mark (Springer-Verlag, New York), pp. 321–331.
Norris, K. S., Würsig, B., Wells, R. S., and Würsig, M. (**1994**). *The Hawaiian Spinner Dolphin* (University of California Press, Berkeley, CA).
Nowacek, D. P., Buckstaff, K. C., Johnson, M. P., and Wells, R. S. (**2001**). "Transmission loss of vessel noise in manatee environments," Florida Fish and Wildlife Conservation Commission Contract 99023, Final Report, pp. 25.
Nowacek, S. M., Wells, R. S., and Solow, A. R. (**2001**). "Short-term effects of boat traffic on bottlenose dolphins, *Tursiops truncatus*, in Sarasota Bay, Florida," Marine Mammal Sci. **17**, 673–688.
Rogers, P. H., and Cox, C. (**1998**). "Underwater sound as a biological stimulus," in *Biology of Aquatic Animals*, edited by J. Atema, R. Fay, A. N. Popper, and W. N. Tavolga (Springer-Verlag, Berlin), pp. 131–149.
Sayigh, L. S., Tyack, P. L., and Wells, R. S. (**1993**). "Recording underwater sounds of free-ranging dolphins while underway in a small boat," Marine Mammal Sci. **9**, 213–220.
Slabbekoorn, H. (**2004**). "Habitat-dependent ambient noise: Consistent spectral profiles in two African forest types," J. Acoust. Soc. Am. **116**, 3727–3733.
Smolker, R. A., Mann, J., and Smuts, B. B. (**1993**). "Use of signature whistles during separations and reunions by wild bottlenose dolphin mothers and infants," Behav. Ecol. Sociobiol. **33**, 393–402.
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 (The University of Chicago Press, Chicago), pp. 270–308.
Urick, R. J. (**1975**). *Principles of Underwater Sound* (Mc-Graw-Hill, New York).
Waples, D. (**1995**). "Activity budgets of free-ranging bottlenose dolphins (*Tursiops truncatus*) in Sarasota Bay, Florida," M.S. thesis, University of California, Santa Cruz, California (unpublished).
Wells, R. S. (**2003**). "Dolphin social complexity: Lessons from long-term study and life history," in *Animal Social Complexity: Intelligence, Culture, and Individualized Societies*, edited by F. B. M. de Waal and P. L. Tyack (Harvard University Press, Massachusetts), pp. 32–56.
Wells, R. S., and Scott, M. D. (**1990**). "Estimating bottlenose dolphin population parameters from individual identification and capture-release techniques," in *Individual Recognition of Cetaceans: Use of Photo-Identification and Other Techniques to Estimate Population Parameters*, edited by P. S. Hammond, S. A. Mizroch, and G. P. Donovan (Reports of the International Whaling Commission. Special Issue 12), pp. 407–415.
Willey, R. H., and Richards, D. G. (**1978**). "Physical constraints on acoustic communication in the atmosphere: Implications for the evolution of animal vocalizations," Behav. Ecol. Sociobiol. **3**, 69–94.
Wilson, B. (**1995**). "The ecology of bottlenose dolphins in the Moray Firth, Scotland: A population at the northern extreme of the species' range," Ph.D. thesis, University of Aberdeen, Scotland (unpublished).

Vocal development in captive harbor seal pups, *Phoca vitulina richardii*: Age, sex, and individual differences[a)]

Christin B. Khan[b)] and Hal Markowitz
Department of Biology, San Francisco State University, 1600 Holloway Avenue, San Francisco, California 94132

Brenda McCowan
Veterinary Medicine Teaching and Research Center, School of Veterinary Medicine, University of California at Davis, 18830 Road 112, Tulare, California 93274

(Received 22 July 2005; revised 8 June 2006; accepted 21 June 2006)

The goal of this research was to determine when harbor seal pup vocalizations become sufficiently distinctive to allow individual recognition. A total of 4593 calls were analyzed from 15 captive pups. Nineteen were harsh, broadband, staccato calls used in an aggressive context. The rest were tonal "mother attraction calls," having an inverted "v"- or "u"-shaped spectrogram with harmonics and a fundamental frequency around 200–600 Hz. Calls were individually distinctive even in pups less than 2 weeks old, suggesting that mothers may be able to recognize pup vocalizations at this early age. Classification rates from discriminant function analysis were generally comparable to those of other phocids and less than in otariids, supporting the theory that recognition is more highly developed in otariids. Significant differences were found between male and female pup calls, and there were significant interactions between pup sex and age. The results of this study should be interpreted with caution until the findings are verified in wild harbor seal pups. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2226530]

PACS number(s): 43.80.Ka, 43.80.Ev [WWA] Pages: 1684–1694

I. INTRODUCTION

Parent-offspring vocal recognition is common in many birds and mammals, and increases reproductive success by restricting parental care to genetically related offspring. Among pinnipeds, vocal recognition is widespread, although it is considered better developed in otariids (fur seals and sea lions) than in phocids (true seals) (see Insley *et al.*, 2003 for a review). An important factor driving the evolution of recognition in pinnipeds may be the maternal strategy, with long periods of dependence and maternal separations correlated with mutual recognition (Insley, 1992; Insley *et al.*, 2003). Otariids have a "foraging cycle strategy" marked by a long period of offspring dependence during which the mother alternates foraging trips to sea with time on land to nurse her pup (Bonner, 1984; Oftedal *et al.*, 1987). In contrast, many phocids have a "fasting strategy" in which mothers remain close to their pups for a short and intense lactation period (Bonner, 1984; Oftedal *et al.*, 1987). The finding that harbor seals forage during lactation challenged the traditional dichotomy of the otariid foraging cycle strategy and the phocid fasting strategy (Bowen *et al.*, 1992). This begs the question of whether harbor seals have faced greater selective pressure on mother-pup recognition than have other phocids.

Female harbor seals give birth to a single pup annually, and the unusually precocial pups follow their mothers both on land and in the water within minutes. Unlike most pinnipeds, female harbor seals do not emit a "pup attraction call." The pups, however, are quite vocal and begin to emit "mother attraction calls" several hours after birth (Lawson and Renouf, 1985). These calls presumably function to aid the mother and pup in maintaining contact and reuniting once separated (Renouf, 1984). Preliminary research found that pup calls appeared to be individually distinctive (Perry and Renouf, 1988; Renouf, 1984), and a captive adult was able to distinguish between playbacks recorded from different pups (Renouf, 1985).

Parents typically begin to recognize offspring calls just after they become individually distinctive and before offspring intermingle with conspecifics (bank swallow: Beecher *et al.*, 1981a; McArthur, 1979; razorbill: Insley *et al.*, 2003; McArthur, 1982; goat: Lenhardt, 1977). In pinnipeds, the ability to recognize pup calls immediately postpartum has not been tested, and factors affecting the ontogeny of recognition are unknown. Insley (2001) suggested that the critical time for otariids is 1–2 weeks postpartum when the mother leaves on her first foraging trip. Subantarctic fur seal mothers that were tested 7 months after parturition recognized playbacks from 1–2-day-old pups (Charrier *et al.*, 2003a), suggesting that pup calls were distinctive at this early age.

Harbor seals are faced with the challenge of maintaining contact with mobile pups on land and in the water immediately after birth. Female harbor seals make diving bouts accompanied by their pups as early as 0–3 days postpartum (Bowen *et al.*, 1999), and a mechanism for long-distance

[a)]Portions of this work were presented in "Aerial vocal development in captive harbor seal pups, *Phoca vitulina richardsi* and *Phoca vitulina concolor*," in Proceedings of the Defenders of Wildlife Conference on Carnivore Biology and Conservation, Monterey, California, November 2002, and "Vocal development in captive harbor seal pups, *Phoca vitulina*," in Proceedings of the 15th Biennial Conference on the Biology of Marine Mammals, Greensboro, North Carolina, December 2003.

[b)]Electronic mail: christin@cbkhan.com

TABLE I. Summary acoustic variables calculated for each vocalization.

Acoustic variable	Description
Start frequency	Frequency at the beginning of the call (Hz)
Finish frequency	Frequency at the end of the call (Hz)
Minimum frequency	Lowest frequency in the call (Hz)
Maximum frequency	Highest frequency in the call (Hz)
Mean frequency	Average frequency across the call (Hz)
Frequency at peak amplitude	Frequency at which peak amplitude occurs (Hz)
Frequency range	Maximum frequency minus the minimum frequency (Hz)
Maximum freq/mean freq	Maximum frequency divided by the mean frequency
Mean freq/minimum freq	Mean frequency divided by the minimum frequency
Duration	The length of time of the call (ms)
Location of minimum frequency	Location of minimum frequency as a percentage of the call
Location of maximum frequency	Location of maximum frequency as a percentage of the call
Start frequency slope	Slope of the initial third of the call (Hz/ms)
Middle frequency slope	Slope of the middle third of the call (Hz/ms)
Finish frequency slope	Slope of the final third of the call (Hz/ms)
Coefficient of variation	Measure of the magnitude of the frequency modulation
Jitter factor	Weighted measure of the frequency modulation
Coefficient of frequency modulation	Unweighted measure of the frequency modulation

recognition should be in place by this time. Breeding rookeries are often in areas with low visibility, strong currents, and high ambient noise. Under these circumstances, the ability to discriminate among pups is important immediately after parturition.

Long-distance pup recognition using acoustic cues is only possible after pups begin to produce individually distinctive vocalizations. The primary goal of this research was to determine when harbor seal pup calls become sufficiently distinctive to allow individual recognition using acoustic cues. Additionally, we expected sex differences in the vocalizations, as well as changes in call structure due to physiological maturity. Therefore, a secondary goal of this research was to explore age and sex information in the calls. The present study is the first quantitative research on vocal development in harbor seal pups.

II. METHODS

A. Subjects and study site

Aerial vocalizations were recorded from 15 captive harbor seal pups, *Phoca vitulina richardii*, at The Marine Mammal Center in Sausalito, California from March 13th until May 18th of 2002. The pups were born in the wild and brought in for rehabilitation at the age of 1–7 days postpartum; they were in acoustic and physical contact with other harbor seal pups (as they are under natural conditions). The Director of Veterinary Science, Dr. Frances Gulland, assigned an estimated date of birth to each pup based on the presence and condition of umbilicus, teeth, and white blood cell counts.

B. Recording vocalizations

Recordings were made with an Audio-technica 835b condenser microphone (frequency response of 40–20 000 Hz) and a Sony TCD-D8 DAT recorder at sampling rates of 44.1 and 48 kHz, allowing for an analysis range of at least 0–20 kHz. Flipper tags allowed individual pups to be easily identified, and recordings were immediately followed by a voice announcement to ensure that calls were assigned to the correct individual. All recordings produced during this study are archived by the Macaulay Library of Natural Sounds at the Cornell Laboratory of Ornithology in Ithaca, New York (www.animalbehaviorarchive.org).

C. Acoustical analysis

Recordings were digitized with a Roland UA-30 USB audio interface with DAT ACS digital connection cord using the software COOL EDIT PRO 1.2 (Syntrillium Software Corp., Phoenix, AZ) and an ESS Maestro sound card at a sampling rate of 44.1 or 48 kHz and 16-bit resolution. Spectrograms were generated using a 1024-point FFT with a Hamming filter. Vocalizations with high signal-to-noise ratio and unambiguous identity were selected and cued for subsequent digital analysis. Background noise was removed in COOL EDIT PRO with a Butterworth bandpass filter using cutoffs of 100 Hz and 15 kHz. Vocalizations were measured using a modified version of the Contour Similarity Technique (for detailed descriptions of this technique, see McCowan, 1995, and McCowan and Reiss, 2001). Sixty time and frequency points were extracted across the duration of each call by following the dominant frequency (the frequency of highest amplitude). Nineteen summary acoustic variables were calculated from these measurements (Table I).

D. Statistical analysis

1. Individual differences

Stepwise, cross-validation discriminant function analysis was performed to classify vocalizations based on the identity of the calling pup using SPSS 10.0 (SPSS Inc., Chicago, IL). To meet the assumptions of this test, continuous variables were tested for normality and log transformed as necessary. When sample sizes were unequal, percent correct expected

by chance was calculated separately for each individual; the overall percent correct expected by chance was determined by adding the number of correctly classified calls expected for each individual and then dividing this sum by the total number of calls analyzed (Tabachnick and Fidell, 2001). Chi-square was then used to test whether correct classification was significantly greater than expected by chance. Holm's sequential Bonferroni method was used to adjust alpha levels for multiple comparisons (Rice, 1989).

Since the fundamental frequency has been implicated as an individually distinctive feature in the vocalizations of harbor seal pups (Renouf, 1984; Perry and Renouf, 1988), cross-validation discriminant function analysis was also run on a subset of the data for which calls were filtered and measured from the fundamental frequency only. The 19 summary acoustic variables were calculated from 60 time and frequency points extracted from the fundamental frequency (Table I). Six pups with large sample sizes were chosen for this analysis, and 40 calls were randomly selected from each pup during each of three age groups ($n=240$ calls per age group). This analysis enabled a comparison of individuality as a function of age.

The Potential for Individuality Coding was calculated from the ratio of the between-individual coefficient of variation (CV_b) and the mean value of the within-individual coefficients of variation (CV_i) (Robisson *et al.*, 1993; Lengagne *et al.*, 1997). Coefficients of variation were calculated by dividing the standard deviation by the mean.

2. Sex differences

Stepwise cross-validation discriminant function analysis was conducted on a subset of the data containing 1500 calls to classify vocalizations based on the sex of the calling pup using SPSS 10.0. Five pups of each sex with large sample sizes were chosen for this analysis, and 150 calls were randomly selected from each pup. Continuous variables were tested for normality and log transformed as necessary. Chi-square was used to test whether correct classification was significantly greater than expected by chance. Mixed-effects linear regression, which is discussed further in the next subsection, was used to identify which variables were significantly different between males and females.

3. Age differences

Differences in vocalizations attributable to age were examined using mixed-effects linear regression with the nested repeated measure (or random grouped effect) of "recording session within id," in S-PLUS 2000 statistical software (Data Analysis Products Division., MathSoft, Seattle, WA; Pinheiro and Bates, 2000). Regression diagnostics confirmed that the variables followed normality and variance assumptions. The covariates or fixed effects included sex and age (in weeks) of the vocalizing pup. The nested repeated measure of recording session within id accounted for any clustering due to individuals or recording session, and was based on the assumption that there would be more variability in call characteristics of the individuals across recording sessions than there would be in recording sessions across individuals.

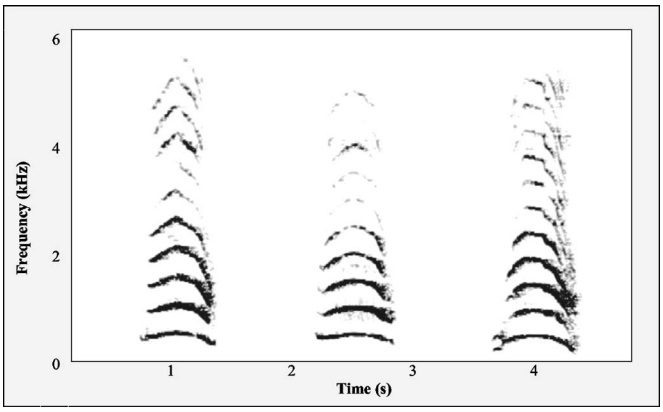


FIG. 1. Spectrogram of typical harbor seal pup vocalizations with inverted "v" or "u" shape.

III. RESULTS

A. General call characteristics

The total data set included 4593 vocalizations recorded from 15 harbor seal pups. Most ($n=4574$) calls were relatively tonal, having an inverted "v-" or "u"-shaped spectrogram with the fundamental frequency around 200–600 Hz and harmonics (Fig. 1). See Table II for descriptive statistics on measured call characteristics made by following the dominant frequency of the call.

B. Aggressive vocalizations

The remaining 19 calls were harsh, broadband, staccato calls used in an aggressive context toward another seal pup or human (Fig. 2). Most of these aggressive calls ($n=15$) were recorded from one individual. Aggressive vocalizations were recorded in a sequence of one to three calls and were produced by both male and female pups. The major energy in these vocalizations was from 200–2000 Hz, and duration ranged from 0.12 to 1.07 s (mean=0.57 s).

C. Individual differences

1. Discriminant function analysis

The 19 calls identified as aggressive vocalizations were excluded from the analysis, resulting in a sample size of 4574 calls. With all age groups combined, stepwise cross-validation discriminant function analysis classified 29.3% of calls correctly, which is significantly more than the 11.5% correct expected by chance (X^2, $P<0.0001$). See Table III. Classification scores for individuals ranged from 3.6% to 53.0% correct, and were significantly greater than expected by chance in 11 of the 15 harbor seals (X^2, $P<0.0001$).

The discriminant function analysis generated ten significant canonical discriminant functions representing the linear combinations of the variables that maximally separate groups in multidimensional space. The first canonical discriminant function accounted for 52.5% of the variation between individuals and loaded most heavily with duration and jitter factor. The second discriminant function, which accounted for 19.3% of the variation, loaded most heavily with mean frequency, the mean frequency divided by the minimum frequency, and call duration.

TABLE II. Descriptive statistics of harbor seal pup calls (n=4593) calculated from 60 time and frequency points extracted by following the dominant frequency of each call.

Acoustic Variable	Minimum	Maximum	Mean	SD	CV
Start frequency	102.63	8067.20	435.44	371.52	0.85
Finish frequency	103.44	7315.00	457.63	304.93	0.67
Minimum frequency	102.63	1145.10	237.47	80.36	0.34
Maximum frequency	343.27	8595.80	1224.88	572.72	0.47
Mean frequency	255.05	1845.95	661.32	208.58	0.32
Frequency at Peak Amplitude	180.73	3106.10	805.39	348.76	0.43
Frequency range	33.71	8411.50	987.41	574.19	0.58
Maximum frequency / mean freq	1.04	10.51	1.89	0.71	0.38
Mean frequency / minimum freq	1.05	11.80	3.04	1.31	0.43
Duration	117.19	2444.04	610.73	203.70	0.33
Location of minimum frequency	0.00	1.00	0.47	0.44	0.95
Location of maximum frequency	0.00	1.00	0.47	0.24	0.52
Start frequency slope	−6.42	3.48	0.17	0.30	—
Middle frequency slope	−3.48	1.49	−0.01	0.22	—
Finish frequency slope	−2.33	7.39	−0.16	0.27	—
Coefficient of variation	0.57	3836.37	177.30	174.87	0.99
Jitter factor	2.08	67.50	15.31	7.57	0.49
Coefficient of frequency modulation	0.04	3.60	0.59	0.35	0.59

2. Fundamental frequency discriminant function analysis

Cross-validation discriminant function analysis was also run on a subset of the data containing 40 calls each from six pups during three different age groups (n=240 per age group) for which calls were filtered and measured from the fundamental frequency only. Overall percent-correct classification scores were significantly greater than expected by chance for each of the three age groups as well as all ages combined (X^2, $P<0.0001$, Table IV). Classification scores using measurements made from just the fundamental frequency were higher than classification scores using measurements made by following the dominant frequency of the call.

The fundamental frequency analysis for all age groups combined generated five significant canonical discriminant functions. The first function accounted for 45.7% of the variation between individuals and loaded most heavily with mean frequency and finish frequency. The second discriminant function, which accounted for 29.6% of the variation, loaded most heavily with maximum frequency and the coefficient of frequency modulation. The third function accounted for an additional 16.2% of the variation between individuals and loaded most heavily with duration and coefficient of variation.

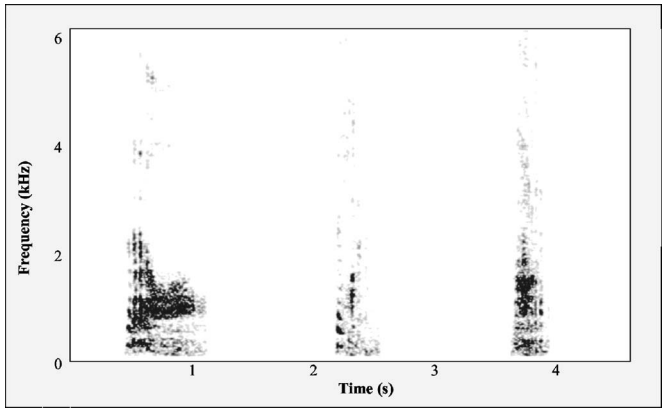


FIG. 2. Spectrogram of three harsh, broadband, staccato calls used in an aggressive context by harbor seal pups (compare with Fig. 1).

3. Potential for individuality coding

Measured call characteristics from the vocalizations of harbor seal pups recorded in this study had Potential of Individuality Coding values ranging from 0.98 to 1.18 (Table V).

D. Sex differences

1. Discriminant function analysis

Cross-validation discriminant function analysis was conducted on a subset of the data containing 1500 calls (150 from each of five male and five female pups) to classify vocalization based on the sex of the calling pup. With all age groups combined, 66.1% of calls were classified correctly, which is significantly more than the 50% correct expected by chance (X^2, $P<0.0001$). The analysis generated one significant canonical discriminant function that accounted for all of the variation between sexes, and loaded most heavily with duration, inflection factor, and mean frequency.

2. Mixed-effects linear regression

When the sex of the vocalizing pup was the only fixed effect considered, only one of the summary acoustic variables was significantly different between males and females: the coefficient of frequency modulation (Table VI). Male harbor seal pup calls had higher coefficients of frequency modulation than did those of female pups.

However, when the interaction between the sex and age of the vocalizing pup was considered, several other variables revealed significant differences: mean frequency, frequency at peak amplitude, duration, location of maximum frequency,

TABLE III. Results of cross-validation discriminant function analysis to classify individual harbor seal pups based on call characteristics measured by following the dominant frequency. Bold numbers are correctly classified calls; others are incorrectly classified (overall 29.3% correct, $n=4574$). Chi-square tested significance. Asterisk indicates significance using Holm's sequential Bonferroni method to adjust alpha levels for multiple comparisons (Rice, 1989).

	Predicted group membership																	
ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	% Correct	% Change	X^2
1	**46**	3	11	4	7	1	2	16	14	5	29	9	19	3	1	27.1%	3.7%	<0.0001*
2	2	**109**	27	1	39	16	37	34	21	4	21	13	23	5	24	29.0%	8.2%	<0.0001*
3	4	4	**63**	2	8	0	7	1	14	3	16	20	7	2	6	40.1%	3.4%	<0.0001*
4	26	8	26	**6**	8	4	8	24	4	1	15	16	7	2	10	3.6%	3.6%	0.9833
5	1	8	4	0	**12**	0	2	3	0	3	5	1	3	2	1	26.7%	1.0%	<0.0001*
6	37	57	38	6	54	**49**	26	52	50	36	67	45	39	25	21	8.1%	13.2%	0.0003
7	15	55	29	3	21	11	**35**	21	15	6	35	13	19	5	19	11.6%	6.6%	<0.0005*
8	14	20	6	5	29	17	18	**42**	14	7	24	12	27	10	4	16.9%	5.4%	<0.0001*
9	7	4	25	6	5	3	6	7	**64**	1	8	23	6	7	2	36.8%	3.8%	<0.0001*
10	5	2	15	0	30	4	4	17	7	**9**	10	16	11	18	4	5.9%	3.3%	0.0738
11	11	3	13	1	9	3	5	3	6	3	**81**	11	0	1	6	51.9%	3.4%	<0.0001*
12	13	17	145	1	13	31	14	20	101	12	32	**571**	59	35	13	53.0%	23.5%	<0.0001*
13	41	26	9	5	34	10	21	55	67	25	19	36	**225**	31	13	36.5%	13.5%	<0.0001*
14	1	16	9	0	17	2	2	9	6	14	13	16	21	**18**	3	12.2%	3.2%	<0.0001*
15	1	31	26	1	26	7	13	8	4	6	22	6	14	8	**12**	6.5%	4.0%	0.0916

In this case, calls were correctly classified significantly *less* often than expected by chance.

TABLE IV. Results of cross-validation discriminant function analysis using a subset of the data containing 40 calls each from six pups during three different age groups ($n=240$ per age group) measured from the fundamental frequency alone. The "% Correct" column contains percentages of calls classified correctly. Chi-square tested whether correct classification was significantly greater than the 16.67% expected by chance. Asterisk indicates significance using Holm's sequential Bonferroni method to adjust alpha levels for multiple comparisons (Rice, 1989).

	0–14 days old		15–28 days old		29–42 days old		All ages combined	
ID	% Correct	X^2	% Correct	X^2	% Correct	X^2	% Correct	X^2
2	72.5%	<0.0001*	60.0%	<0.0001*	80.0%	<0.0001*	74.2%	<0.0001*
6	15.0%	0.7773	42.5%	<0.0001*	42.5%	<0.0001*	30.8%	<0.0001*
7	55.0%	<0.0001*	42.5%	<0.0001*	70.0%	<0.0001*	50.0%	<0.0001*
8	40.0%	0.0001*	42.5%	<0.0001*	30.0%	0.0237	38.3%	<0.0001*
13	67.5%	<0.0001*	52.5%	<0.0001*	57.5%	<0.0001*	54.2%	<0.0001*
15	15.0%	0.7773	47.5%	<0.0001*	62.5%	<0.0001*	39.2%	<0.0001*
Overall	44.2%	<0.0001*	47.9%	<0.0001*	57.1%	<0.0001*	47.8%	<0.0001*

TABLE V. Potential for individuality coding (PIC) of harbor seal pup calls. The potential for individuality coding is the ratio of the between-individual coefficient of variation (CV_b) and the mean value of the within-individual coefficients of variation (CV_i).

Acoustic variable	Mean±SD	CV_b	Mean CV_i	PIC
Start frequency	435.44±371.52	0.85	0.76	1.12
Finish frequency	457.63±304.93	0.67	0.61	1.10
Minimum frequency	237.47±80.36	0.34	0.33	1.03
Maximum frequency	1224.88±572.72	0.47	0.44	1.06
Mean frequency	661.32±208.58	0.32	0.29	1.12
Freq at peak amplitude	805.39±348.76	0.43	0.40	1.08
Frequency range	987.41±574.19	0.58	0.55	1.05
Maximum freq / mean freq	1.89±0.71	0.38	0.36	1.04
Mean freq / minimum freq	3.04±1.31	0.43	0.41	1.04
Duration	610.73±203.70	0.33	0.29	1.15
Location of minimum frequency	0.47±0.44	0.95	0.97	0.98
Location of maximum frequency	0.47±0.24	0.52	0.51	1.02
Coefficient of freq modulation	0.59±0.35	0.59	0.58	1.01
Coefficient of variation	177.30±174.87	0.99	0.84	1.18
Jitter factor	15.31±7.57	0.49	0.49	0.99

TABLE VI. Results of linear mixed effects regression with the random effect of "recording session within id" on harbor seal pup calls. Covariates include: fixed effect of "age" (as a continuous variable), fixed effect of "sex," and the interaction between the fixed effects of "age" and "sex" (* indicates significance at the alpha <0.05 level). Female is referent.

	Age		Sex		Sex and Age	
Acoustic Variable	Coef	p(df=298)	Coef	p(df=13)	Coef	p(df=298)
Minimum frequency	−4.7697	<0.0001*	−1.2063	0.7364	−0.0119	0.9876
Maximum frequency	−19.402	0.0044*	−50.2890	0.2133	−1.244	0.8542
Mean frequency	2.1601	0.5246	−9.9786	0.6189	−9.4408	0.0057*
Freq at peak amplitude	5.4630	0.2370	−16.8019	0.5283	−12.2869	0.0081*
Frequency range	−16.0775	0.0190*	−49.0871	0.2334	1.1875	0.8618
Duration	−0.7073	0.7815	−33.1473	0.1145	9.3709	0.0003*
Location of min freq	−0.0019	0.0025*	−0.0003	0.9116	0.0002	0.7532
Location of max freq	−0.0128	0.0001*	−0.0287	0.1079	0.0074	0.0195*
Start frequency slope	−0.0003	0.9544	−0.0025	0.9311	−0.0018	0.7161
Middle frequency slope	−0.0045	0.0772	−0.0189	0.1095	0.0022	0.3942
Finish frequency slope	−0.0052	0.1908	−0.0029	0.9025	0.0061	0.1308
Coefficient of freq mod	−0.0182	<0.0001*	−0.0405	0.0479*	0.0047	0.1689
Coefficient of Variation	−5.3022	<0.0001*	−11.1505	0.1239	2.5226	0.0472*
Jitter factor	−0.4796	<0.0001*	−0.7555	0.1422	0.2358	0.0034*

coefficient of variation, and jitter factor (Table VI). See Fig. 3 for graphs generated from the linear mixed-effects models demonstrating the interaction between the age and sex of the vocalizing pup. Young harbor seal pups of both sexes tended to have similar values for mean frequency, but as the pups matured, female pup vocalizations became higher in mean frequency while male pup vocalizations became lower in mean frequency. Similarly, young pups of both sexes had similar values for frequency at peak amplitude, but as pups matured, female pup vocalizations became higher while male pup vocalizations became lower. The mean duration of female pup vocalizations remained relatively constant throughout the study period, but male pups exhibited an increase in the duration of their calls as they matured. The location of maximum frequency, coefficient of variation, and jitter factor values were all higher in female pup calls during the first few weeks of age, becoming gradually lower in females and higher in males so that male values were higher than females in pups over 5 weeks of age.

E. Age differences

In addition to maturational changes in male and female pup calls (See Sec. III D above), there were also significant age differences in pup calls regardless of sex. When the age of the pup was the only fixed effect considered in the repeated measures linear regression, the following summary acoustic variables remained significant: location of the maximum frequency, coefficient of variation, and jitter factor (Table VI, Fig. 3). In addition, the following variables decreased significantly with age in pups of both sexes: minimum frequency, maximum frequency, frequency range, location of the minimum frequency, and coefficient of frequency modulation (Table VI). See Fig. 4 for graphs generated from the linear mixed-effects models demonstrating the effect of age.

IV. DISCUSSION

A. General call characteristics

Most calls produced by harbor seal pups during this study were relatively tonal, having an inverted "v-" or "u"-shaped spectrogram with the fundamental frequency around 200–600 Hz and harmonics. These calls had acoustic features similar to those commonly described in the literature and presumed to function as "mother attraction calls" (Scheffer and Slipp, 1944; Bishop, 1967; Newby, 1973; Renouf, 1984; Ralls *et al.*, 1985; Perry and Renouf, 1988). Pup vocalizations of the western Atlantic harbor seal, *Phoca vitulina concolor*, have a fundamental frequency at about 350 Hz and harmonics (Ralls *et al.*, 1985). The mean (±SD) duration of calls in this study (0.57±0.30 s) falls between the reported values for the Atlantic subspecies of 0.31±0.14 s (Perry and Renouf, 1988), 0.81±0.19 s (Ralls *et al.*, 1985), and 1.1±0.7 s (Van Parijs and Kovacs, 2002).

B. Aggressive vocalizations

Nineteen of the calls recorded in this study were harsh, broadband, staccato calls used in an aggressive context towards another seal pup or human. These calls resemble "growls" or "hisses" reported in previous studies (Bishop, 1967; Sullivan, 1982; Van Parijs and Kovacs, 2002), although no quantitative data was provided. The earliest age at which a pup was recorded vocalizing aggressively in this study was 16 days postpartum. Further research is needed to explore the development of aggressive behavior in harbor seal pups.

C. Individual differences

The results of this study indicate that harbor seal pup calls are individually distinctive, confirming preliminary research (Perry and Renouf, 1988; Renouf, 1984). Classifica-

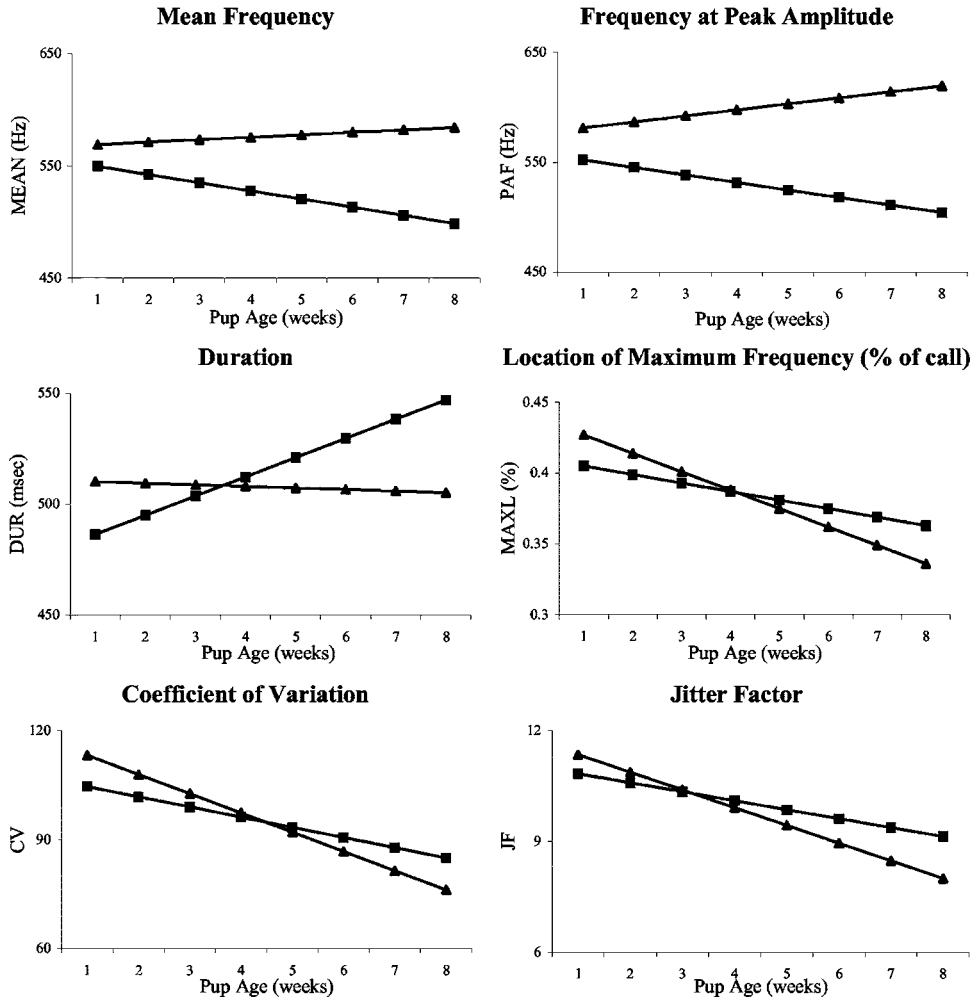


FIG. 3. Graphs generated from repeated measures linear regression demonstrating the interaction between the fixed effects of "age" and "sex" for variables describing harbor seal pup, *Phoca vitulina richardii*, vocalizations that were significant ($\alpha < 0.05$) with the random effect of "recording session within id." Squares represent males, and triangles represent females.

tion scores from discriminant function analysis were 29% correct overall. For effective maternal recognition under natural conditions, female harbor seals need to discriminate between pups at a much higher rate of success; several factors may contribute to this discrepancy. Classification scores from discriminant function analysis are based solely on the information contained in a single vocalization, whereas harbor seal mothers have an entire calling bout available to them as well as geographic, visual, and olfactory cues.

Individually distinctive contact calls have been found in all pinnipeds studied to date (for a review, see Insley *et al.*, 2003). Discriminant function analysis has resulted in a wide range of percent-correct classification scores (see Table VII), but whether this reflects species differences in stereotypy or is an artifact of sample size differences is unknown. Percent-correct classification scores have been shown to increase corresponding to a decrease in the number of individuals and the number of signals per individual (Bee *et al.*, 2001). Despite these limitations, the range of classification scores does conform to current ideas regarding recognition abilities in pinnipeds. Percent-correct classification scores tend to be higher for otariids than phocids (see Table VII), supporting the theory that recognition is more highly developed in otariids. Harbor seals are phylogenetically and behaviorally similar to grey seals, with females of both species making regular separations from their pups, and the two species have similar percent-correct classification scores in discriminant function analysis (Table VII). The ability of females to recognize pup calls has been studied in two different populations of grey seals with differing results. Grey seal females on Sable Island in Canada were able to recognize the calls of their own pups, whereas on the Isle of May in Scotland females failed to recognize pups (McCulloch *et al.*, 1999; McCulloch and Boness, 2000). The only test of recognition in harbor seals was with a single captive female who was able to discriminate between recordings of two different pups (Renouf, 1985). Based on what we have learned about grey seals, recognition abilities in harbor seals should be studied in several populations. It is also interesting to note that Hawaiian monk seals, the only other pinniped in which mothers were found not to recognize their pups, had the lowest classification scores reported for any species at 14% correct (Job *et al.*, 1995).

Harbor seal contact calls were individually distinctive even in pups less than 2 weeks old. Classification scores were higher using measurements from the fundamental fre-

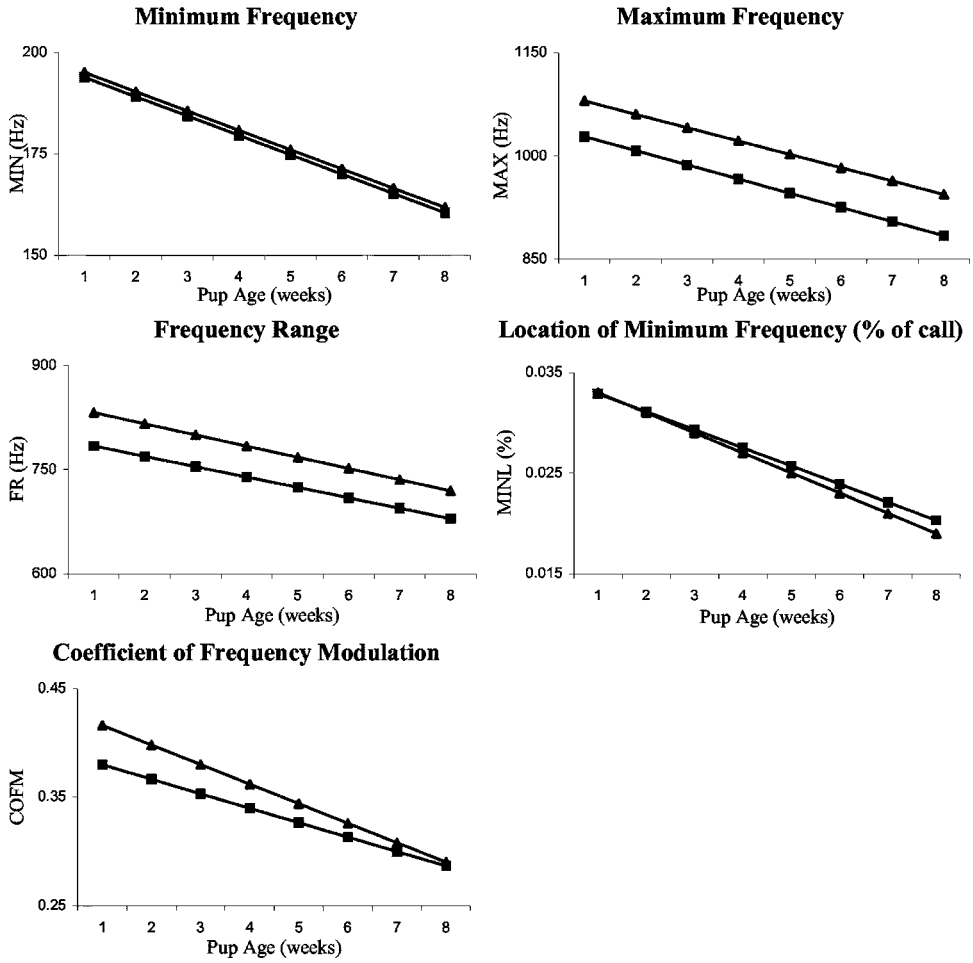


FIG. 4. Graphs generated from repeated measures linear regression demonstrating the fixed effect of "age" for variables describing harbor seal pup, *Phoca vitulina richardii*, vocalizations that were significant ($\alpha < 0.05$) with the random effect of "recording session within id." Squares represent males, and triangles represent females.

quency than from the dominant frequency of the call. This result is consistent with previous findings that the fundamental frequency is a reliable indicator of identity in harbor seals (Renouf, 1984; Perry and Renouf, 1988) and other pinnipeds (South American fur seals: Phillips and Stirling, 2000; subantarctic fur seals: Charrier *et al.*, 2002; northern fur seals: Insley, 1992; grey seals: McCulloch *et al.*, 1999; northern elephant seals: Insley, 1992; Hawaiian monk seals: Job *et al.*,

TABLE VII. Empirical studies of "mother attraction calls" in pinnipeds demonstrating distinctive calls and/or maternal recognition. (%"Correct," "Seals," "Calls" refers to the results of discriminant function analysis; Y=yes, N=no).

Species	%Correct	#Seals	#Calls	Are pup calls individually distinctive?		Do mothers recognize pup calls?	
Otariids							
Southern sea lions	89%	33	99	Y	Fernandez-Juricic *et al.*, 1999		
Subantarctic fur seals	83%	10	100	Y	Page *et al.*, 2002; Charrier *et al.*, 2002; Roux and Jouventin, 1987	Y	Charrier *et al.*, 2002, 2003a; Roux and Jouventin, 1987
New Zealand fur seals	79%	10	100	Y	Page *et al.*, 2002		
Northern fur seals	79%	8	160	Y	Insley, 1992	Y	Insley, 2000, 2001
Antarctic fur seals	52%	9	90	Y	Page *et al.*, 2002		
S. American fur seals	51%	13	260	Y	Phillips and Stirling, 2000		
Phocids							
Northern elephant seals	64%	8	126	Y	Insley, 1992	Y	Petrinovich, 1974
Grey seals	32%	20	216	Y	McCulloch *et al.*, 1999;	Y	McCulloch and Boness, 2000
					Caudron *et al.*, 1998	N	McCulloch *et al.*, 1999
Harbor seals	29%	15	4574	Y	*Present study*; Perry and Renouf, 1988; Renouf, 1984	Y	Renouf, 1985[a]
Harp seals	…	…	…	Y	Van Opzeeland and Van Parijs, 2004		
Hawaiian monk seals	14%	15	428	Y	Job *et al.*, 1995	N	Job *et al.*, 1995

[a]Based on the ability of one captive female to discriminate between recordings of two different pups.

1995; southern sea lions: Fernandez-Juricic *et al.*, 1999). Duration appears to be somewhat stereotyped in harbor seal pup calls, and has also been implicated as a distinctive feature in other pinnipeds (South American fur seals: Phillips and Stirling, 2000; grey seals: McCulloch *et al.*, 1999; northern elephant seals: Insley, 1992; Hawaiian monk seals: Job *et al.*, 1995; southern sea lions: Fernandez-Juricic *et al.*, 1999; harp seals: Van Opzeeland and Van Parijs, 2004).

The Potential for Individuality Coding is a means of comparison across studies with different methodologies; values greater than 1 indicate the potential for individual recognition since the between-individual variability is greater than the within-individual variability (Robisson *et al.*, 1993; Lengagne *et al.*, 1997). The Potential for Individuality Coding ratios for call characteristics measured in this study are all very close to 1, suggesting that harbor seal pup calls are weakly individualized. Mother attraction calls from subantarctic fur seal pups had ratios from 0.89 to 3.36, with the highest values for fundamental frequency, frequency at the first peak amplitude, and frequency modulation (Charrier *et al.*, 2002). Playback experiments confirmed that these calls are individually distinctive and that mothers respond selectively to the calls of their own pup (Charrier *et al.*, 2002). Call duration was the only variable that could be directly compared to the present study, and appears to be more individualized in the vocalizations of harbor seal pups (1.15) than in subantarctic fur seal pups (0.89).

D. Sex differences

Significant differences were found between male and female pup calls. Male pup calls had higher coefficients of frequency modulation than calls of female pups. Young pups of both sexes had similar values for mean frequency, but as they matured, female calls became higher in mean frequency while male calls became lower. Likewise, young animals of both sexes had similar values for frequency at peak amplitude, but as they matured, females became higher while males became lower. The mean duration of female vocalizations remained relatively constant throughout the study period, but the mean duration of male calls increased as they matured. The location of maximum frequency, coefficient of variation, and jitter factor values were all higher in female pup calls during the first few weeks of age, becoming gradually lower in females and higher in males so that male values were highest in pups over 5 weeks of age.

Sex differences have also been found in the vocalizations of harp seal pups, *Phoca groenlandica*; males and females were separated on the first split of the classification tree, and female calls were classified correctly more often than male calls (Van Opzeeland and Van Parijs, 2004). Sex differences in vocalizations may function as a means for parents to allocate resources differentially based on offspring sex, and/or reflect early development of sexual dimorphism in vocal behavior (Saino *et al.*, 2003). There is no evidence for differential resource allocation in harbor seals. The sex ratio is 1:1 both at birth and weaning (Ellis, 1998), and although male harbor seal pups do weigh more than females at weaning, this reflects greater birth mass of males rather than differential maternal investment during lactation (Ellis, 1998; Bowen *et al.*, 2001). Sex differences in pup calls are also unlikely to reflect early development of sexual dimorphism in vocal behavior, since this call disappears from the vocal repertoire at weaning. However, sex differences in harbor seal pup vocalizations may reflect early development of the sexual size dimorphism characteristic of adults. This size dimorphism is already apparent at birth, with male pups weighing more than females (Ellis, 1998).

E. Age differences

Pups produced calls with well-defined acoustic features as young as 2 days postpartum. These mother attraction calls appear to encode identity information from a young age, although call structure does change throughout maturation. Discriminant function analysis classified calls correctly significantly more often than expected by chance even when pups were less than 2 weeks old. As discussed in the previous subsection, age-related changes in many call characteristics varied according to the sex of the pup. The following variables decreased significantly over time in pups of both sexes: location of the maximum frequency, coefficient of variation, jitter factor, minimum frequency, maximum frequency, frequency range, location of the minimum frequency, and coefficient of frequency modulation. These ontogenetic changes in frequency are similar to those reported in other mammals (vervet monkey: Hauser, 1989; pigtail macaque: Gouzoules and Gouzoules, 1989), and likely reflect the gradual enlargement of the vocal tract.

There has not been much research on the ontogeny of pup calls in pinnipeds. California sea lions may recognize vocalizations from their 2-week old pups, suggesting that the calls are individually stereotyped at this time (Gisiner and Schusterman, 1991; Schusterman *et al.*, 1992), although visual and olfactory cues could not be ruled out in these studies. Research on subantarctic fur seal pups showed significant age-related changes in the fundamental frequency and the percentage of quavering in pup calls (Charrier *et al.*, 2003a). Calls of young pups emphasized some high frequencies, while the spectral energy in the calls of older pups was concentrated on the first harmonics (Charrier *et al.*, 2003a).

F. Directions for future research

Further research is needed to confirm that harbor seal mothers recognize the calls of their pups under natural conditions in the wild. In addition to investigation of wild pup calls, observational studies of natural separations could determine whether vocal activity in pups facilitates successful reunions upon the return of the mother to the breeding rookery.

It is unknown which acoustic characteristics harbor seal mothers use to recognize the vocalizations of their own offspring; call characteristics measured in this study may not accurately reflect perceptually salient features. Playback studies using artificially manipulated signals such as those recently conducted by Charrier and colleagues on black-

headed gulls and subantarctic fur seals (Charrier *et al.*, 2001, 2002, 2003b) could help to elucidate the cues used by harbor seals in voice discrimination.

The present study did not examine the effects of motivational state on characteristics of harbor seal pup vocalizations, although recordings were made under similar conditions in an attempt to minimize the possible influence of motivation. Preliminary observations suggest that distressed pups have a greater number of calls per calling bout, a faster rate of call emission, and more harmonics (Renouf, 1984; Perry and Renouf, 1988). Future research should take motivational state information into account.

The harbor seal and the closely related spotted seal provide a unique opportunity for comparative study of the effects of breeding density on the vocal recognition system. Spotted seals, which breed on isolated ice floes, are expected to have a less reliable system of mother-offspring recognition than harbor seals, which breed in rookeries where there is greater risk of confusion. The absence of mother-pup recognition in spotted seals is suggested by a pup-exchange experiment in which a female spotted seal accepted a strange pup (Burns *et al.*, 1972). However, further research on wild populations of both harbor seals and spotted seals is needed to test these predictions.

ACKNOWLEDGMENTS

We would like to thank Dr. Frances Gulland and the volunteers of The Marine Mammal Center for providing logistical support. This research was conducted under IACUC protocol #01-712 approved by the IACUC of San Francisco State University on June 28th 2001.

Bee, M. A., Kozich, C. E., Blackwell, K. J., and Gerhardt, H. C. (**2001**). "Individual variation in advertisement calls of territorial male green frogs, *Rana clamitans*: Implications for individual discrimination," Ethology **107**, 65–84.

Beecher, M. D., Beecher, I. M., and Hahn, S. (1981). "Pavent-offspring recognition in bank swallows , (*Ripavia ripavia*): II. Development and acoustic bias," Anim. Behav. **29**, 95–101.

Bishop, R. H. (1967). "Reproduction, age determination, and behavior of the harbor seal, *Phoca vitulina* L., in the Gulf of Alaska," M.S. thesis, Univ. of Alaska. [Available at the NOAA National Marine Mammal Laboratory Library, Seattle, WA].

Bonner, W. N. (**1984**). "Lactation strategies in pinnipeds: problems for a marine mammalian group," Symposia of the Zoological Society of London. **51**, 253–272.

Bowen, W. D., Boness, D. J., and Iverson, S. J. (**1999**). "Diving behaviour of lactating harbour seals and their pups during maternal foraging trips," Can. J. Zool. **77**, 978–988.

Bowen, W. D., Ellis, S. L., Iverson, S. J., and Boness, D. J. (**2001**). "Maternal effects on offspring growth rate and weaning mass in harbour seals," Can. J. Zool. **79**, 1088–1101.

Bowen, W. D., Oftedal, O. T., and Boness, D. J. (**1992**). "Mass and energy transfer during lactation in a small phocid, the harbor seal (*Phoca vitulina*)," Physiol. Zool. **65**, 844–866.

Burns, J. J., Ray, G. C., Fay, F. H., and Shaughnessy, P. D. (**1972**). "Adoption of a strange pup by the ice-inhabiting harbour seal, *Phoca vitulina largha*," J. Mammal. **53**, 594–598.

Caudron, A. K., Kondakov, A. A., and Siryanov, S. V. (**1998**). "Acoustic structure and individual variation of grey seal (Halichoerus grypus) pup calls," J. Mar. Biol. Assoc. U.K. **78**, 651–658.

Charrier, I., Mathevon, N., and Jouventin, P. (**2002**). "How does a fur seal mother recognize the voice of her pup? An experimental study of *Arctocephalus tropicalis*," J. Exp. Biol. **205**, 603–612.

Charrier, I., Mathevon, N., and Jouventin, P. (**2003a**). "Fur seal mothers memorize subsequent versions of developing pups' calls: Adaptation to long-term recognition or evolutionary by-product?" Biol. J. Linn. Soc. **80**, 305–312.

Charrier, I., Mathevon, N., and Jouventin, P. (**2003b**). "Vocal signature recognition of mothers by fur seal pups," Anim. Behav. **65**, 543–550.

Charrier, I., Mathevon, N., Jouventin, P., and Aubin, T. (**2001**). "Acoustic communication in a black-headed gull colony: How do chicks identify their parents?" Ethology **107**, 961–974.

Ellis, S. L. (1998). "Maternal effects on offspring traits from birth through weaning in the harbour seal, *Phoca vitulina*," Ph.D. dissertation, Dalhousie Univ. [Available online through National Library of Canada at www.collectionscanada.ca/thesescanada/].

Fernandez-Juricic, E., Campagna, C., Enriquez, V., and Ortiz, C. L. (**1999**). "Vocal communication and individual variation in breeding South American sea lions," Behaviour **136**, 495–517.

Gisiner, R., and Schusterman, R. J. (**1991**). "California seal lion pups play an active role in reunions with their mothers," Anim. Behav. **41**, 364–366.

Gouzoules, H., and Gouzoules, S. (**1989**). "Design features and developmental modification of pigtail macaque, *Macaca nemestrina*, agonistic screams," Anim. Behav. **37**, 383–401.

Hauser, M. D. (**1989**). "Ontogenetic changes in the comprehension and production of vervet monkey (*Cercopithecus aethiops*) vocalizations," J. Comp. Physiol. **103**, 149–158.

Insley, S. J. (**1992**). "Mother-offspring separation and acoustic stereotypy: A comparison of call morphology in two species of pinnipeds," Behaviour **120**, 103–121.

Insley, S. J. (**2000**). "Long-term vocal recognition in the northern fur seal," Nature (London) **406**, 404–405.

Insley, S. J. (**2001**). "Mother-offspring vocal recognition in northern fur seals is mutual but asymmetrical," Anim. Behav. **61**, 129–137.

Insley, S. J., Phillips, A. V., and Charrier, I. (**2003**). "A review of social recognition in pinnipeds," Aquat. Mamm. **29**, 181–201.

Job, D. A., Boness, D. J., and Francis, J. M. (**1995**). "Individual variation in nursing vocalizations of Hawaiian monk seal pups, *Monachus schauinslandi* (Phocidae, Pinnipedia), and lack of maternal recognition," Can. J. Zool. **73**, 975–983.

Lawson, J. W., and Renouf, D. (**1985**). "Parturition in the Atlantic harbor seal, *Phoca vitulina concolor*," J. Mammal. **66**, 395–398.

Lengagne, T., Lauga, J., and Jouventin, P. (**1997**). "A method of independent time and frequency decomposition of bioacoustic signals: Inter-individual recognition in four species of penguins," C. R. Acad. Sci. III **320**, 885–891.

Lenhardt, M.L. (1997). "Vocal contour cues in maternal recognition of goat kids," Applied Animal Ethology **3**, 211–219..

McArthur, P.D. (1979). "Parent-young recognition in the piron jay: Mechanisms, ontogeny, and survival value," Ph.D. thesis, Northern Arizona University, Flagstaff, Arizona.

McArthur, P.D. (**1982**). "Mechanisms and development of parent-young vocal recognition in the piron jay (*Gymnorhinus cyanocephalus*," Anim. Behav. **30**, 62–74.

McCowan, B. (**1995**). "A new quantitative technique for categorizing whistles using simulated signals and whistles from captive bottlenose dolphins (Delphinidae, *Tursiops truncatus*)," Ethology **100**, 177–193.

McCowan, B., and Reiss, D. (**2001**). "The fallacy of 'signature whistles' in bottlenose dolphins: A comparative perspective of 'signature information' in animal vocalizations," Anim. Behav. **62**, 1151–1162.

McCulloch, S., and Boness, D. J. (**2000**). "Mother-pup vocal recognition in the grey seal (*Halichoerus grypus*) of Sable Island, Nova Scotia, Canada," J. Zool. **251**, 449–455.

McCulloch, S., Pomeroy, P. P., and Slater, P. J. B. (**1999**). "Individually distinctive pup vocalizations fail to prevent allo-suckling in grey seals," Can. J. Zool. **77**, 716–723.

Newby, T. C. (**1973**). "Observations on the breeding behavior of the harbor seal in the state of Washington," J. Mammal. **54**, 540–543.

Offedal, O. T., Boness, D. J., and Tedman, R. A. (1987). "The behavior, physiology, and anatomy of lactaction in the pinnipedia," Curr. Mammal. 1175–245.

Page, B., Goldsworthy, S. D., and Hindell, M. A. (**2002**). "Individual vocal traits of mother and pup fur seals," Bioacoustics 13, 121–143.

Perry, E. A., and Renouf, D. (**1988**). "Further studies of the role of harbour seal (*Phoca vitulina*) pup vocalizations in preventing separation of mother-pup pairs," Can. J. Zool. **66**, 934–938.

Petrinovich, L. (**1974**). "Individual recognition of pup vocalization by Northern elephant seal mothers," Z. Tierpsychol **34**, 308–312.

Phillips, A. V., and Stirling, I. (**2000**). "Vocal individuality in mother and pup South American fur seals, *Arctocephalus australis*," Marine Mammal Sci. **16**, 592–616.
Pinheiro, J. C., and Bates, D. M. (**2000**). *Mixed-Effects Models in S and S-Plus* (Springer, New York).
Ralls, K., Fiorelli, P., and Gish, S. (**1985**). "Vocalizations and vocal mimicry in captive harbor seals, *Phoca vitulina*," Can. J. Zool. **63**, 1050–1056.
Renouf, D. (**1984**). "The vocalization of the harbour seal pup (*Phoca vitulina*) and its role in the maintenance of contact with the mother," J. Zool. **202**, 583–590.
Renouf, D. (**1985**). "A demonstration of the ability of the harbour seal, *Phoca vitulina* (L.) to discriminate among pup vocalizations," J. Exp. Mar. Biol. Ecol. **87**, 41–46.
Rice, W. R. (**1989**). "Analyzing tables of statistical tests," Evolution (Lawrence, Kans.) **43**, 223–225.
Robisson, P., Aubin, T., and Bremond, J.-C. (**1993**). "Individuality in the voice of the emperor penguin *Aptenodytes forsteri*: Adaptation to a noisy environment," Ethology **94**, 279–290.
Roux, J.-P., and Jouventin, P. (1987). "Behavioral cues to individual recognition in the subantarctic fur seal, *Arctocephalus tropicalis*," NOAA Technical Report NMFS **51**, 95–102.
Saino, N., Galeotti, P., Sacchi, R., Boncoraglio, G., Martinelli, R., and Moller, A. P. (**2003**). "Sex differences in begging vocalizations of nestling barn swallows, *Hirundo rustica*," Anim. Behav. **66**, 1003–1010.
Scheffer, V. B., and Slipp, J. W. (**1944**). "The harbor seal in Washington State," Am. Midl. Nat. **32**, 373–416.
Schusterman, R. J., Hanggi, E. B., and Gisiner, R. (**1992**). "Acoustic signalling in mother-pup reunions, interspecies bonding, and affiliation by kinship in California sea lions (*Zalophus californianus*)," in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Y. Supin (Plenum, New York), pp. 533–551.
Sullivan, R. M. (**1982**). "Agonistic behavior and dominance relationships in the harbor seal, *Phoca vitulina*," J. Mammal. **63**, 554–569.
Tabachnick, B. G., and Fidell, L. S. (**2001**). *Using Multivariate Statistics* (Allyn & Bacon, Boston).
Van Opzeeland, I. C., and Van Parijs, S. M. (**2004**). "Individuality in harp seal, *Phoca groenlandica*, pup vocalizations," Anim. Behav. **68**, 1115–1123.
Van Parijs, S. M., and Kovacs, K. M. (**2002**). "In-air and underwater vocalizations of eastern Canadian harbour seals, *Phoca vitulina*," Can. J. Zool. **80**, 1173–1179.

Preliminary evidence for signature vocalizations among free-ranging narwhals (*Monodon monoceros*)[a)]

Ari D. Shapiro[b)]
Biology Department, Woods Hole Oceanographic Institution, MS #50, Woods Hole, Massachusetts 02543

(Received 14 March 2006; revised 9 June 2006; accepted 21 June 2006)

Animal signature vocalizations that are distinctive at the individual or group level can facilitate recognition between conspecifics and re-establish contact with an animal that has become separated from its associates. In this study, the vocal behavior of two free-ranging adult male narwhals (*Monodon monoceros*) in Admiralty Inlet, Baffin Island was recorded using digital archival tags. These recording instruments were deployed when the animals were caught and held onshore to attach satellite tags, a protocol that separated them from their groups. The signature content of two vocal categories was considered: (1) combined tonal/pulsed signals, which contained synchronous pulsatile and tonal content; (2) whistles, or frequency modulated tonal signals with harmonic energy. Nonparametric comparisons of the temporal and spectral features of each vocal class revealed significant differences between the two individuals. A separate, cross-correlation measure conducted on the whistles that accounted for overall contour shape and absolute frequency content confirmed greater interindividual compared to intraindividual differences. These data are consistent with the hypothesis that narwhals produce signature vocalizations that may facilitate their reunion with group members once they become separated, but additional data are required to demonstrate this claim more rigorously. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2226586]

PACS number(s): 43.80.Ka, 43.80.Ev [WAA] Pages: 1695–1705

I. INTRODUCTION

Signature vocalizations of animals acoustically encode individual or group identity and are characterized by unique sets of spectral and/or temporal attributes. The specific acoustic features required to distinguish between individuals according to their vocalizations have been measured in a host of taxa [e.g., birds: macaroni penguins (*Eudyptes chrysolophus*), Searby *et al.*, 2004, chiropterans: evening bats (*Nycticeius humeralis*), Scherrer and Wilkinson, 1993, canids: timber wolves (*Canis lupus*), Goldman *et al.*, 1995, primates: common marmosets (*Callithrix jacchus*), Jones *et al.*, 1993, pinnipeds: subantarctic fur seals (*Arctocephalus tropicalis*), Charrier *et al.*, 2001, 2003, cetaceans: bottlenose dolphins (*Tursiops truncatus*), Janik, 1999, Watwood *et al.*, 2005]. Playback experiments have demonstrated that animals can recognize signature signals and have illustrated the diversity of contexts in which signature vocalizations are used, including facilitating recognition between an infant and one or both of its parents [cliff swallows (*Hirundo pyrrhonota*): Stoddard and Beecher, 1983, tree swallows (*Tachycineta bicolor*): Leonard *et al.*, 1997, Mexican free-tailed bats (*Tadarida brasiliensis mexicana*): Balcombe, 1990, fur seals (*Arctocephalus tropicalis*): Charrier *et al.*, 2001, 2003, bottlenose dolphins: Sayigh *et al.*, 1998, Janik *et al.*, 2006], mate-pair recognition [king penguins (*Aptenodytes patagonicus*): Lengagne *et al.*, 2000], and group affiliation associated with territorial defense (North American bullfrog (*Rana catesbeiana*): Bee and Gerhardt, 2002, Arctic foxes (*Alopex lagopus*): Frommolt *et al.*, 2003). There are selective benefits for the signals produced in these contexts. Recognition is very important when one or both parents must allocate a finite amount of resources to their offspring, a scenario in which confusion is associated with high fitness costs. Mates or groups of individuals that consistently defend one another, their young, or their territory can benefit from individual or group recognition because it provides a system for remembering with whom they have shared mutual investments.

The proximate methods for achieving signature recognition can include imprinting, habituation, associative learning, and vocal learning. Vocal learning occurs when the respiratory, phonatory, and/or filter systems are employed to render signals more or less similar to acoustic models that are encountered through experience with other individuals (Janik and Slater, 1997, 2000). Contact calls, generally used by animals when they become separated from their social partner(s) or group to first locate one another and then mediate reunion, contain signature content in certain species and appear to be vocally learned. When placed in social groupings of unfamiliar individuals, the contact calls of male budgerigars (*Melopsittacus undulatus*), for example, initially converged and subsequently underwent continuous and synchronous changes (Farabaugh *et al.*, 1994). Evidence is accumulating for some species of nonhuman primates to possess vocal plasticity during adulthood despite its apparent absence during development (see review by Egnor and Hauser, 2004). Male chimpanzees (*Pan troglodytes*) produce pant hoots, long-distance vocalizations that seem to function in maintaining contact with and attracting allied individuals

[a)]Portions of this work were presented in "Vocal behavior of free-ranging Arctic narwhals (*Monodon monoceros*)," Proceedings of the 16th Biennial Conference on the Biology of Marine Mammals, San Diego, CA, December 2005.

[b)]Author to whom correspondence should be addressed. Electronic mail: ashapiro@whoi.edu

(Mitani and Nishida, 1993). Pant hoot convergence was observed among both chorusing dyadic pairs (Mitani and Gros-Louis, 1998) and larger groups containing 3–11 adult males (Marshall *et al.*, 1999). Similarly, several spectral and temporal parameters of the contact calls among pygmy marmosets (*Cebuella pygmaea*), referred to as trills, underwent parallel or convergent shifts between new adult mate pairs (Snowdon *et al.*, 1997; Snowdon and Elowson, 1999). Comparable observations were made when two naïve pygmy marmoset social groups of mixed-age composition were introduced (Elowson and Snowdon, 1994).

In principle, it is possible to conclude that signals contain signature content by demonstrating more interindividual than intraindividual variability. To offer sufficient proof, this result must be shown for at least several exemplars of the signature signals of each of several individuals. One of the most striking examples of signature vocalizations is found among bottlenose dolphins whose signature whistles, first identified by Caldwell and Caldwell (1965), appear to function as vocally-learned contact calls. Among bottlenose dolphins, whistle imitation appears to be an important agent in the ontogeny of this signature vocalization (Sayigh *et al.*, 1990, Miksis *et al.*, 2002) and in social communication both in captivity (Tyack, 1986) and in the wild (Janik, 2000, Fripp *et al.*, 2005). Bottlenose dolphin mother-calf pairs were more likely to orient towards one another's signature whistles when separated from each other and temporarily restrained than those of other individuals of the same corresponding age cohort (Sayigh *et al.*, 1998, Janik *et al.*, 2006). Captive animals produced their signature whistles most often when they separated themselves voluntarily and spontaneously from their mixed-age group by swimming into an adjacent tank compared to when they were swimming together (Janik and Slater, 1998). The remaining dolphins were also more likely to produce their signature whistles when an individual left the main tank. Finally, adult males that had strong social bonds with another male were most likely to use signature whistles when they were separated either due to temporary, artificial restraint or voluntarily when they were free-ranging, presumably to facilitate an eventual reunion (Watwood *et al.*, 2005). Collectively, these studies reveal the importance of signature whistles in maintaining contact between bottlenose dolphin individuals in artificial and natural settings alike and in both involuntary and voluntary contexts.

Despite some debate (see McCowan and Reiss, 1995, 2001; Janik, 1999 for discussion; Smolker and Pepper, 1999), the studies just described have helped to solidify the case for signature whistles in bottlenose dolphins and suggest the possibility of signature vocalizations among other long-lived, social odontocetes in an underwater environment where acoustic signaling is the most reliable and efficient form of communication. In addition, signature information can also be encoded at the group level, which can form the basis for acoustic badges of membership (e.g., Boughman, 1997). For example, five of six shared call types among killer whales (*Orcinus orca*) revealed distinctive structural differences between matrilineal units (Miller and Bain, 2000) and it is possible that other signature details allow differentiation between individual animals as well (Nousek *et al.*, in press). Cultural divergence of discrete call types appears to account for some of the subtle differences in the temporal and spectral features across these matrilineal units and even within pods (Deecke *et al.*, 2000).

The principal challenge for studying the signature signals of marine mammals involves the difficulty of assigning vocalizations in the wild unambiguously to the individual animal that produced them. In this manuscript, the possibility of signature signals among free-ranging narwhals (*Monodon monoceros*) was examined by recording the acoustic activity of two individuals with digital archival tags. These gregarious, long-lived Arctic odontocetes migrate distances of thousands of kilometers in large numbers with subpopulations moving in a coordinated fashion (Hay and Mansfield, 1989; Dietz and Heide-Jørgensen, 1995; Laidre *et al.*, 2004). They travel in groups that are often sex segregated and range in size from a few animals to dozens of individuals, although the stability or fluidity and interconnectedness of these assemblages remain unknown (reviewed in Hay and Mansfield, 1989).

Narwhals produce echolocation clicks with repetition rates between 2 and $>500/s$ (Ford and Fisher, 1978, Møhl *et al.*, 1990), maximum frequencies reaching at least 160 kHz (Miller *et al.*, 1995) and maximal source levels reaching 218 dB re 1 μPa (Møhl *et al.*, 1990). Miller *et al.*(1995) arbitrarily divided clicking into the two categories of train clicks produced at $\leqslant 30$ clicks/s and burst clicks produced at $\geqslant 40$ clicks/s. Pulsatile sounds featuring a repetition rate high enough to possess a tonal character with harmonically related sidebands (see Watkins, 1967) were called longer click series by Watkins *et al.* (1971) and pulsed tones by Ford and Fisher (1978). Characterized as narrow-band, these signals had durations between 0.56 and several seconds and spectral energy ranging from 500 Hz to 24 kHz. The repetition rate was generally constant although Watkins *et al.* (1971) reported a tendency for the repetition rate of these vocalizations to increase at the very beginning and slow down towards the end. In this manuscript, these signals will be referred to as combined tonal/pulsed signals. Finally, narrow-band, frequency modulated (FM) whistles have been described that generally last <1.0 s (range: 0.1–6.0 s) and have a frequency range between 300 Hz and 18 kHz (Ford and Fisher, 1978; Møhl *et al.*, 1990).

Although Ford and Fisher (1978) did not find any evidence for signature content among whistles, they speculated that the different pulsed tones in their recordings were produced by separate individuals as signature calls in a social context. They recorded series of the same tone growing louder and then softer, concluding that this resulted from one individual producing each series as it approached and then swam past a stationary hydrophone. This possibility was not conclusive since groups of animals were swimming by the recorder and multiple individuals could have been producing each tone. In addition, no data on differences in acoustic parameters were available to quantify the distinctiveness of the calls. In this study, we examined the possibility of signature vocalizations among free-ranging narwhals more closely. The results support this hypothesis for both combined tonal/pulsed signals and whistles, suggesting a social

FIG. 1. Narwhal shown with Crittercam (contained within the dashed ellipse) and DTAG (contained within the dashed rectangle) attached immediately before release. Photograph courtesy of Rune Dietz.

function for vocal production that is distinctive either at the individual or group level. Further work is recommended to confirm signature vocal production among additional animals and to ascertain the natural function of these vocalizations in the wild.

II. METHODS

A. Study area

Field work was conducted from 8–23 August 2004 at Kakiak Point, Admiralty Inlet on Baffin Island in Nunavut, Canada (73°40'N, 86°40'W). The inlet has a maximum depth of 720 m. Groups of narwhals ranging from approximately 5–30 individuals (pers. obs.) traveled into the inlet at this time of year once the ice had mostly melted. The field camp occupied a position about 500 m from a site used intermittently by the Inuit to hunt narwhals.

B. Equipment

This experiment employed a digital archival tag (DTAG) developed by Johnson and Tyack (2003) featuring a single hydrophone, pressure and temperature sensors, and a triaxial accelerometer and magnetometer, which recorded to flash memory. The sampling rate of the hydrophone was set to 96 kHz while the other sensors sampled at 50 Hz. A 16 bit ADC was used. Sigma delta conversion provided an effective antialiasing filter, dispensing with aliasing caused by energy exceeding the Nyquist frequency of 48 kHz. The tag attached noninvasively to individual animals via suction cups and its release was coupled to the release mechanism of the National Geographic Crittercam (see Marshall, 1998) that was deployed simultaneously. A VHF transmitter signaled the location of an attached tag intermittently as the animal surfaced and then regularly once the tag was released and floated to the water's surface.

C. Capturing and tagging protocol

The DTAG was deployed in collaboration with a satellite tagging project that required working with the animals onshore. As described by Dietz *et al.* (2001), a 50 m long and 10 m deep black net with 20×20 cm mesh was oriented perpendicular to the shore and kept afloat with 7–8 white buoys. The net was secured to the shore and in the water.

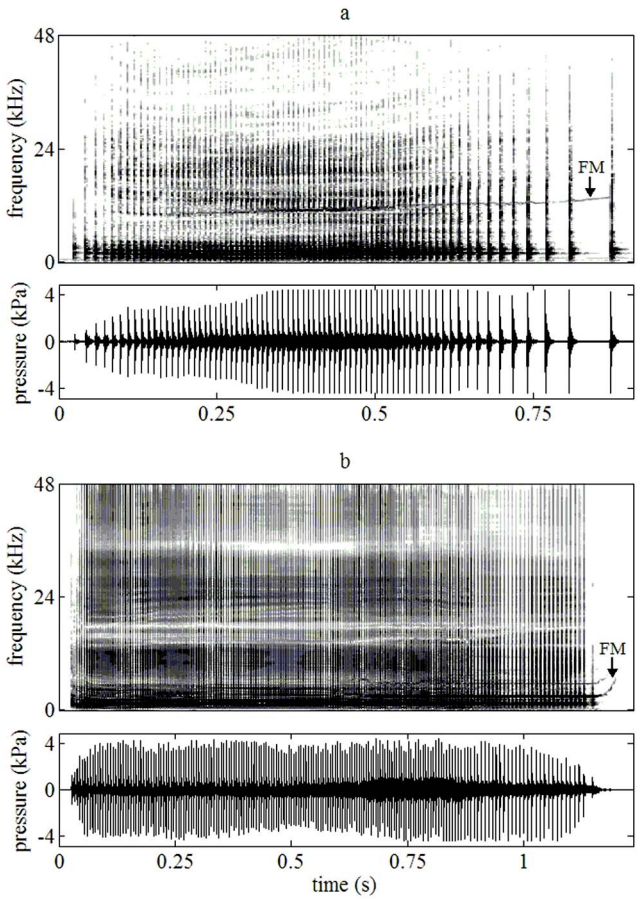


FIG. 2. Sample spectrograms (larger, top plots) and waveforms (smaller, bottom plots) of a combined tonal/pulsed signal produced by individual (a) mm224 and (b) mm226 with a FFT size and frame length of 512 points, 50% window overlap, and a maximum frequency displayed of 48 kHz. The low frequency energy associated with most of the pulses is likely due to the resonance of the air sacs involved in sound production or transmission. The solid arrows in both spectrograms indicate the synchronous FM tonal component produced by the tagged animal.

When weather conditions permitted, the net was deployed and monitored constantly for caught animals, signaled by the submergence of at least one of the buoys. As soon as a whale became entangled, two boats were dispatched immediately to bring the animal to the surface to breathe and tow it to shore with the assistance of a land-based crew hauling on the net. Once an animal was caught, the remainder of its group moved out of visual range, presumably continuing their migration deeper into the inlet. A fluke belt was used to keep the animal ashore and oriented with its head submerged and pointed into the water while its blowhole had access to the air at all times. Three males and five females were captured in all. During satellite tag attachment, blood samples were collected to assess overall health and stress levels.

Just before two of the adult males and one of the adult females were guided back to deeper water, a DTAG was attached to their dorsal sides ~0.5 m caudal to the blowhole (Fig. 1). These animals were not followed visually once they were released from shore so it was not possible to determine whether they eventually reunited with their group members. The VHF signal was monitored from the field camp on shore using two handheld yagi antennae. Once a regular VHF sig-

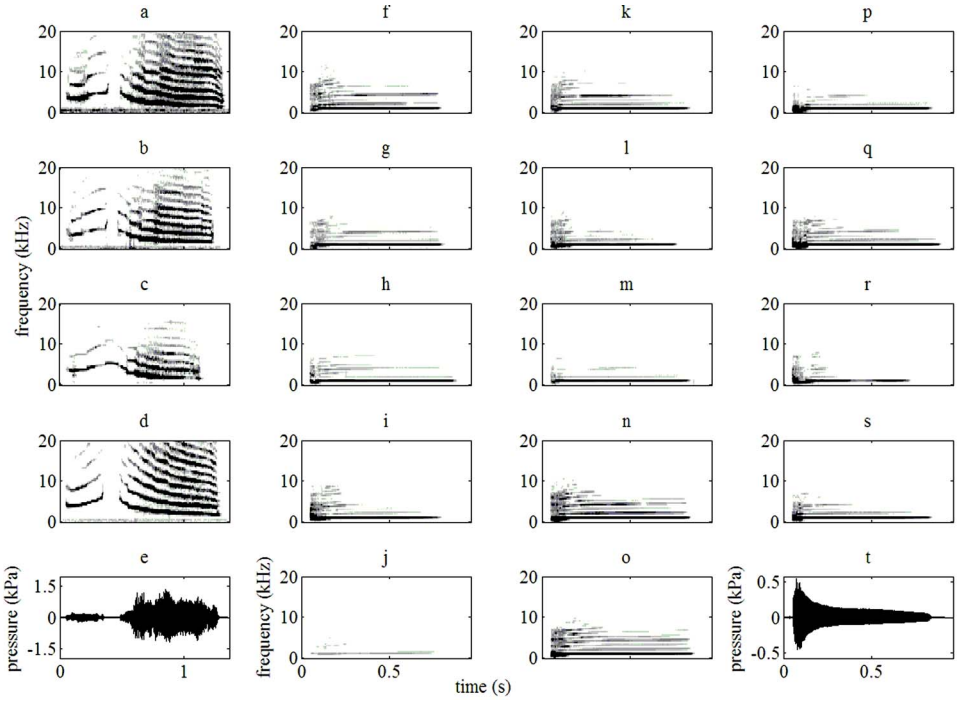


FIG. 3. Spectrogram composite of all four whistles of mm224 (a–d) and 14 of the 17 whistles of mm226 (f–s) with a FFT size and frame length of 512 points, 50% window overlap, and a maximum frequency displayed of 20 kHz. The remaining 3 whistles of mm226 resembled those displayed here but were excluded for graphical convenience. The waveforms displayed in subplots e and t are of the same whistles used to generate subplots d and s, respectively.

nal was detected from a tag that had released from the animal and the weather permitted, a boat was dispatched for recovery. If the VHF signal grew too faint to detect from shore, tracking was conducted from a higher altitude on the nearby cliffs for improved range. The first tag recorded for 2.54 h (male mm224), the second tag for 12.14 h (male mm226), and the third tag was not recovered. These two tagged males entered the inlet two days apart, strongly suggesting that they belonged to different social groups. The data were offloaded and burned to CD in duplicate in the field.

D. Vocalization extraction

The 14.68 h of recordings were audited by listening to and visually examining the spectrograms in 15 s segments. Focal (tagged animal) vocalizations were marked according to their starting time and vocal category. It was assumed that vocalizations with a relatively high signal to noise ratio (SNR) belonged to the focal animal and not a neighboring nonfocal animal. Although this assumption could not be verified visually because the tagged narwhals were not followed, it was true for at least the first few dives since no group members were observed in the immediate vicinity. Much softer sounds were often heard on the recordings, presumably from more distant, vocalizing nonfocal animals. The SNR of these focal vocalizations was computed by comparing the root of the mean of the squared pressure (RMS) along the window containing 90% of the signal energy to a segment of noise of the same duration immediately preceding the signal (Madsen, 2005). The analysis presented here only excluded echolocation clicks, or broadband pulses of energy with interclick intervals usually greater than 100 ms. Nearly all remaining vocalizations were considered that could be divided into the two discrete categories of (1) combined tonal/pulsed signals (Fig. 2), defined as uninterrupted pulsatile vocalizations with a synchronously produced FM tonal component, a low mean interpulse interval (IPI $<$ 13 ms), and high pulse number ($>$49), and (2) whistles (Fig. 3), which were characterized by FM, tonal energy with several harmonics. See the discussion in this manuscript for a comparison of these designations to earlier classification schemata. All vocalizations were saved as separate wav files. Individual mm224 produced 42 combined tonal/pulsed signals and 4 whistles and mm226 produced 31 combined tonal/pulsed signals and 17 whistles.

E. Analysis of combined tonal/pulsed signals

Pulses were located automatically using customized Matlab 7.1 (The MathWorks, Inc.) software that, given a user-provided threshold value, triggered on and marked abrupt peaks in the pressure waveform. Subsequent inspections of all waveforms were made to select undetected and remove erroneously marked pulses. A nonparametric Wilcoxon rank sum test with a Bonferroni correction was used to examine whether the four parameters of average IPI, duration, number of pulses, and pulse repetition rate were significantly distinguishable between the two individuals. The pulse repetition rate and the normalized pulse number were also plotted as functions of the normalized duration to provide a visual means of comparing these sounds.

F. Whistle extraction and analysis

The fundamental frequency contour of each whistle spectrogram (FFT size and frame length of 2048 points with 50% window overlap) was traced by hand with customized Matlab software (Fig. 4). One hundred equally spaced points were extracted from these contours and normalized to a time axis between 0 and 1 (see Watwood *et al.*, 2004, 2005). Two tests of similarity were conducted on these whistle contours:

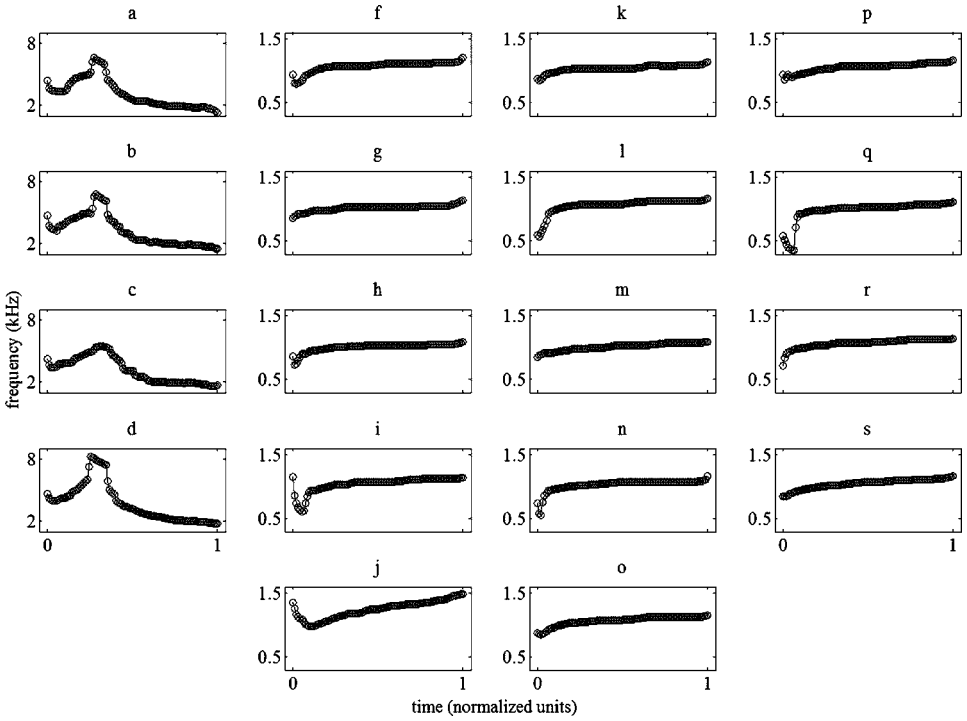


FIG. 4. Digitized traces of the fundamental frequency contours of the whistles displayed in Fig. 3. Each trace is shown with 100 equally spaced points that have been normalized on a horizontal time axis from 0 to 1. Again, panels a–d correspond to the whistles produced by mm224 and f–s to those by mm226. Note the difference in the frequency ranges for the two individuals.

1. Nonparametric comparison

One temporal (original duration before normalization) and five spectral (minimum, maximum, mean, initial, and ending frequencies) features were determined for every whistle (Fig. 5). These parameters were selected because they summarized the timing and coarse frequency content of the whistles. The differences between the finer aspects of the frequency contours were reserved for the cross-correlation comparison. Again, a nonparametric Wilcoxon rank sum test with a Bonferroni correction was used to compare these features between individuals.

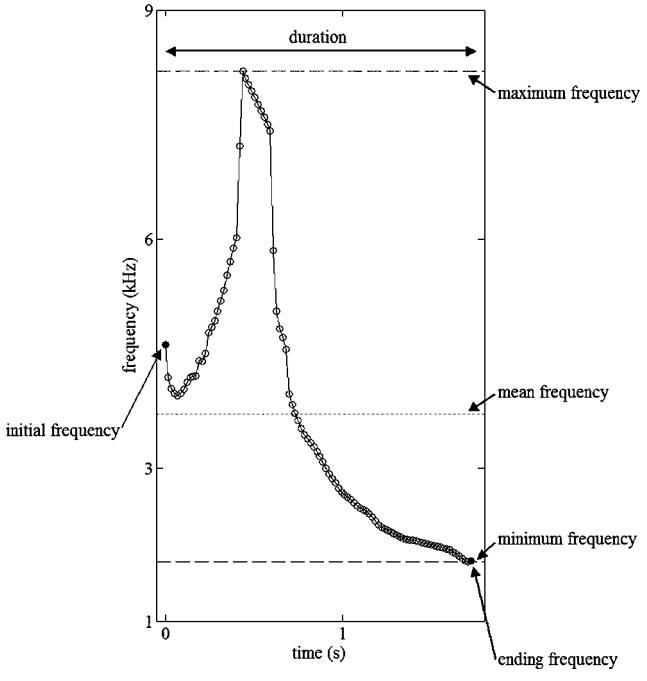


FIG. 5. Illustration of temporal and spectral features extracted from a traced whistle produced by mm224 (subplot d in Figs. 3 and 4). The initial and ending frequencies are indicated by filled circles (●) while the other spectral components are marked with horizontal dashed lines. Note that the mean frequency is closer to the minimum frequency because ~60% of the whistle's frequency content lies below 3700 Hz.

2. Cross-correlation comparison

Cross-correlation is often used in signal processing as a tool for determining the similarity between two signals. Because the frequency ranges of the whistles from the two individuals were distinct (e.g., the average maximum frequency of mm226 was less than the average minimum frequency of mm224), the cross-correlation measurement used here was designed to account for overall contour shape and absolute frequency content. In the equation

$$\sum_{i=1}^{100} \frac{|(f_A(i) - f_B(i))(f_A(i) - f_{B_m}(i))|}{\overline{f_A(i) + f_B(i)}}, \quad (1)$$

i is the sample number that ranges between 1 and 100, $f_A(i)$ and $f_B(i)$ correspond to the ith frequency value of contours A and B, respectively, and $f_{B_m}(i)$ is the ith frequency value of contour B after it has been slid along the frequency axis to minimize the frequency differences between contours A and B. These terms are illustrated in Fig. 6. Larger values of this cross-correlation measure indicated greater differences between contours than smaller values. A value of 0 would reveal no difference at all in contour shape. A constrained, nonlinear minimization routine was used to determine $f_{B_m}(i)$. The first term in the product of the numerator of (1) is the difference between points along the actual contours normalized in time [Fig. 6(a)]. The second term, however, returns a smaller number if the frequency modulation pattern is similar between the whistles regardless of the absolute frequency offset of the two [Fig. 6(b)]. Whistles therefore could have achieved a higher similarity

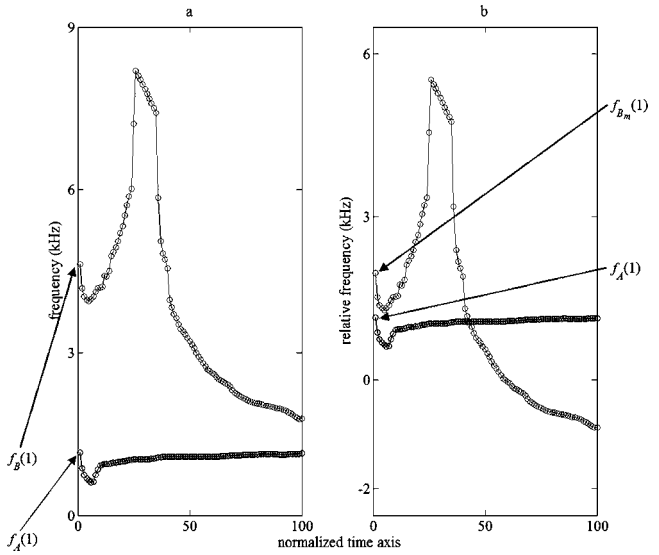


FIG. 6. Illustration of points used for cross-correlation comparison of whistles (see text for the equation). In (a), contour *A* (darker, from mm226: subplot i in Figs. 3 and 4) and *B* (lighter, from mm224: subplot d in Figs. 3 and 4) are depicted normalized in time with their original frequency content. In (b), contour *B* has been shifted along the frequency axis to minimize the frequency difference between the two contours. All 100 points along the contours were used to compute Eq. (1).

ranking [a smaller value of (1)] by overlapping in absolute frequency, possessing similar overall contour shapes or both.

III. RESULTS

Tables I and II list the summary measurements of the combined tonal/pulsed signals and whistles. With the exception of a single whistle assigned to mm226 with a SNR of 13.9 dB, the remaining vocalizations produced by both animals were characterized by a SNR of at least 28.2 dB. Combined tonal/pulsed signals were produced throughout the water column but tended to concentrate at particular depths (roughly 70 m for mm224 and 20 m for mm226, Fig. 7). Whistle production occurred between 20 and 100 m for mm224 but was confined to the upper 30 m for mm226 (Fig. 7). Both vocal categories were recorded throughout the diving sequence, indicating that the behavioral or environmental contexts in which these vocalizations occurred were not generally restricted to a very narrow depth or time. The two animals responded differently immediately after handling. Many combined pulsed/tonal signals (17 of 42) and one whistle were produced by individual mm224 just after release on his first dive lasting only 10.8 min. Individual mm226 was vocally active, however, between hours 4 and 10 of the deployment where he reached a maximum depth of about 125 m (data not shown). He did not produce any combined tonal/pulsed signals or whistles for the first 24 dives that exceeded roughly 10 m following his release, a response more closely resembling the silent reaction observed and discussed by Finley *et al.* (1990) of narwhals exposed to envi-

TABLE I. Summary statistics of the acoustic features of combined tonal/pulsed signals.

	Mean	Std.	Min.	Max.
mm224, $n=42$				
Duration (s)	1.6	0.7	0.6	2.7
Average IPI (ms)	12.9	4.3	9.0	36.1
Number of pulses	128.1	45.4	49.0	202.0
Pulse repetition rate (pulses/s)	82.3	14.2	28.1	112.8
mm226, $n=31$				
Duration (s)	1.2	0.1	1.0	1.2
Average IPI (ms)	6.3	0.3	5.6	6.8
Number of pulses	186.2	9.6	156.0	201.0
Pulse repetition rate (pulses/s)	160.8	7.4	147.5	180.5

TABLE II. Summary statistics of the acoustic features of whistles.

	Mean	Std.	Min.	Max.
mm224, $n=4$				
Duration (s)	1.19	0.08	1.09	1.26
Minimum frequency (Hz)	1549	201	1292	1775
Maximum frequency (Hz)	7181	1386	5460	8844
Mean frequency (Hz)	3638	370	3405	4190
Initial frequency (Hz)	4773	277	4496	5145
Ending frequency (Hz)	1572	216	1292	1788
mm226, $n=17$				
Duration (s)	0.78	0.04	0.68	0.85
Minimum frequency (Hz)	718	156	360	980
Maximum frequency (Hz)	1177	111	1095	1501
Mean frequency (Hz)	1012	71	895	1240
Initial frequency (Hz)	939	236	641	1501
Ending frequency (Hz)	1160	90	1095	1486

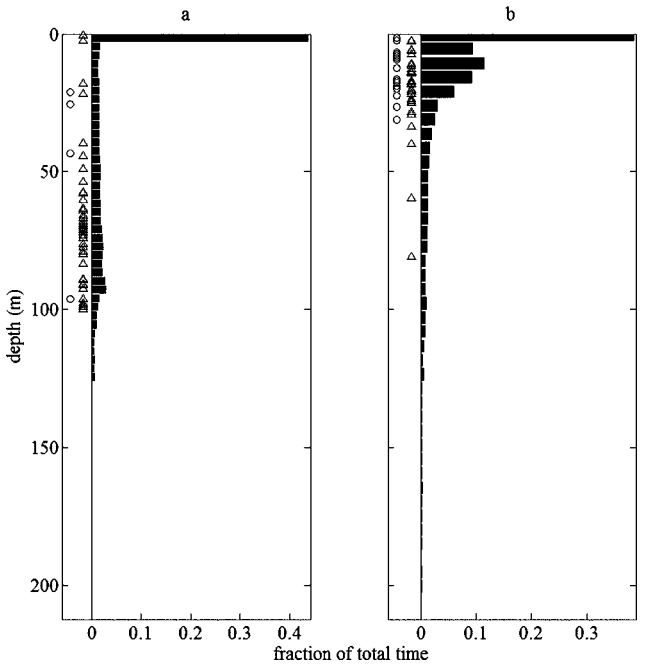


FIG. 7. Approximate depths where combined tonal/pulsed signals (triangles, △) and whistles (circles, ○) were produced adjacent to a frequency histogram of depth bins (bars) for mm224 (a) and mm226 (b). The frequency plotted on the abscissa is expressed as a fraction of the total amount of time spent at all depths. The maximum depths achieved for mm224 and mm226 during the DTAG deployments were roughly 125 and 210 m, respectively.

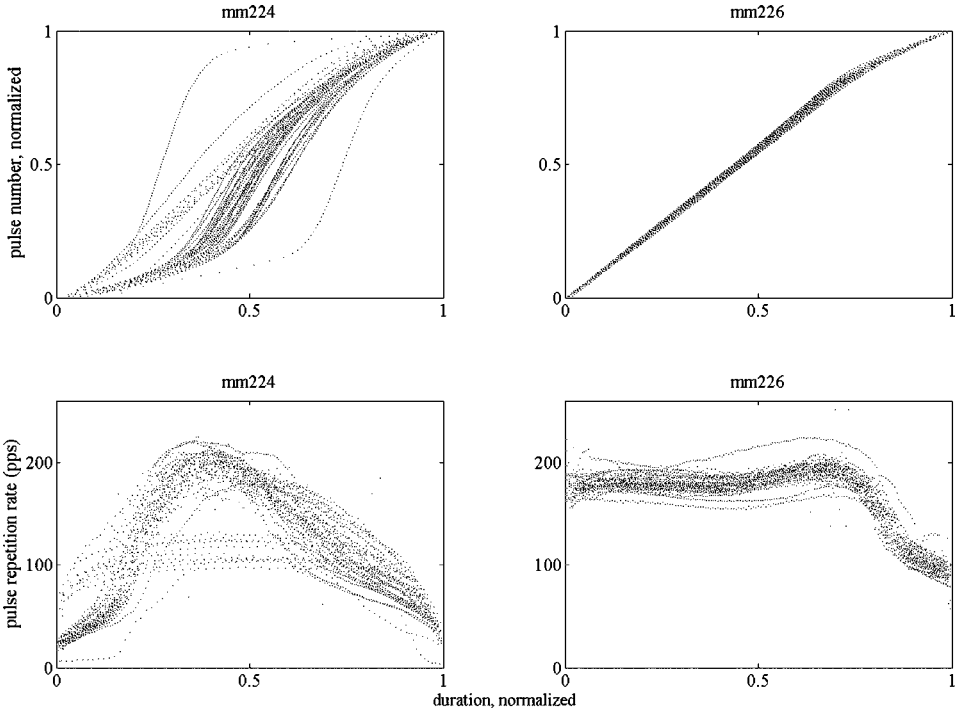


FIG. 8. Visual representations of temporal features of combined tonal/pulsed signals. Normalized pulse number (top two panels) and pulse repetition rate (pulses per second, bottom two panels) as a function of normalized duration. Note the clear differences in general morphology of these plots between the two individuals.

ronmental disturbances. Whistles were less common than combined tonal/pulsed signals, as reported in earlier studies (Ford and Fisher, 1978; Miller *et al.*, 1995).

The combined tonal/pulsed signals lasted between 0.55 and 2.68 s and contained between 49 and 202 pulses. Spectrograms revealed the synchronous production of both pulsatile energy in the form of repeated broadband impulses and a tonal, FM component by the tagged animals (Fig. 2). The FM component was not an analytical artifact of the pulsatile energy (see Watkins, 1967) because the fundamental frequency of the tonal feature was inconsistent with the repetition rate of the pulses. It is likely that at least two sound generating apparatuses are required to produce these combined tonal/pulsed signals to achieve pulsatile and FM energy content simultaneously. Combined tonal/pulsed signals were characterized by pulse rates between 28 and 113 pulses/s for mm224 and between 148 and 180 pulses/s for mm226. Figure 2 reveals additional low frequency energy associated with each pulse, which is likely the consequence of resonance of the air sacs of the tagged animal and not reverberations from or echoes off of elements along the inlet bottom. Indeed, no echoes consistent with target localization or monitoring position in the water column were detected in the audio record. All whistles were between 0.68 and 1.26 s with frequencies ranging between 360 and 8844 Hz. Three of the four whistles produced by mm224 were characterized by a brief (0.18–0.24 s) upsweep, followed by a pause and longer (0.62–0.66 s) downsweep [Figs. 3(a), 3(b), and 3(d)]. The fourth whistle was continuous but still showed an upsweep preceding the downsweep [Fig. 3(c)]. All seventeen whistles produced by mm226 contained a brief (0.026–0.091 s) broadband segment with energy that peaked between about 500 and 700 Hz and then decayed steadily until disappearing above 8–10 kHz followed by a flat, constant frequency tone that lasted for the remainder of the whistle [Figs. 3(f)–3(s), 3 whistles are not shown]. The whistles were all of about the same intensity except for a quieter one recorded on the tag attached to mm226 [Fig. 3(j)], which may have been softer, may have been produced by a nonfocal animal located further from the tag, or may have had a transmission path that was partially obscured by the animal or tag components before reaching the hydrophone.

The uniqueness of each of the two sets of combined tonal/pulsed signals and whistles was apparent from simple visual inspection. Among the combined tonal/pulsed signals the patterns of how the relative timing and repetition rate of the pulses varied as a function of normalized duration differed between the two individuals (Fig. 8). All measured features for both the combined tonal/pulsed signals and whistles differed significantly between the two animals (Wilcoxon rank sum test with a Bonferroni correction, $P=0.002$ for combined tonal/pulsed signals duration, $P<0.001$ for remaining combined tonal/pulsed signals measurements, $P=0.011$ for whistle ending frequency and $P=0.008$ for remaining whistle measurements). The whistles produced by mm224 were longer and higher in every measurement compared to those belonging to mm226, which were shorter and lower. Indeed, the minimum frequency of mm224 was 1549 ± 201 Hz ($\bar{x}\pm$sd) and the maximum frequency of mm226 was 1177 ± 111 Hz ($\bar{x}\pm$sd), values that did not overlap even a single standard deviation away (Table II). The cross-correlation test on the whistles revealed dramatic differences for the interindividual comparisons (between mm224 and mm226) and only slight differences among the intraindividual comparisons (Table III). The interindividual results were more different than the intraindividual results by 1–2 orders of magnitude.

TABLE III. Cross-correlation comparison of whistles between the same and different individuals. These data were computed in arbitrary units with higher values indicating a greater difference between the contours being compared. The intraindividual comparisons are italicized.

	mm226	mm224
mm224	127 250	*5 348*
mm226	*716*	

IV. DISCUSSION AND CONCLUSIONS

Two free-ranging narwhals each produced an acoustically distinctive set of combined tonal/pulsed signals and whistles. Visual and aural inspection and nonparametric and cross-correlation analyses all demonstrated striking interindividual differences among these vocalizations and intraindividual temporal and spectral fidelity. These results support the claim of Ford and Fisher (1978) that narwhals produce individually distinctive signature vocalizations. The recordings analyzed here also contained numerous faint combined tonal/pulsed signals and whistles produced by nonfocal animals. These observations are consistent with the conclusion that these vocal categories are regularly produced by free-ranging narwhals in this area.

The function of these vocalizations remains uncertain, but they do not appear to facilitate foraging. When feeding, some odontocetes produce a sequence of regularly spaced echolocation clicks that precede a buzz, or a series of clicks characterized by a dramatically elevated repetition rate [e.g., sperm whales (*Physeter macrocephalus*): Miller *et al.*, 2004, Blainville's beaked whales (*Mesoplodon densirostris*): Madsen *et al.*, 2005]. It seems unlikely that the narwhal vocalizations quantified here were used for foraging purposes since no echolocation clicks were detected immediately before the combined tonal/pulsed signals or the whistles. Indeed, the kind of clicking behavior characterized by changes in repetition rate and amplitude that is associated with foraging has been recorded from narwhals in previous studies (see Møhl *et al.*, 1990; Miller *et al.*, 1995), but was not observed here.

Combined tonal/pulsed signals and whistles may play a role in social communication based on their stereotypy (Ford and Fisher, 1978) and the signature content shown by the limited dataset presented here. These distinctive vocalizations might serve as contact calls to facilitate reunions of individuals with their group members in a manner similar to that observed in captive and free-ranging bottlenose dolphins (Janik and Slater, 1998; Watwood *et al.*, 2005). Unlike the studies conducted with bottlenose dolphins to identify pair bonds or alliances among males (Connor *et al.*, 1992, 2001; reviewed in Wells, 2003), little work has been completed to describe the social structure and group relationships among individual narwhals. Based on personal observations, the narwhals entering Admiralty Inlet traveled in groups ranging in size from roughly 5 to 30 animals. The group members traveling with the tagged animals vacated the area while their companions were detained on shore. The vocalizations of more distant animals that were recorded in this study were usually faint, suggesting that for the tagged animals, the dives occurring after their capture were likely solitary events. In addition, no other animals appeared in accompanying video footage recorded from a Crittercam (with a visual range extending between 3 and 20 m depending on the light level) attached to mm226 during the first hour following its release. These observations are consistent with the hypothesis that the combined tonal/pulsed signals and/or whistles were used by these two narwhals as contact calls in an effort to regain contact with their groups. However, actual reunions with other animals were not obvious from the acoustic record.

Because these two animals were likely members of different groups traveling into Admiralty Inlet, an alternative explanation consistent with the results presented here is that these vocalization classes may have been distinctive at the level of the social group (see Terhune *et al.*, 2001; Weiß *et al.*, 2006). Another possibility is that combined tonal/pulsed signals and whistles are actually used as signature vocalizations to cue conspecifics about individual identity. The whistle of mm226 was characterized by nearly constant frequency except for the brief noisy segment at the very beginning. From an information theory perspective, a flat whistle encodes less information compared to a frequency modulated whistle. It is possible, however, that even flat whistles of consistently distinctive durations or pitches could be used to distinguish between individuals. The contours of the whistles of mm226 appeared very similar to the FM component of the combined tonal/pulsed signals of this animal. The whistles produced by mm224, however, contained the frequency modulation expected of signature vocalizations and did not resemble this animal's combined tonal/pulsed signals FM component.

The combined tonal/pulsed signals described here most closely resembled the longer click series and the pulsed sounds described by Watkins *et al.* (1971) and Ford and Fisher (1978), respectively. All of these vocalizations were characterized by a combination of pulses and a tonal signal. Watkins *et al.* (1971) described the repetition rate of their longer click series tending to increase before becoming constant and eventually slowing down, somewhat similar to the trend observed in the combined tonal/pulsed signals described here (Fig. 8). In both the combined tonal/pulsed signals recorded in this study and their equivalents described in Watkins *et al.* (1971) and Ford and Fisher (1978), the synchronously-produced FM component creating the tonal quality in these sounds was not due exclusively to harmonically-related sidebands of the repetition rate (see Watkins, 1967). Note in Fig. 2 that the pitch of the FM component does not always correspond to the repetition rate of the pulses. For the combined tonal/pulsed signals produced by mm226, for example, the FM component begins over halfway through the signal without any observable change in repetition rate. Also, the tonal energy persists even as the pulses slow down at the end of the combined tonal/pulsed signals attributed to individual mm224 [Fig. 2(a)] and continues beyond the conclusion of the pulses in the signal assigned to mm226 [Fig. 2(b)]. The pulsatile component of

the combined tonal/pulsed signals quantified here was characterized by a higher upper frequency limit (up to 48 kHz) compared to earlier recordings.

Previous studies (Watkins *et al.*, 1971; Ford and Fisher, 1978) reported observing tonal signals with properties that were both similar to and different from the whistles analyzed here. The frequency ranges overlapped but the whistles that were recorded here had higher harmonics, extending the upper bound of these tonal vocalizations to at least 48 kHz. Earlier published tones were either constant in frequency or swept upwards or downwards, again consistent with the whistles presented here (Fig. 3). Both earlier works, however, described whistles as narrow-band signals lacking additional detectable harmonic content. All of the whistles recorded for this study were typified by a fundamental FM component and harmonic energy, a difference which may have resulted from the higher sampling rate of the recording equipment and/or the elevated signal to noise ratio due to the close proximity of the hydrophone to the whale.

The cross-correlation test used in this analysis was modified slightly from those described in other studies of signature vocalizations (e.g., Buck and Tyack, 1993; McCowan, 1995; Janik, 1999; Watwood *et al.*, 2005). In general, similarity between whistle contours can result either coarsely from a general overlap in frequency range and/or more finely from comparable frequency modulation (e.g., loop number, overall shape). The time-invariant cross-correlation test used here incorporated both of these components into its final measurement. A continuum was possible ranging from dissimilar (minimal frequency overlap and contour resemblance) to very similar (maximal frequency overlap and contour resemblance). Because the first term of the product ranked similarity according to both overlap and contour and the second term according to contour only, intermediate scores of similarity were also possible. This was particularly important since the frequency ranges of the whistles from the two narwhals were mostly nonoverlapping. The possibility of contour shape resemblance was excluded by the unambiguous results of the cross-correlation test (Table III).

These findings suggest possible directions for future work. Tagging and recording the combined tonal/pulsed signals and whistles of multiple narwhals from other groups would provide data that could support or reject the conclusions made here. If these signals do possess signature content, further study could ascertain whether they are distinctive at the individual or group level. Critical to determining the function of these vocalizations will be an assessment of the behavioral context in which these vocalizations are produced and the stability of group composition and size over short and long time scales. If narwhals are capable of differentiating between individuals acoustically, quantifying these aspects of group dynamics would provide starting estimates for the number of animals with which a single individual is interacting and therefore between which it should be able to distinguish. Playback experiments would be useful for identifying the temporal and spectral features of the combined tonal/pulsed signals and whistles that the animals may be using to facilitate differentiation. An understanding of the ontogeny of these sounds to determine if vocal learning plays any role in their acquisition or development requires acoustic data from the same animals and their groups collected longitudinally over many years.

Signature whistles appear to be used by bottlenose dolphins as contact calls in a variety of contexts (Janik and Slater, 1998; Sayigh *et al.*, 1998; Watwood *et al.*, 2005). If narwhals, another gregarious odontocete, similarly use their combined tonal/pulsed signals and whistles as contact calls when separated from conspecifics, the procedure described here affords an opportunity to make recordings in this context while the animals are detained ashore. Under this hypothesis, an involuntary separation of the sort imposed here would cause the animals and/or their group members to vocalize in an effort to regain contact.

Despite the small sample size, the data presented in this manuscript provide supportive evidence for at least two classes of signature vocalizations among free-ranging narwhals at the individual or group level. Future work focused on the ontogeny, function, and acoustic characteristics of the combined tonal/pulsed signals and whistles produced by narwhals is required to develop an improved understanding of the vocal and social behavior of this elusive Arctic animal.

ACKNOWLEDGMENTS

This manuscript benefited from the insights, thorough revisions, and helpful suggestions provided by Peter Tyack. Special thanks to Mark Johnson, Kira Barton, Alex Bocconcelli, Amanda Hansen, Tom Hurst, and Alex Shorter for DTAG and field support and to Rune Dietz, Jack Orr, Mehdi Bakhtiari, Sila Akikuluk, Sandie Black, Moe Keenainak, Martin Nweeia, Hans-Christian Schmidt, Seemee Tunraq, and Keith Yip for their generous assistance in the field. I am grateful to Amanda Searby and Stephanie Watwood for comments on the contour comparisons and to two anonymous reviewers whose suggestions helped to improve this manuscript. I thank the WHOI Academic Programs office, the National Science Foundation Research Fellowship, and the National Defense Science and Engineering Graduate Fellowship for financial support. This field operation was funded by the Greenland Institute of Natural Resources, the National Environmental Research Institute, Department of Fisheries and Oceans, Nunavut Wildlife Management Board and the Danish Cooperation for the Environment in the Arctic (DANCEA). Additional logistical support was provided by the Polar Continental Shelf Project. This work was conducted under permit No. SLE-04/05-9 issued by the Canadian Department of Fisheries and Oceans and animal use protocol FWI2004-2005-10.

Balcombe, J. P. (**1990**). "Vocal recognition of pups by mother Mexican free-tailed bats, *Tadarida brasiliensis mexicana*," Anim. Behav. **39**, 960–966.

Bee, M. A., and Gerhardt, H.C. (**2002**). "Individual voice recognition in a territorial frog (*Rana catesbeiana*)," Proc. R. Soc. London, Ser. B **269**, 1443–1448.

Boughman, J. W. (**1997**). "Greater spear-nosed bats give group-distinctive calls," Behav. Ecol. Sociobiol. **40**, 61–70.

Buck, J. R., and Tyack, P.L. (**1993**). "A quantitative measure of similarity for *Tursiops truncatus* signature whistles," J. Acoust. Soc. Am. **94**, 2497–2506.

Caldwell, M. C., and Caldwell, D.K. (**1965**). "Individualized whistle contours in bottlenose dolphins (*Tursiops truncatus*)," Nature (London) **207**, 434–435.
Charrier, I., Mathevon, N., and Jouventin, P. (**2001**). "Mother's voice recognition by seal pups," Nature (London) **412**, 873.
Charrier, I., Mathevon, N., and Jouventin, P. (**2003**). "Vocal signature recognition of mothers by fur seal pups," Anim. Behav. **65**, 543–550.
Connor, R. C., Heithaus, M. R., and Barre, L. M. (**2001**). "Complex social structure, alliance stability and mating access in a bottlenose dolphin 'super-alliance'," Proc. R. Soc. London, Ser. B **268**, 263–267.
Connor, R. C., Smolker, R. A., and Richards, A. F. (**1992**). "Two levels of alliance formation among male bottlenose dolphins (*Tursiops* sp.)," Proc. Natl. Acad. Sci. U.S.A. **89**, 987–990.
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**, 629–638.
Dietz, R., and Heide-Jørgensen, M. P. (**1995**). "Movements and swimming speed of narwhals *Monodon monoceros*, equipped with satellite transmitters in Melville Bay, northwest Greenland," Can. J. Zool. **73**, 2106–2119.
Dietz, R., Heide-Jørgensen, M. P., Richard, P. R., and Acquarone, M. (**2001**). "Summer and fall movements of narwhals (Monodon monoceros) from northeastern Baffin Island towards northern Davis Strait," Arctic **54**, 244–261.
Egnor, S. E. R., andHauser, M. D. (**2004**). "A paradox in the evolution of primate vocal learning," Trends Neurosci. **27**, 649–654.
Elowson, A. M., andSnowdon, C. T. (**1994**). "Pygmy marmosets, *Cebuella pygmaea*, modify vocal structure in response to changed social environment," Anim. Behav. **47**, 1267–1277.
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. Comput. Phys. **108**, 81–92.
Finley, K. J., Miller, G. W., Davis, R. A., and Greene, C. R. (**1990**). "Reactions of belugas, *Delphinapterus leucas*, and narwhals, *Monodon monoceros*, to ice-breaking ships in the Canadian high Arctic," Can. Bull. Fish. Aquat. Sci. **224**, 97–117
Ford, J. K. B., and Fisher, H. D. (**1978**). "Underwater acoustic signals of the narwhal (*Monodon monoceros*)," Can. J. Zool. **56**, 552–560.
Fripp, D., Owen, C., Quintana-Rizzo, E., Shapiro, A., Buckstaff, K., Jankowski, K., Wells, R., and Tyack, P. (**2005**). "Bottlenose dolphin (*Tursiops truncatus*) calves appear to model their signature whistles on the signature whistles of community members," Anim. Cogn. **8**, 17–26.
Frommolt, K. H., Goltsman, M. E., and MacDonald, D. W. (**2003**). "Barking foxes, *Alopex lagopus*: Field experiments in individual recognition in a territorial mammal," Anim. Behav. **65**, 509–518.
Goldman, J. A., Phillips, D. P., and Fentress, J. C. (**1995**). "An acoustic basis for maternal recognition in timber wolves (*Canis lupus*)," J. Acoust. Soc. Am. **97**, 1970–1973.
Hay, K. A., and Mansfield, A. W. (**1989**). "Narwhal: *Monodon monoceros* Linnaeus, 1758," in *Handbook of Marine Mammals: Volume 4, River Dolphins and the Larger Toothed Whales*, edited by S. H. Ridgway and R. Harrison (Academic, London), pp. 145–176.
Janik, V. M. (**1999**). "Pitfalls in the categorization of behaviour: a comparison of dolphin whistle classification methods," Anim. Behav. **57**, 133–143.
Janik, V. M. (**2000**). "Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*)," Science **289**, 1355–1357.
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.
Janik, V. M., and Slater, P. J. B. (**1997**). "Vocal learning in mammals," Adv. Study Behav. **26**, 59–99.
Janik, V. M., and Slater, P. J. B. (**1998**). "Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls," Anim. Behav. **56**, 829–838.
Janik, V. M., and Slater, P. J. B. (**2000**). "The different roles of social learning in vocal communication," Anim. Behav. **60**, 1–11.
Johnson, M. P. and Tyack, P. L. (**2003**). "A digital acoustic recording tag for measuring the response of wild marine mammals to sound," IEEE J. Ocean. Eng. **28**, 3–12.
Jones, B. S., Harris, D. H. R., and Catchpole, C. K. (**1993**). "The stability of the vocal signature in phee calls of the common marmoset, *Callithrix jacchus*," Am. J. Primatol **31**, 67–75.
Laidre, K. L., Heide-Jørgensen, M. P., Logdson, M. L., Hobbs, R. C., Heagerty, P., Dietz, R., Jørgensen, O. A., and Treble, M. A. (**2004**). "Seasonal narwhal habitat associations in the high Arctic," Mar. Biol. (Berlin) **145**, 821–831.
Lengagne, T., Aubin, T., Jouventin, P., and Lauga, J. (**2000**). "Perceptual salience of individually distinctive features in the calls of adult king penguins," J. Acoust. Soc. Am. **107**, 508–516.
Leonard, M. L., Horn, A. G., Brown, C. R., and Fernandez, N. J. (**1997**). "Parent-offspring recognition in tree swallows, *Tachycineta bicolor*," Anim. Behav. **54**, 1107–1116.
Madsen, P. T. (**2005**). "Marine mammals and noise: Problems with root mean square sound pressure levels for transients," J. Acoust. Soc. Am. **117**, 3952–3957.
Madsen, P. T., Johnson, M., de Soto, N. A., Zimmer, W. M. X., and Tyack, P. (**2005**). "Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*)," J. Evol. Biol. **208**, 181–194.
Marshall, A. J., Wrangham, R. W., and Arcadi, A. C. (**1999**). "Does learning affect the structure of vocalizations in chimpanzees?," Anim. Behav. **58**, 825–830.
Marshall, G. J. (**1998**). "Crittercam: An animal-borne imaging and data logging system," Mar. Technol. Soc. J. **32**, 11–17.
McCowan, B. (**1995**). "A new quantitative technique for categorizing whistles using simulated signals and whistles from captive bottlenose dolphins (Delphinidae, *Tursiops truncatus*)," Ethology **100**, 177–193.
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**, 194–209.
McCowan, B., and Reiss, D. (**2001**). "The fallacy of 'signature whistles' in bottlenose dolphins: A comparative perspective of 'signature information' in animal vocalizations," Anim. Behav. **62**, 1151–1162.
Miksis, J. L., Tyack, P. L., and Buck, J. R. (**2002**). "Captive dolphins, *Tursiops truncatus*, develop signature whistles that match acoustic features of human-made model sounds," J. Acoust. Soc. Am. **112**, 728–739.
Miller, L. A., Pristed, J., Møhl, B., and Surlykke, A. (**1995**). "The click-sounds of narwhals (*Monodon monoceros*) in Inglefield Bay, Northwest Greenland," Marine Mammal Sci. **11**, 491–502.
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., Johnson, M. P., 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.
Mitani, J. C., and Gros-Louis, J. (**1998**). "Chorusing and call convergence in chimpanzees: Tests of three hypotheses," Behaviour **135**, 1041–1064.
Mitani, J. C. and Nishida, T. (**1993**). "Contexts and social correlates of long distance calling by male chimpanzees," Anim. Behav. **45**, 735–746.
Møhl, B., Surlykke, A., and Miller, L. A. (**1990**). "High intensity narwhal clicks," in *Sensory Abilities of Cetaceans*, edited by J. Thomas and R. Kastelein (Plenum, New York), pp. 295–303.
Nousek, A. E., Slater, P. J. B., and Miller, P. J. O. (**in press**). "The influence of social affiliation on individual vocal signatures of northern resident killer whales (Orcinus orca)," Proc. R. Soc. London, Ser B.
Sayigh, L. S., Tyack, P. L., Wells, R. S., and Scott, M. D. (**1990**). "Signature whistles of free-ranging bottenose dolphins, Tursiops truncatus: Mother-offspring comparisons," Behav. Ecol. Sociobiol. **26**, 247–260.
Sayigh, L. S., Tyack, P. L., Wells, R. S., Solow, A. R., Scott, M. D., and Irvine, A. B. (**1998**). "Individual recognition in wild bottlenose dolphins: a field test using playback experiments," Anim. Behav. **57**, 41–50.
Scherrer, J. A., and Wilkinson, G. S. (**1993**). "Evening bat isolation calls provide evidence for heritable signatures," Anim. Behav. **46**, 847–860.
Searby, A., Jouventin, P., and Aubin, T. (**2004**). "Acoustic recognition in macaroni penguins: An original signature system," Anim. Behav. **67**, 615–625.
Smolker, R., and Pepper, J. W. (**1999**). "Whistle convergence among allied male bottlenose dolphins (Delphinidae, *Tursiops* sp.)," Ethology **105**, 595–617.
Snowdon, C. T., and Elowson, A. M. (**1999**). "Pygmy marmosets modify call structure when paired," Ethology **105**, 893–908.
Snowdon, C. T., Elowson, A. M., and Roush, R. S. (**1997**). "Social influences on vocal development in New World primates," in *Social Influences on Vocal Development*, edited by C. T. Snowdon and M. Hausberger (Cambridge University Press, Cambridge), pp. 234–248.
Stoddard, P. K., and Beecher, M. D. (**1983**). "Parental recognition of offspring in the cliff swallow," Auk **100**, 795–799.
Terhune, J. M., Healey, S. R., and Burton, H. R. (**2001**). "Easily measured

call attributes can detect vocal differences between Weddell seals from two areas," Bioacoustics 11, 211–222.
Tyack, P. (**1986**). "Whistle repertoires of two bottlenosed dolphins, *Tursiops truncatus*: Mimicry of signature whistles?" Behav. Ecol. Sociobiol. **18**, 251–257.
Watkins, W. A. (**1967**). "The harmonic interval: fact or artifact in spectral analysis of pulse trains," in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 15–43.
Watkins, W. A., Schevill, W. E., and Ray, C. (**1971**). "Underwater sounds of *Monodon* (Narwhal)," J. Acoust. Soc. Am. **49**, 595–599.
Watwood, S. L., Owen, E. C. G., Tyack, P. L., and Wells, R. S. (**2005**). "Signature whistle use by temporarily restrained and free-swimming bottlenose dolphins, *Tursiops truncatus*," Anim. Behav. **69**, 1373–1386.
Watwood, S. L., Tyack, P. L., and Wells, R. S. (**2004**). "Whistle sharing in paired male bottlenose dolphins, Tursiops truncatus," Behav. Ecol. Sociobiol. **55**, 531–543.
Weiß, B. M., Ladich, F., Spong, P., and Symonds, H. (**2006**). "Vocal behavior of resident killer whale matrilines with newborn calves: The role of family signatures," J. Acoust. Soc. Am. **119**, 627–635.
Wells, R. S. (**2003**). "Dolphin Social Complexity: Lessons from Long-Term Study and Life History," in *Animal Social Complexity: Intelligence, Culture, and Individualized Societies*, edited by F. B. M. de Waal and P. L. Tyack (Harvard University Press, Cambridge, Massachusetts), pp. 32–56.

Pitch cue learning in chinchillas: The role of spectral region in the training stimulus

William P. Shofner
Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue, Bloomington, Indiana 47405 and Parmly Hearing Institute, Loyola University Chicago, 6525 North Sheridan Road, Chicago, Illinois 60626

William M. Whitmer[a)]
Parmly Hearing Institute, Loyola University Chicago, 6525 North Sheridan Road, Chicago, Illinois 60626

(Received 6 February 2006; revised 24 May 2006; accepted 21 June 2006)

Chinchillas were trained to discriminate a cosine-phase harmonic tone complex (COS) from wideband noise (WBN) and tested in a stimulus generalization paradigm with tone complexes in which phase differed between frequency regions. In this split-phase condition, responses to complexes made of random-phase low frequencies, cosine-phase high frequencies were similar to responses to the COS-training stimulus. However, responses to complexes made of cosine-phase low frequencies, random-phase high frequencies were generally lower than their responses to the COS-training stimulus. When tested with sine-phase (SIN) and random-phase (RND) tone complexes, responses were large for SIN, but were small for RND. Chinchillas were then trained to discriminate infinitely-iterated rippled noise (IIRN) from WBN and tested with noises in which the spectral ripple differed between frequency regions. In this split-spectrum condition, responses were large to noises made of rippled-spectrum low frequencies, flat-spectrum high frequencies, whereas responses were generally lower to noises made of flat-spectrum low frequencies, rippled-spectrum high frequencies. The results suggest that chinchillas listen across all frequencies, but attend to high frequencies when discriminating COS from WBN and attend to low frequencies when discriminating IIRN from WBN. [DOI: 10.1121/1.2225969]

PACS number(s): 43.80.Lb, 43.66.Gf, 43.66.Hg [JAS] Pages: 1706–1712

I. INTRODUCTION

Pitch is a fundamental perception, and a variety of perceptual attributes are associated with pitch (see Plack *et al.*, 2005). One important attribute of pitch perception is that similar pitches can be evoked by a variety of complex sounds that differ in their spectral and temporal characteristics (Fastl and Stoll, 1979). Terhardt (1974) proposed that there is a learning stage associated with pitch perception, presumably related in part to the learning of speech sounds. Divenyi (1979) has described anecdotal evidence in support of this learning model, but has also pointed out that testing the learning aspects of Terhardt's model directly would be difficult. For example, it would be virtually impossible to study normal-hearing listeners who have not been previously exposed to speech and music. One alternative pointed out by Divenyi would be to compare pitch attributes between hearing-impaired listeners with acquired hearing loss (having previous exposure to speech and music) to those with congenital hearing loss (having no previous exposure to speech and music). More recent studies have described pitch-cue learning that is specific to the training stimuli (e.g., Demany and Semal, 2002; Grimault *et al.*, 2003; Fitzgerald and Wright, 2005). These studies show that when human listeners are trained in a pitch discrimination task with a specific sound, the improvement in discrimination performance does not generalize to other sounds, particularly if those sounds have different spectral and temporal characteristics. For example, Fitzgerald and Wright (2005) specifically trained listeners to discriminate the rate of a 150 Hz sinusoidally amplitude modulated noise (SAM). Trained listeners showed a significant improvement in rate discrimination, but listeners trained for the rate discrimination of the 150 Hz SAM noise did not show significant improvement when placed in a 150 Hz pure tone discrimination task or a 150 Hz rippled noise pitch discrimination task. In other words, the learning that occurred for the SAM noise was specific for the pitch cues evoked by that stimulus and did not generalize to the pure tone or rippled noise stimuli.

We have recently described a similar learning effect for pitch saliency (i.e., pitch strength) in chinchillas. When animals are trained to discriminate a cosine-phase harmonic tone complex (COS) from wideband noise (WBN) and tested in a stimulus generalization paradigm, the dominant pitch-cue appears to be the periodicity strength in the stimulus envelope (Shofner, 2002). When animals are retrained in the task using an infinitely-iterated rippled noise (IIRN) as the training stimulus instead of the COS, then the dominant pitch-cue appears to the periodicity strength in the waveform fine structure (Shofner *et al.*, 2005b). The changes in the generalization gradients obtained under the two training conditions suggest that there is a change in listening strategy. That is, animals use the periodicity information in the stimu-

[a)]Present address: GN ReSound, 2601 Patriot Boulevard, Glenview, Illinois 60026.

lus envelope when trained with COS, but they can learn to use the periodicity information in the fine structure when trained with IIRN.

The large influence of the stimulus envelope when COS is used as the training stimulus suggests that animals are attending to high frequencies more so than low frequencies. In contrast, the large influence of the fine structure when IIRN is used as the training stimulus suggests that animals have changed the focus of their attention to low frequencies more so than high frequencies. The experiments described in the present study were designed to test directly whether chinchillas attend differently to low and high frequency regions depending on the training stimulus. In the present study, 3 kHz was chosen as the boundary between low and high frequency regions for two reasons. First, if a representation of fine structure is dependent on synchrony in the auditory nerve, then 3 kHz is an appropriate demarcation of phase-locking ability (e.g., Palmer and Russell, 1986). Second, for the 250-Hz fundamental frequency used in the present study, it would be expected that harmonics below 3 kHz are resolved, whereas those above 3 kHz are unresolved (e.g., Shofner *et al.*, 2005a).

II. METHODS

The methods, behavioral procedure, and equipment are the same as those used in our previous studies (see Shofner, 2002; Shofner *et al.*, 2005b for details).

Adult chinchillas (*Chinchilla laniger*) were generally maintained at 80%–90% of their normal body weight and received food pellet rewards during behavioral testing. Their diets were supplemented with chinchilla chow to maintain their body weights. The care and use of the chinchillas and the procedures employed were approved by the Institutional Animal Care and Use Committee of Loyola University Chicago.

Stimulus presentation and data acquisition were controlled by a Gateway computer system with Tucker-Davis Technologies System II modules. Stimuli used for behavioral training consisted of WBN, COS, and IIRN (see Shofner, 2002; Shofner *et al.*, 2005b for details). For this study, the harmonic complex tones consisted of a 250 Hz fundamental frequency (F0) and all higher harmonics up to and including 10 kHz. Individual components were of equal amplitude and either added in cosine-starting phase, sine-starting phase (SIN) or random-starting phase (RND). Harmonic tone complexes were generated on a digital array processor at a sampling rate of 50 kHz and stored on a disk. The delay used for generating the IIRN was fixed at 4 ms resulting in spectral peaks at integer multiples of 250 Hz, and the delayed-noise attenuation was fixed at −1 dB. The bandwidths of the WBN and IIRN were 10 kHz. For all noises used in the present study, 5 s of each waveform were sampled at 50 kHz and stored on disk. The root-mean-squared amplitudes of stimuli were scaled to be equal in order to eliminate the use of overall level as a cue during testing. Sound level was monitored by placing a condenser microphone (Ivie 1133) in the approximate position of an animal's head and measuring the A-weighted sound pressure level with a sound level meter (Ivie IE-30-A Audio Spectrum Analyzer). Overall sound level was fixed at 73 dB SPL for all stimuli.

Chinchillas were placed inside a cage in a single-walled, sound-attenuating animal test chamber lined with acoustic foam during a testing session. At one end of the cage was a pellet dispenser with a reward chute attached to a response lever, and a loudspeaker was placed near the pellet dispenser approximately 6 in. in the front of the animal at approximately 30° to the right of center.

Chinchillas were tested in a stimulus generalization procedure that was based on an operant conditioning paradigm using positive reinforcement (see Shofner, 2002; Shofner *et al.*, 2005b for details). Chinchillas discriminated a signal stimulus from a WBN-standard stimulus. Bursts of the WBN (500 ms with 10 ms rise/fall times) were presented continually once per second throughout the testing session regardless of whether the animal initiated a trial. The animal initiated a trial by pressing down on the response lever. When the lever was pressed down, the WBN was presented continually for an additional 1–8 bursts resulting in a holdtime of 1150–8150 ms for each trial. The number of WBN bursts presented after a trial was initiated varied randomly for each trial and was determined from a rectangular probability distribution. If the animal released the lever before the random holdtime, the countdown for the holdtime was halted; that trial began again with the next press of the lever using the same holdtime. If the animal held the response lever down for the duration of the holdtime, then either a signal interval or a blank interval (i.e., catch interval) occurred. A signal interval consisted of two bursts of the signal stimulus, while a blank interval consisted of two additional bursts of WBN. The response window was coincident with the duration of the signal/blank interval, but began 150 ms after the onset of the first burst and lasted until the onset of the next WBN burst. Thus, the duration of the response window was 1850 ms. Lever releases during the signal interval were treated as hits, whereas lever releases during a blank interval were treated as false alarms. If the animal continued to hold the lever down, during a blank interval, then that nonresponse was treated as a correct rejection. Animals received a food pellet reward for hits and correct rejections. In a block of 40 testing intervals, 60% were signal intervals, 20% were blank intervals. On each of the two sets of 4 intervals (10% of the intervals each) two different split-test stimuli were presented (see below). Animals received no food pellet rewards for behavioral responses to test stimuli, regardless of whether the response was a release of the lever or continuing to press the lever down. Data were collected for a minimum of 50 blocks, resulting in at least 200 trials for each test stimulus.

In experiment 1, chinchillas were trained to discriminate COS (i.e., signal) from WBN. The two test stimuli in this experiment were harmonic tone complexes with the same 40 frequency components, but with different phase characteristics in the lower and higher spectral regions (Fig. 1). We will refer to these test tone complexes as split-phase test stimuli. For one split-phase test stimulus, harmonic components 1–12 were added in cosine-starting phase, whereas harmonic com-

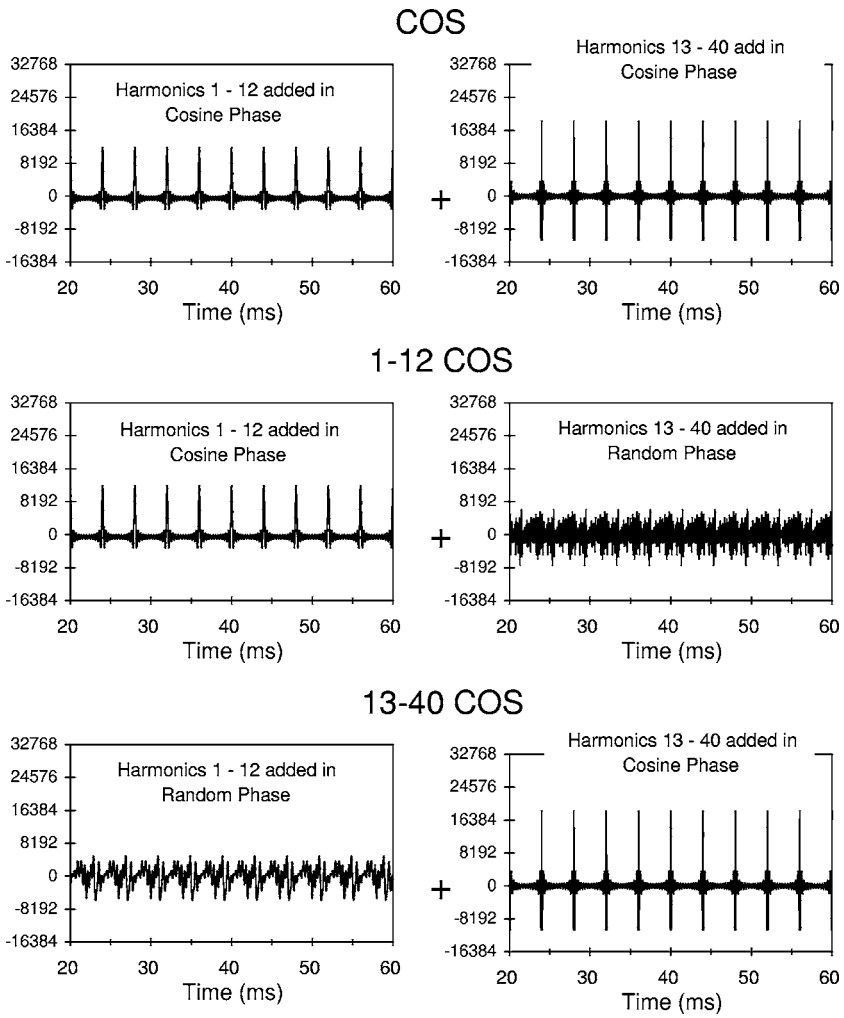


FIG. 1. Illustrations of the waveforms presented for training and test stimuli in experiment 1. The waveforms in the left- and right-hand panels show the low and high frequencies separately for conceptual purposes and are not meant to imply that two separate stimuli were presented. The time between 20–60 ms illustrates a window during the steady state portion of the stimulus. *Top panel* illustrates the COS training stimulus in which all 40 components are added in cosine-starting phase. *Middle panel* illustrates the test stimulus of 1–12 COS in which components 1–12 are added in cosine-starting phase and components 13–40 are added in random-starting phase. *Bottom panel* illustrates the test stimulus 13–40 COS in which components 1–12 are added in the random-starting phase and components 13–40 are added in the cosine-starting phase.

ponents 13–40 were added in random-starting phase. We will call this split-phase stimulus 1–12 COS (middle panels in Fig. 1). For the other split-phase test stimulus, harmonic components 1–12 were added in random-starting phase, whereas harmonic components 13—40 were added in cosine-starting phase. We will call this split-phase stimulus 13–40 COS (bottom panels in Fig. 1). If animals are attending to high frequencies when discriminating the cosine-phase tone complex from wideband noise, then we would predict that behavioral responses to the 13–40 COS test stimulus will be large, but responses to the 1–12 COS test stimulus will be reduced. As part of this experiment, chinchillas were also tested with SIN and RND harmonic tone complexes.

In experiment 2, chinchillas were trained to discriminate IIRN (i.e., signal) from WBN. The two test stimuli in this experiment were rippled noises that differed in the spectral regions of the ripples (Fig. 2). We will refer to these test rippled noises as split-spectrum test stimuli. One split-spectrum test stimulus was generated by adding together a low-pass filtered version of the infinitely iterated rippled noise to a high-pass filtered version of the wideband noise. We will call this spit-spectrum stimulus LP-IIRN (middle panel in Fig. 2). The other split-spectrum test stimulus was generated by adding together a low-pass filtered version of the wideband noise to a high-pass filtered version of the infinitely-iterated rippled noise. We will call this split-spectrum stimulus HP-IIRN (bottom panel in Fig. 2). The cutoff frequencies of the low-pass and high-pass filters were set at 3 kHz, and the slopes of the filter functions were 48 dB/octave. Note that the cutoff frequency of 3 kHz corresponds to the 12th harmonic in the tone complexes of experiment 1. If animals are attending to low frequencies when discriminating the infinitely-iterated rippled noise from wideband noise, then we would predict that behavioral responses to the LP-IIRN test stimulus will be large, but responses to the HP-IIRN test stimulus will be reduced.

III. RESULTS

Figure 3(a) shows the behavioral responses obtained from experiment 1 in which 8 chinchillas discriminated COS from WBN and were tested with the split-phase stimuli. Behavioral responses are shown as percent generalization which is the percentage of intervals the chinchilla released the lever during the response window for each stimulus presented. For each animal, the percent generalization to the COS-signal is high, whereas the percent generalization to the WBN-standard is low. The percent generalization to the 13–40 COS test stimulus was approximately equal to the COS response in 5/8 chinchillas. For 3/8 chinchillas (C29, C42, C43), the percent generalization to the 13–40 COS was well above the response to WBN, but was lower than the

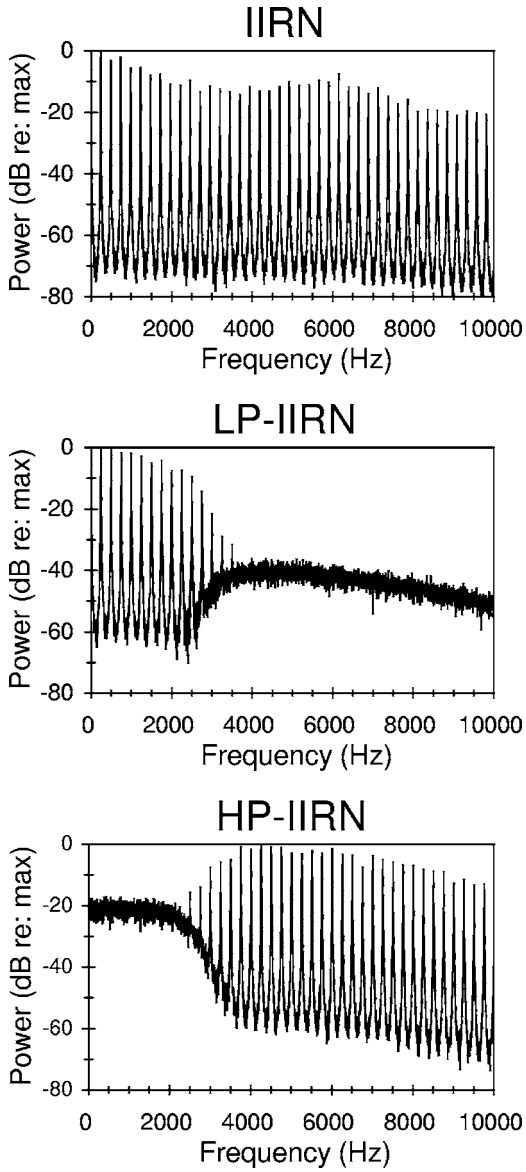


FIG. 2. (A) Example spectrum of the stimulus used for training the animals to discriminate IIRN from WBN. (B) Example spectrum of the test stimulus LP-IIRN. (C) Example spectrum of the test stimulus HP-IIRN.

response to COS. For all animals, the responses to 1–12 COS were below those obtained for 13–40 COS. Figure 3(b) shows the average and 95% confidence intervals determined using Tukey's standard error. A repeated-measures analysis of variance showed a significant effect of stimuli ($F=104.6$; $p \ll 0.0005$). Pairwise comparisons based on Tukey's test showed there was no significant difference between 13–40 COS and COS ($q=2.25$; $p>0.05$). There was a significant difference between 1–12 COS and COS ($q=12.55$; $p<0.001$) and between 1–12 COS and 13–40 COS ($q=10.30$; $p<0.001$).

Figure 3(c) shows the behavioral responses obtained from experiment 1 in which 7 chinchillas discriminated COS from WBN and were tested with SIN and RND. Again, for each animal, the percent generalization to the COS-signal is high, whereas the percent generalization to the WBN-standard is low. The percent generalization to the SIN-test stimulus was approximately equal to the COS response in all chinchillas. For all chinchillas, the percent generalization to the RND-test stimulus was above the response to WBN, but was much lower than the response to COS or SIN. Figure 3(d) shows the average and 95% confidence intervals determined using Tukey's standard error. A repeated-measures analysis of variance showed a significant effect of stimuli ($F=115.4$; $p \ll 0.0005$). Pairwise comparisons based on Tukey's test showed there was no significant difference between SIN and COS ($q=-0.08$; $p>0.05$), but there was a significant difference between RND and COS ($q=14.20$; $p<0.001$) and between RND and SIN ($q=14.28$; $p<0.001$).

Figure 3(e) shows the behavioral responses obtained from experiment 2 in which 7 chinchillas discriminated IIRN from WBN and were tested with the split-spectrum stimuli. For each animal, the percent generalization to the IIRN-signal is high, whereas the percent generalization to the WBN-standard is low. For all chinchillas, the percent generalization to the LP-IIRN test stimulus was large relative to the responses to IIRN. For 6/7 chinchillas, the percent generalization to the LP-IIRN test stimulus was above the response to the HP-IIRN test stimulus. For the remaining chinchilla (C16), the percent generalization to LP-IIRN was slightly below that to HP-IIRN. Figure 3(f) shows the average and 95% confidence intervals determined using Tukey's standard error. A repeated-measures analysis of variance showed a significant effect of stimuli ($F=57.0$; $p \ll 0.0005$). Pairwise comparisons based on Tukey's test showed there was no significant difference between LP-IIRN and IIRN ($q=3.47$; $p>0.05$), but there was a significant difference between HP-IIRN and IIRN ($q=9.62$; $p<0.001$) and between LP-IIRN and HP-IIRN ($q=6.15$; $p<0.005$).

IV. DISCUSSION

The present study used a stimulus generalization paradigm to examine whether chinchillas attend to different frequency regions during different pitch-cue discriminations. In our stimulus generalization paradigm, chinchillas are trained to discriminate a signal stimulus from a standard stimulus. One of the key elements of this procedure is that the signal is presented on most of the trials; food reward is given for correct responses, because these are considered to be objective responses. Consequently, it is likely that the chinchilla expects to hear the signal and receive a reward during most of the trials. Test stimuli were presented infrequently during the session, similar to the probe-signal method of Greenberg and Larkin (1968), but these responses were not rewarded, because they are considered to be subjective responses (i.e., they are neither correct nor incorrect). Behavioral responses to test stimuli that are similar to the responses to signal stimuli suggest a perceptual equivalence or perceptual invariance between the test and signal stimulus (Hulse, 1995). That is, there is some feature that is similar across these stimuli that leads to a similarity in the perception.

In human listeners, the COS tone complex used in the present study evokes a salient pitch perception, whereas the WBN only evokes a noise perception. This COS has a band-

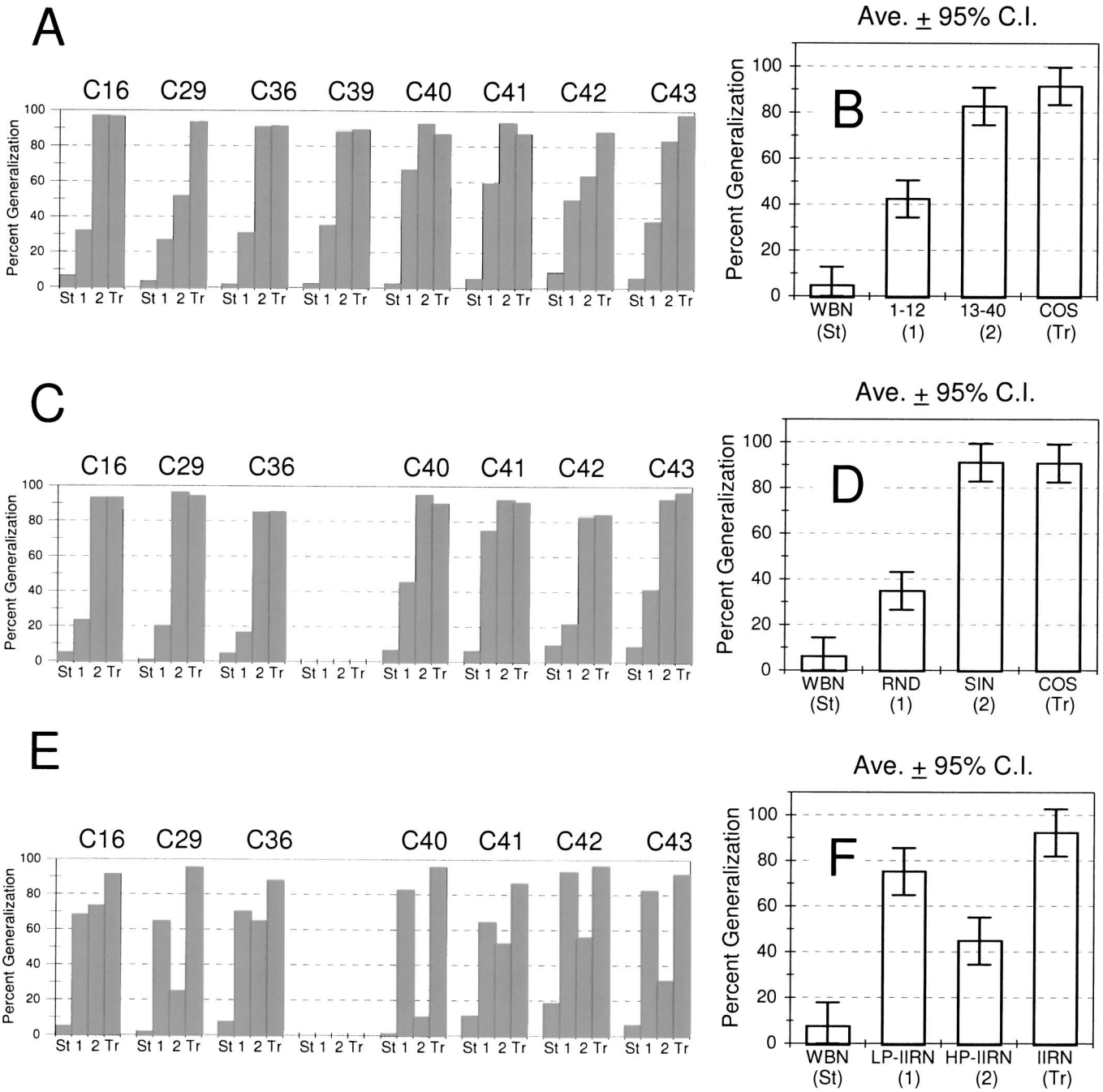


FIG. 3. Filled-bar graphs (A, C, E) show behavioral responses from individual animals when trained to discriminate the training stimulus (Tr) from the standard stimulus (St) and tested with stimuli 1 and 2. White-bar graphs (B, D, F) show average behavioral responses and the 95% confidence intervals based on Tukey's standard error for the individual data. A, B show data when animals are trained to discriminate COS (Tr) from WBN (St) and tested with 1–12 COS (1) and 13–40 COS (2). C, D show data when animals are trained to discriminate COS (Tr) from WBN (St) and tested with RND (1) and SIN (2). E, F show data when animals are trained to discriminate IIRN (Tr) from WBN (St) and tested with LP-IIRN (1) and HP-IIRN (2).

width that is similar to that of WBN and is comprised of both resolved and unresolved harmonics. For this particular tone complex with a 250 Hz fundamental frequency, human listeners judge the pitch strength or saliency of the RND tone complex to be only slightly less than that of the COS tone complex (Shofner and Selas, 2002). In contrast, the behavioral responses of the chinchillas to RND tone complexes were greatly reduced compared to the COS tone complex indicating that these two stimuli were not perceptually equivalent. However, the responses to the SIN tone complex were as large as those obtained to the COS indicating that these two stimuli are perceptually equivalent in the chinchilla. Both SIN and COS tone complexes have highly modulated waveforms; the autocorrelation functions of the envelopes extracted with a Hilbert transform are identical. However, the waveform of RND is much less modulated than either of the SIN or COS tone complexes. When discriminating COS from WBN, chinchillas attend to the features of the tone complex that distinguish it from the noise. Previous behavioral studies in chinchillas indicate that the most salient feature in this discrimination is the periodicity found in the stimulus envelope. (Shofner, 2002). That is, the

animal listens across all frequencies, but presumably attends to high frequency auditory filters where information about the envelope is dominant.

The results of the present study demonstrate that chinchillas do indeed attend to high frequencies more so than low frequencies when discriminating COS from WBN. When tested infrequently with split-phase tone complexes in which harmonics less than or equal to 3 kHz were added in random phase while harmonics above 3 kHz were added in cosine phase (13–40 COS), large behavioral responses were obtained. That is, tone complexes having identical phases in the high frequencies, but not in the low frequencies, were perceptually equivalent to the cosine-phase harmonic tone complex. In contrast, when tested infrequently with split-phase tone complexes in which harmonics above 3 kHz were added in random phase while harmonics less than or equal to 3 kHz were added in cosine phase (1–12 COS), small behavioral responses were obtained. In this case, tone complexes having identical phases in the low frequencies, but not in the high frequencies, were not perceptually equivalent to the cosine-phase harmonic tone complex. Thus, the perceptual equivalence of 13–40 COS to the cosine-phase tone complex can be attributed to the similar phase characteristics in the high frequency auditory filters (see Fig. 1). The 1–12 COS stimulus is not perceptually equivalent to the cosine-phase tone complex, because the phase characteristics are different in the high frequency filters between the two sounds.

When discriminating IIRN from WBN, chinchillas attend to the features of the rippled noise that distinguish it from wideband noise. In human listeners, this IIRN evokes both a pitch perception and a noise perception (see Patterson *et al.*, 1996). Previous behavioral studies in chinchillas indicate that the most salient feature in this discrimination is the periodicity found in the waveform fine structure. (Shofner *et al.*, 2005b). That is, the animal listens again across all frequencies, but presumably attends to low frequency auditory filters where information about the fine structure is dominant. The results of the present study demonstrate that chinchillas do indeed attend to low frequencies more so than high frequencies when discriminating IIRN from WBN. When tested infrequently with split-spectrum stimuli which had a flat-spectrum for frequencies equal to and below 3 kHz, but had a rippled-spectrum above 3 kHz (HP-IIRN), small behavioral responses were obtained. That is, split-spectrum noises having identical spectra in the high frequencies, but not in the low frequencies, were not perceptually equivalent to the IIRN-signal. In contrast, when tested infrequently with split-spectrum stimuli which had a flat-spectrum above 3 kHz, but a rippled-spectrum for frequencies equal to and below 3 kHz (LP-IIRN), large behavioral responses were obtained. That is, split-spectrum noises having identical spectra in the low frequencies, but not in the high frequencies, were perceptually equivalent to the IIRN-signal. Thus, the perceptual equivalence of LP-IIRN to the IIRN-signal can be attributed to the similar ripple characteristics in the low frequency auditory filters (see Fig. 2). The HP-IIRN stimulus is not perceptually equivalent to the IIRN-signal, because the ripple characteristics are different in the low frequency filters between the two sounds.

In summary, our previous studies showed that the perception of pitch strength in chinchillas is dependent on the type of stimulus used in the training of the animal. A COS-tone complex possesses a large periodicity at the reciprocal of the fundamental frequency in both the fine structure and envelope. When trained to discriminate COS from WBN, the cues for discrimination appear to be the periodicity in the envelope (Shofner, 2002). The results of the present study show that for this discrimination, chinchillas attend to high frequencies more so than low frequencies. In contrast, IIRN has a strong fine structure periodicity at the reciprocal of the delay, but a weak periodicity in the envelope. When trained to discriminate IIRN from WBN, the cues for discrimination appear to be the periodicity in the fine structure (Shofner *et al.*, 2005b). The results of the present study show that for this discrimination, chinchillas attend to low frequencies more so than high frequencies. Thus, there is a change in the ability of the chinchilla to use different pitch-related cues for the discrimination. That is, chinchillas normally attend to high frequency information, but can learn to attend to the low frequency information. This change in listening strategy from "high frequency listening" to "low frequency listening" reflects stimulus-specific, pitch-cue based discrimination learning.

Pitch-cue based discrimination that is specific to the training stimulus has been reported in human listeners (Demany and Semal, 2002; Grimault *et al.*, 2003; Fitzgerald and Wright, 2005) as well as goldfish (Fay, 2005). That is, discrimination based on pitch-cues specific to the training stimulus does not necessarily generalize to other stimuli, particularly if the pitch-cues of the test stimuli are different from those that were trained on. For example, the learning acquired for rate discrimination of SAM noise does not transfer to other nontraining stimuli (e.g., pure tone frequency discrimination or rippled noise delay discrimination), because the element of pitch processing affected by the learning acquired for the SAM rate discrimination is different from that for the nontraining stimuli (Fitzgerald and Wright, 2005). In the present context, chinchillas learn to discriminate COS (i.e., training stimulus) from WBN by analyzing high frequency auditory channels, but when tested with IIRNs (i.e., nontraining stimuli), the outputs of high frequency channels are more WBN-like than COS-like and thus, there is no generalization to the IIRN (Shofner, 2002). However, chinchillas can learn to discriminate IIRN from WBN by analyzing low frequency auditory channels. The outputs of low frequency channels for IIRN (i.e., training stimulus) are more similar to COS than to WBN, and therefore generalization is observed when animals are tested with COS (Shofner *et al.*, 2005b).

The neural substrate responsible for this pitch-cue learning in chinchillas is currently unknown. Learning effects are certainly known to occur in the auditory cortex (e.g., Recanzone *et al.*, 1993; Bao *et al.*, 2004; Polley *et al.*, 2006). For example, the topographic representation of a target frequency is expanded in the auditory cortex of rats trained in a frequency recognition task (Polley *et al.*, 2006). In addition to cortex, learning effects in lower auditory centers may also occur. Kettner and Thompson (1985) found physiological differences in unit responses in the medial geniculate body

and inferior colliculus of conditioned rabbits when the animal detected or failed to detect a white-noise stimulus, but found no physiological effect in cochlear nucleus unit responses. These findings suggest that learning effects may be observed in the thalamus and inferior colliculus, but not observed as far down as the cochlear nucleus. However, Woody *et al.* (1992; 1994) have shown that cochlear nucleus unit responses to clicks are larger from conditioned cats than from nonconditioned cats, suggesting that plasticity due to learning may also occur at the level of the cochlear nucleus. Neural synchrony as measured by the frequency following response to the fundamental frequency of Mandarin Chinese monosyllables is larger for Mandarin Chinese speakers than for native English speakers (Krishnan *et al.*, 2005), and these authors argue that synchrony in the brainstem to voice pitch is enhanced by language-dependent learning. It is possible that neural synchrony in the brainstem nuclei to iterated rippled noises (e.g., Shofner, 1999; Winter *et al.*, 2001; Griffiths *et al.*, 2001; Neuert *et al.*, 2005) is stronger in chinchillas following training with IIRN. Thus, the neural representations of IIRN in the chinchilla auditory cortex, medial geniculate, and perhaps even brainstem nuclei are presumably enhanced in some unknown manner following training with the IIRN stimulus.

ACKNOWLEDGMENTS

The authors thank Jennifer Lentz for helpful suggestions regarding data presentation. Supported by NIDCD Grant No. R01 DC005596.

Bao, S., Chang, E. F., Woods, J., and Merzenich, M. M. (**2004**). "Temporal plasticity in the primary auditory cortex induced by operant perceptual learning," Nat. Neurosci. **7**, 974–981.

Demany, L., and Semal, C. (**2002**). "Learning to perceive pitch differences," J. Acoust. Soc. Am. **111**, 1377–1388.

Divenyi, P. L. (**1979**). "Is pitch a learned attribute of sounds? Two points in support of Terhardt's pitch theory," J. Acoust. Soc. Am. **66**, 1210–1213.

Fastl, H., and Stoll, G. (**1979**). "Scaling of pitch strength," Hear. Res. **1**, 293–301.

Fay, R. R. (**2005**). "Perception of pitch by goldfish," Hear. Res. **205**, 7–20.

Fitzgerald, M. B., and Wright, B. A. (**2005**). "A perceptual learning investigation of the pitch elicited by amplitude-modulated noise," J. Acoust. Soc. Am. **118**, 3794–3803.

Greenberg, G. Z., and Larkin, W. D. (**1968**). "Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: The probe-signal method," J. Acoust. Soc. Am. **44**, 1513–1523.

Griffiths, T. D., Uppenkamp, S., Johnrude, I., Josephs, O., and Patterson, R. D. (**2001**). "Encoding of temporal regularity of sound in the human brainstem," Nat. Neurosci. **4**, 633–637.

Grimault, N., Micheyl, C., Carlyon, R. P., Bacon, S. P., and Collet, L. (**2003**). "Learning in discrimination of frequency or modulation rate: generalization to fundamental frequency discrimination," Hear. Res. **184**, 41–50.

Hulse, S. H. (**1995**). "The discrimination-transfer procedure for studying auditory perception and perceptual invariance in animals," in *Methods in Comparative Psychoacoustics*, edited by G. M. Klump, R. J. Dooling, R. R., Fay, and W. C. Stebbins (Birkhauser Verlag, Basel, Switzerland), pp. 319–330.

Kettner, R. E., and Thompson, R. F. (**1985**). "Cochlear nucleus, inferior colliculus and medial geniculate reponses during the behavioral detection of threshold-level auditory stimuli in the rabbit," J. Acoust. Soc. Am. **77**, 2111–2127.

Krishnan, A., Xu, Y., Gandour, J., and Cariani, P. (**2005**). "Encoding of pitch in the human brainstem is sensitive to language experience," Brain Res. Cognit. Brain Res. **25**, 161–168.

Neuert, V., Verhey, J., and Winter, I. M. (**2005**). "Temporal representation of the delay of iterated rippled noise in the dorsal cochlear nucleus," J. Neurophysiol. **93**, 2766–2776.

Palmer, A. R., and Russell, I. J. (**1986**). "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells," Hear. Res. **24**, 1–15.

Patterson, R. D., Handel, S., Yost, W. A., and Datta, A. J. (**1996**). "The relative strength of the tone and noise components in iterated rippled noise," J. Acoust. Soc. Am. **100**, 3286–3294.

Plack, C. J., Oxenham, A. J., Fay, R. R., and Popper, A. N. (**2005**). *Pitch: Neural Coding and Perception*, Vol. **24** in *Springer Handbook of Auditory Research* (Springer, New York).

Polley, D. B., Steinberg, E. E., and Merzenich, M. M. (**2006**). "Perceptual learning directs auditory cortical map reorganization through top-down influences," J. Neurosci. **26**, 4970–4982.

Recanzone, G. H., Schreiner, C. E., and Merzenich, M. M. (**1993**). "Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys," J. Neurosci. **13**, 87–103.

Shofner, W. P. (**1999**). "Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: Analysis of neural autocorrelograms," J. Neurophysiol. **81**, 2662–2672.

Shofner, W. P. (**2002**). "Perception of the periodicity strength of complex sounds by the chinchilla," Hear. Res. **173**, 69–81.

Shofner, W. P., and Selas, G. (**2002**). "Pitch strength and Stevens' Power Law," Percept. Psychophys. **64**, 437–450.

Shofner, W. P., Sparks, K., Wu, Y. E., and Pham, E. (**2005a**). "Similarity of spectral resolvability in chinchillas and human listeners based on phase discrimination," ARLO **6**, 35–40.

Shofner, W. P., Whitmer, W. M., and Yost, W. A. (**2005b**). "Listening experience with iterated rippled noise alters the perception of 'pitch' strength of complex sounds in the chinchilla," J. Acoust. Soc. Am. **118**, 3187–3197.

Terhardt, E. (**1974**). "Pitch, consonance, and harmony," J. Acoust. Soc. Am. **55**, 1061–1069.

Winter, I. M., Wiegrebe, L., and Patterson, R. D. (**2001**). "The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea-pig," J. Physiol. (London) **537.2**, 553–566.

Woody, C. D., Wang, X.-F., Gruen, E., and Landeira-Fernandez, J. (**1992**). "Unit activity to click CS changes in dorsal cochlear nucleus," NeuroReport **3**, 385–388.

Woody, C. D., Wang, X.-F., and Gruen, E. (**1994**). "Response to acoustic stimuli increases in the ventral cochlear nucleus after stimulus pairing," NeuroReport **5**, 513–515.

A comparison of underwater hearing sensitivity in bottlenose dolphins (*Tursiops truncatus*) determined by electrophysiological and behavioral methods

Dorian S. Houser[a)]
Biomimetica, 7951 Shantung Drive, Santee, California 92071

James J. Finneran
U.S. Navy Marine Mammal Program, Space and Naval Warfare Systems Center, San Diego, Code 2351, 53560 Hull Street, San Diego, California 92152

(Received 24 March 2006; revised 19 June 2006; accepted 26 June 2006)

Variable stimulus presentation methods are used in auditory evoked potential (AEP) estimates of cetacean hearing sensitivity, each of which might affect stimulus reception and hearing threshold estimates. This study quantifies differences in underwater hearing thresholds obtained by AEP and behavioral means. For AEP estimates, a transducer embedded in a suction cup (jawphone) was coupled to the dolphin's lower jaw for stimulus presentation. Underwater AEP thresholds were obtained for three dolphins in San Diego Bay and for one dolphin in a quiet pool. Thresholds were estimated from the envelope following response at carrier frequencies ranging from 10 to 150 kHz. One animal, with an atypical audiogram, demonstrated significantly greater hearing loss in the right ear than in the left. Across test conditions, the range and average difference between AEP and behavioral threshold estimates were consistent with published comparisons between underwater behavioral and in-air AEP thresholds. AEP thresholds for one animal obtained in-air and in a quiet pool demonstrated a range of differences of −10 to 9 dB (mean=3 dB). Results suggest that for the frequencies tested, the presentation of sound stimuli through a jawphone, underwater and in-air, results in acceptable differences to AEP threshold estimates. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2229286]

PACS number(s): 43.80.Lb, 43.64.Ri [WWA] Pages: 1713–1722

I. INTRODUCTION

Auditory evoked potential (AEP) techniques currently provide the best opportunity for successfully responding to National Research Council (NRC) recommendations to obtain population-level audiograms for marine mammals and estimate the hearing sensitivity of rare or hard to maintain marine mammals. Specifically, the NRC stated that "population-level audiograms of many individuals are necessary for establishing baseline hearing capabilities and normal hearing loss in marine mammals" and that there should be coordinated efforts to "obtain audiometric data from stranded or ensnared whales" (NRC, 1994; 2000). Hearing information remains to be obtained on most marine mammal species and the natural variation in hearing sensitivity within a population (i.e., as a function of age and/or sex) is unknown for all marine mammal species.

A number of AEP techniques exist for measuring auditory thresholds. These include the measurement of auditory brainstem responses (ABRs) resulting from exposure to clicks, tone pips, and sinusoidal amplitude modulated (SAM) tones. Clicks have broad spectral content and are suitable for testing hearing sensitivity across a range of frequencies simultaneously. Tone pips are short bursts of a tone, typically less than a few ms, which improve frequency specificity of the evoked response because of a narrower spectral content (compared to a click). However, the technique gaining favor for use with small odontocetes is the recording of the envelope following response (EFR). This technique often uses SAM tones as stimuli in order to elicit rhythmic ABRs. The magnitude of the EFR depends on the SAM tone carrier frequency and sound pressure level, as well as the amplitude modulation rate and the depth of modulation. Studies on several odontocete species have demonstrated optimal amplitude modulation rates and depths for detection of the EFR (Dolphin *et al.*, 1995; Supin and Popov, 1995, 2000; Mooney and Nachtigall, 2006). The rhythmic nature of the EFR makes it amenable to analysis in the frequency domain and provides greater frequency specificity than do tests measuring evoked responses to either clicks or tone pips. The EFR has been used to estimate the hearing sensitivity of several marine mammal species (Dolphin, 2000; Klishin *et al.*, 2000; Andre *et al.*, 2003; Nachtigall *et al.*, 2005; Popov *et al.*, 2005; Yuen *et al.*, 2005), but quantification of differences between behavioral and EFR thresholds obtained within the same animal has only recently been performed (Yuen *et al.*, 2005; Finneran and Houser, 2006). Prior to broad acceptance of the EFR technique for use in estimating hearing sensitivity in marine mammals, appropriate quantification of the differences that exist between more commonly accepted behavioral approaches to estimating hearing sensitivity and those determined with EFR should be performed.

[a)]Electronic mail: biomimetica@cox.net

The conditions under which marine mammals are available for study influence how SAM stimuli are presented or how EFRs are recorded. For example, a stranded delphinid on the beach may require use of a "jawphone" (suction cup embedded sound projector) for stimulus presentation (Brill *et al.*, 2001; Finneran and Houser, 2006), where a rehabilitating or trained delphinid tested underwater in a pool may allow direct field presentation of acoustic stimuli (Nachtigall *et al.*, 2005). The differences in threshold estimates from alternative methods are currently not quantified, but may vary as a function of equipment used (e.g., jawphone versus direct field) or the sound propagation characteristics of the testing environment (e.g., stimuli emitted underwater versus those emitted in air). Threshold estimates resulting from differing approaches to EFR data collection should be cautiously applied and placed in the proper context of how the data were collected. Ultimately, variation in threshold estimates resulting from these different approaches should be quantified to facilitate comparisons in threshold estimation across studies.

Finneran and Houser (2006) quantified differences in EFR and behavioral thresholds for the same dolphins and at the same frequencies. The electrophysiological approach was used with four animals, all of which were out of the water and quietly resting on a padded mat. Behavioral estimates of hearing sensitivity were obtained underwater, either in San Diego Bay (SD Bay) or in a quiet pool. Compared to behavioral thresholds obtained in SD Bay at unmasked frequencies (>40 kHz), AEP estimates were on average 6 dB higher. Comparison across the range of hearing for a single animal whose behavioral estimates were obtained in a quiet pool indicated that AEP threshold estimates were from 0 to 18 dB higher, with the greatest differences occurring at the lowest (10 kHz) and highest frequencies (80 and 100 kHz) at which that individual was tested. These differences must necessarily be placed in context of the data collection design used in that study. Dolphins were in air during the procedure and stimuli were presented to them via a jawphone coupled to the pan region of the lower jaw. Differences between behavioral and EFR threshold estimates obtained in this manner may differ from those obtained underwater or in which stimulus presentation occurred in the direct field.

The jawphone has been used for stimulus presentation in the mapping of the sensitivity of the dolphins' lower jaw to click-like signals (Møhl *et al.*, 1999). It has also been used to preferentially stimulate one ear over the other during behavioral audiometry conducted with the animal underwater (Brill *et al.*, 2001). The results of Brill *et al.* suggested that behavioral audiograms obtained from an animal with the jawphone compared favorably to hearing sensitivity measured in the direct field. The study presented here extends the use of the jawphone for presenting SAM stimuli and applying the EFR technique to submerged animals. EFRs were recorded at multiple carrier frequencies and analyzed to estimate hearing thresholds. The results are compared to behavioral thresholds within the same dolphins to determine the differences between the two approaches. EFR-derived threshold estimates obtained on submerged dolphins are further compared to estimates on the same dolphins in air (Finneran and Houser, 2006) to examine how threshold predictions obtained with EFRs varied as a function of the medium in which the data are collected.

II. METHODS

A. Subjects

Three bottlenose dolphins were used: BEN (male, 41 yr, 324 kg), BLU (female, 39 yr, 200 kg), and WEN (male, 21 yr, 210 kg). Subjects were housed in floating netted enclosures, 9×9 to 12×24 m, located in SD Bay, CA. The study followed a protocol approved by the Institutional Animal Care and Use Committee of the Biosciences Division, SSC San Diego, and followed all applicable U.S. Department of Defense guidelines for the use and treatment of animals in science.

B. AEP measurements

1. Hardware

All AEP measurements were conducted underwater. BEN, BLU, and WEN were tested in SD Bay; BLU was also tested in an above-ground vinyl-walled pool. Representative ambient noise levels for the pool and SD Bay test sites are provided in Finneran *et al.* (2005). Within SD Bay, pressure spectral densities were approximately 75 dB $re:1\ \mu\text{Pa}^2/\text{Hz}$ at 5 kHz and declined to approximately 63 dB $re:1\ \mu\text{Pa}^2/\text{Hz}$ as frequency increased to 50 kHz. Noise levels in the pool were approximately 20 dB below ambient noise levels in SD Bay.

The equipment used to generate sound stimuli and acquire data in SD Bay has been previously detailed (Finneran and Houser, 2006). Briefly, a multifunction data acquisition card (National Instruments PCI-MIO-16E-1) was used to generate sound stimuli and digitize the evoked responses via a personal computer. Sound stimuli were attenuated (Tucker-Davis Technologies PA-5), bandpass filtered (Krohn-Hite 3C module, 1–150 kHz), and amplified (Hafler P1000) before being projected to the dolphin. Stimuli were presented to the subjects via a jawphone consisting of a piezoelectric sound projector embedded in a Rhodia V-1065 silicon rubber suction cup (Brill *et al.*, 2001; Finneran and Houser, 2006). The jawphone was coupled to the dolphin's lower jaw over the pan region, a site which is an important entryway to the auditory system for high frequency sounds (Bullock *et al.*, 1968; McCormick *et al.*, 1970, 1980; Brill *et al.*, 1988; Møhl *et al.*, 1999). A diagram of the system, reproduced from Finneran and Houser (2006), is provided in Fig. 1.

Two jawphones were used in this study: JP4 was fabricated using a commercially available sound projector (ITC 1042) while JP5 was constructed using a custom piezoelectric sphere. JP4 was used for all subjects tested in SD Bay and JP5 was used for BLU when she was tested in the pool. Jawphones were calibrated by measuring underwater rms sound pressure levels at a distance of 15 cm from the jawphone (Finneran and Houser, 2006). The 15-cm distance was based on the measured distance between the jawphone attachment point and the ipsilateral auditory bulla as determined from a computed tomography (CT) scan of WEN (Houser *et al.*, 2004). Sound pressure levels were measured

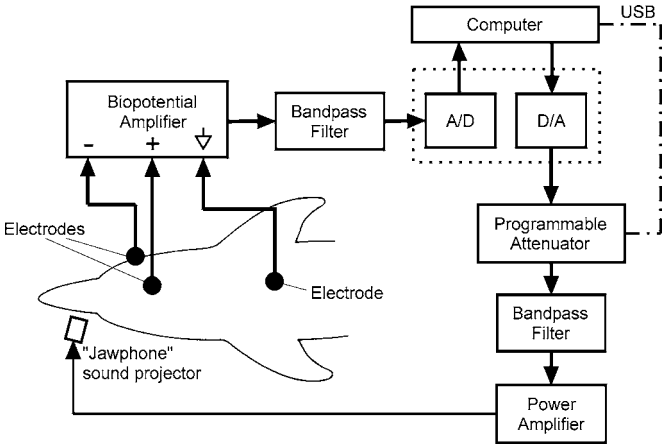


FIG. 1. AEP measurement system configuration (reproduced from Finneran and Houser, 2006).

using a calibrated hydrophone (B&K 8105) and charge amplifier (B&K 2692 or 2635). Figure 2 shows the resulting transmitting voltage responses (TVR). Note that the TVR differs from that of a free transducer; the response of the transducer is damped by its being embedded in the silicon rubber. The straight lines represent linear regressions of TVR versus log-frequency performed over a range of 5–100 kHz for JP4 and 5–40 kHz for JP5 (the frequency range over which each TVR was approximately linear with log-frequency). Regression slopes were 12.5 and 12.3 dB/octave and r^2 values were 0.95 and 0.99 for JP4 and JP5, respectively.

Gold cup electrodes (6 and 10 mm diameter) embedded in 25-mm-diam Rhodia V-1065 silicon rubber suction cups were used as surface electrodes. Prior to electrode placement, the dolphin positioned the region of its body where the electrode was to be placed out of the water and the attachment site was dried using gauze pads. The noninverting electrode (+) was placed near the vertex of the dolphin's head, approximately 10 cm posterior of the blowhole and offset ~2 cm contralateral of the ear being tested. The exact position of the (+) electrode varied depending on the shape of the subject's head. Curvature of the head sometimes prevented adequate sealing of the suction cups at the preferred attachment site and the suction cup had to be moved slightly to

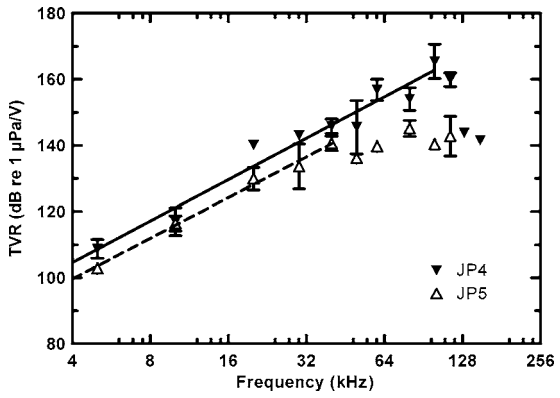


FIG. 2. Transmitting voltage response (TVR) from jawphones JP4 and JP5 measured 15 cm from the transducer. The lines represent linear regressions of TVR vs log-frequency.

ensure an adequate seal. The inverting electrode (–) was placed contralateral to the ear being tested, and just behind the external auditory meatus. A ground (common) electrode was placed either on the subject's back near the dorsal fin or on the dorsal fin itself. Subjects were positioned near the surface so that the (+) and ground electrodes were normally above the waterline.

2. Stimuli

SAM tones were used as stimuli to generate the EFRs. Carrier frequencies ranged from 5 kHz (pool only) or 10 kHz up to 150 kHz. Modulation depth was 100%. Modulation rates were 1 kHz for all carrier frequencies except 5 kHz, where the modulation rate was lowered to 500 Hz to keep the stimulus bandwidth within the anticipated width of the auditory filter centered at the carrier frequency (Au and Moore, 1990; Finneran *et al.*, 2002). Prior research on bottlenose dolphins has indicated that modulation rates from approximately 500 to 600 and 1000 to 1200 Hz evoke strong responses (Dolphin *et al.*, 1995; Supin and Popov, 1995; Popov and Supin, 1998).

All dolphins were tested using intermittent SAM tones with 1 ms rise/fall times. Interstimulus intervals (quiet period between stimuli) were ⩾6 ms. Preliminary data showed no significant changes in EFR amplitudes for ISIs ⩾6 ms. The specific durations depended on the need to balance the speed achieved with shorter stimuli versus the better frequency resolution and lower noise levels achieved with longer stimuli. Initial tests in SD Bay used durations of 13 ms; this was increased to 23 ms for the majority of tests. Tests in the pool used durations of 23 and 32 ms except at 5 kHz, where the stimulus duration was extended to 62 ms.

3. Evoked responses and response detection

Evoked responses were differentially amplified and filtered using a Grass IP-511 biopotential amplifier. The amplifier gain was fixed at 100 000. The high- and low-pass filters were set at 300 Hz and 3 kHz, respectively, except for measurements at 5 kHz where the high-pass filter was set at 100 Hz. Additional filtering (Krohn-Hite 3C, 100 Hz and 5 kHz) was sometimes used, depending on signal quality. The resulting signal was digitized (National Instruments PCI-MIO-16E-1) at either 15 or 20 kHz. Signals greater than 20 μV were rejected from analysis. For each frequency tested, 500 epochs were recorded. Frequency analysis was performed on 11–30 ms epochs, except for measurements at 5 kHz where the epoch duration was set to either 21 or 60 ms. The portions of the evoked response corresponding to the stimulus rise and fall were not included in the frequency analysis. To avoid spectral leakage in the frequency domain, durations for frequency analysis were constrained to integral multiples of 1 ms (1 kHz modulation rate) or 2 ms (500 Hz modulation rate).

Magnitude-squared coherence (MSC) was calculated from the total collection of epochs obtained for each frequency and stimulus level tested. The MSC was used to assess whether the spectral component at the modulation frequency was statistically different from noise (Dobie and

Wilson, 1989; Dobie, 1993; Dobie and Wilson, 1996). The MSC calculation used 20 subaverages and critical values for MSC, using $\alpha=0.01$, were obtained from Amos and Koopmans (1963) and Brillinger (1978). Evoked responses with MSC values larger than the critical MSC (a function of α and the number of subaverages) were considered to be detected.

4. Data collection

Data collection began with a stimulus SPL of 80 dB re 1 μPa. The same automated modified staircase technique described in Finneran and Houser (2006) was used to adjust stimulus levels to quickly reach threshold. The staircase was terminated when the step size was reduced below 4 dB. A linear regression was then performed on the AEP amplitude versus stimulus SPL data. Data included in the regression were all detected responses except those exceeding 400 nV. If the regression value (r^2) from a minimum of four detected responses reached 0.9, the test was concluded. If the minimum r^2 value was not obtained, additional measurements and regression analyses were conducted until the criterion r^2 was met or a maximum of eight detections was made.

For hearing tests conducted in SD Bay, all frequencies were tested three to seven times for both ears of each test subject. For pool recordings, all frequencies were tested four to five times for both of BLU's ears. The exception for pool testing was at 100 kHz, where testing was performed once on the left ear and twice on the right ear. The decision to reduce the number of trials at 100 kHz was due to BLU's insensitivity at this frequency.

5. Threshold estimation

Hearing thresholds were estimated for each frequency using a linear regression technique (Supin *et al.*, 2001). The regression was first performed on the four detected responses with the lowest amplitudes. If the regression r^2 was less than 0.9, additional points obtained at consecutively higher stimulus levels were added to the regression analysis until the r^2 criterion of 0.9 was met, or an obvious plateau was reached in the AEP amplitude data. The threshold was estimated by extrapolating the regression line to the 0 V level and determining the SPL for the zero-crossing. Occasionally, only three points could be included in the regression because of the appearance of a plateau in the amplitude response curve that occurred at relatively low stimulus levels. If a minimum of a three point regression could not be performed at a particular frequency, estimates for that frequency were not performed.

C. Behavioral measurements

Behavioral thresholds were measured underwater in SD Bay (all subjects) and the above-ground pool (BLU only). Subjects were trained to whistle in response to test tones and to stay quiet otherwise. Stimulus levels were adjusted using a modified up/down staircase technique with 500 ms pure tone stimuli. Details of the behavioral test methods for WEN and BLU may be found in Finneran and Houser (2006). Behavioral thresholds for BEN were acquired as part of a prior study on noise-induced temporary threshold shifts (Finneran *et al.*, 2005). Testing for WEN was performed in context of a single interval experiment. The behavioral threshold procedures for BEN and BLU were based on the Method of Free Response (MFR; Egan *et al.*, 1961).

D. Analysis

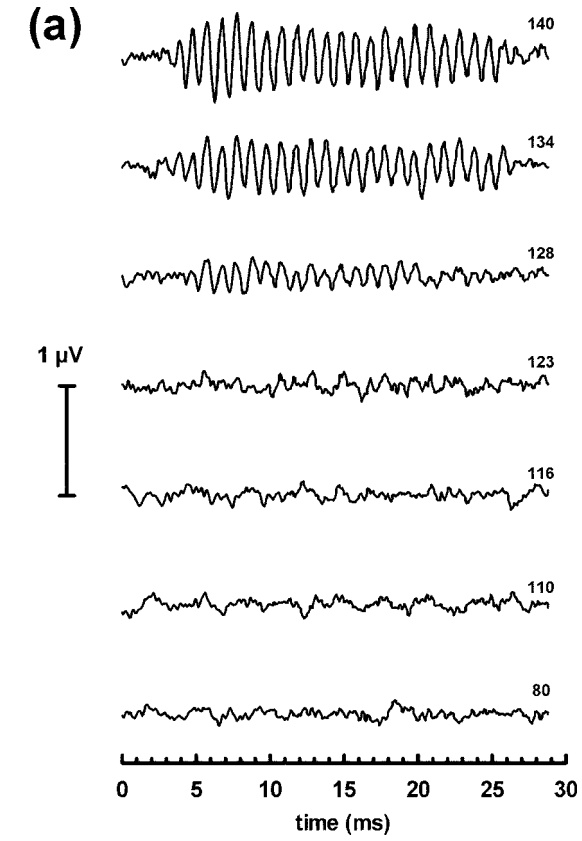
T-tests for independent samples ($\alpha=0.05$) were used to compare AEP threshold estimates obtained from stimulation of the left and right ears. Statistical analyses were performed with the software package STATISTICA©(StatSoft, Inc.).

III. RESULTS

Figure 3(a) shows examples of the EFR wave form obtained from BLU, while submerged, in response to an intermittent 50 kHz SAM tone. Responses are easily observed in the wave forms at high stimulus levels. As the stimulus SPL is lowered, the magnitude of the response decreases until it is eventually indistinguishable from noise in the time domain. Figure 3(b) shows the frequency spectra of the wave forms from Fig. 3(a). The feature of interest for detection of the EFR is the spectral peak at 1 kHz, which corresponds to the amplitude modulation frequency of the SAM tone.

Figure 4 shows the underwater AEP audiogram and behavioral audiogram for BLU derived from measurements made in SD Bay. AEP responses could not be detected above 100 kHz at the highest SPL the jawphone could generate. As previously demonstrated for BLU, sensitivity above 30 kHz becomes progressively poorer with a reduction in sensitivity of ~47 dB/octave (Finneran and Houser, 2006). The pattern of hearing sensitivity with frequency is similar regardless of whether AEP estimates or behavioral estimates of threshold are used. The AEP threshold estimates tended to be lower than the behavioral estimates and there was a strong correlation between estimates ($r=0.93$). When the sensitivity of each ear was considered independently, differences between estimates (AEP threshold estimates minus behavioral thresholds) ranged from −18 to +22 dB re 1 μPa. The mean underestimate of threshold was 7 dB while the mean overestimate of threshold was 11 dB. When the sensitivities of the ears were averaged, differences between estimates ranged from −16 to +20 dB and the mean underestimate and overestimate of threshold was −5 and +12 dB, respectively. Differences between the left and right ear AEP audiograms were insignificant across all frequencies tested, except at 50 kHz ($t=3.8$, $p=0.02$). At this frequency the left ear threshold was ~4 dB higher than that obtained for the right ear.

Figure 5 compares underwater AEP threshold estimates for WEN to behavioral thresholds measured in SD Bay. At frequencies below 60 kHz, AEP threshold estimates consistently underestimated behavioral thresholds obtained in SD Bay. Conversely, above 60 kHz, AEP threshold estimates overestimated behavioral thresholds. Considering each ear independently, differences between estimates ranged from −28 to +23 dB, the mean underestimate of threshold was 16 dB, and the mean overestimate of threshold was 13 dB. The correlation between threshold estimates was 0.50. When the sensitivities of the ears were averaged, differences be-



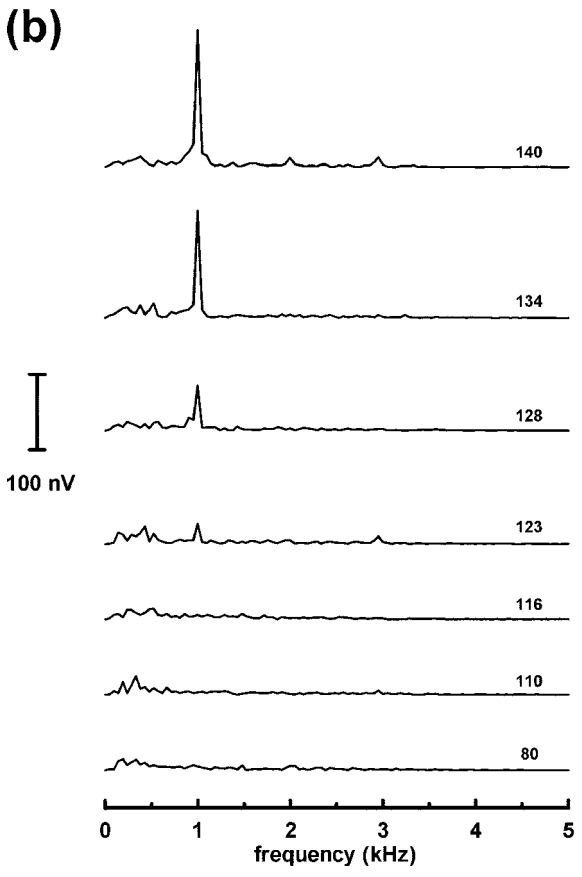


FIG. 3. EFR (a) wave forms and (b) frequency spectra measured at different stimulus SPLs for BLU while she was submerged. The modulation rate of the SAM tone was 1 kHz and the carrier frequency was 50 kHz. Time $t=0$ corresponds to the stimulus onset. The frequency spectra were calculated from a 10 ms analysis window. The numbers next to each series indicate the stimulus SPL (dB re 1 μPa).

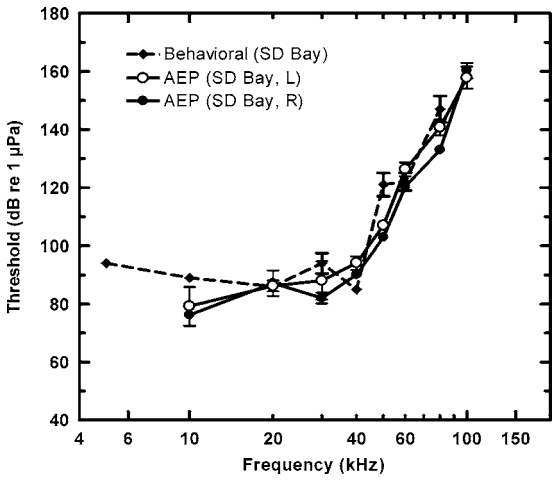


FIG. 4. Comparison of EFR threshold estimates obtained from left (L) and right (R) ears to behavioral thresholds for subject BLU. All EFR tests and behavioral tests were performed in SD Bay. Symbols represent mean values; error bars show the mean ± one standard deviation.

tween estimates ranged from −20 to +20 dB and the mean underestimate and overestimate of threshold was −14 and +14 dB, respectively. Differences between the left and right ear AEP audiograms were insignificant across all frequencies tested, except at 115 kHz ($t=-5.2$, $p<0.01$). At this frequency the right ear threshold was ~6 dB higher than that obtained for the left ear.

Figure 6 compares AEP threshold estimates for BEN to his behavioral thresholds measured in SD Bay. In contrast to BLU and WEN, AEP thresholds were generally higher than behavioral estimates. Behavioral thresholds for BEN were much higher than either WEN or BLU from 15 to 30 kHz (note the notch at these frequencies in Fig. 6). Differences between estimates ranged from −20 to +29 dB. The mean underestimate of threshold was 8 dB while the mean overestimate of threshold was 13 dB. The correlation between threshold estimates was 0.90. When the sensitivities of the ears were averaged, differences between estimates ranged from −8 to +21 dB and the mean underestimate and overes-

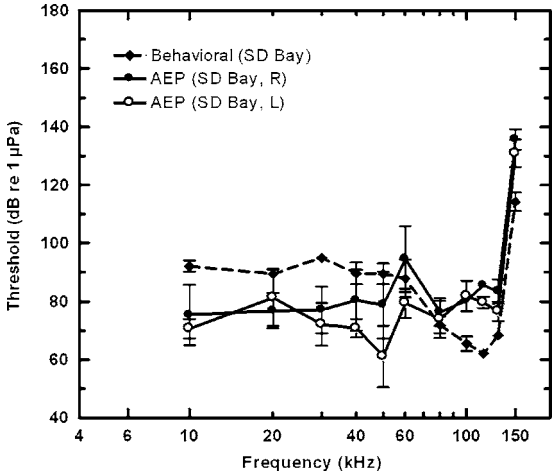


FIG. 5. Comparison of EFR threshold estimates obtained from left (L) and right (R) ears to behavioral thresholds for subject WEN. All EFR tests and behavioral tests were performed in SD Bay. Symbols represent mean values; error bars show the mean ± one standard deviation.

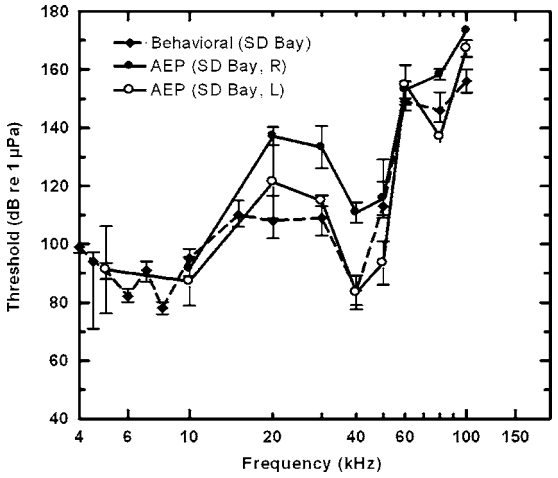


FIG. 6. Comparison of EFR threshold estimates obtained from left (L) and right (R) ears to behavioral thresholds for subject BEN. All EFR tests and behavioral tests were performed in SD Bay. Symbols represent mean values; error bars show the mean ± one standard deviation.

timate of threshold was −7 and +12 dB, respectively. AEP threshold estimates obtained for the right ear of BEN were significantly higher than those obtained from the left for frequencies of 20 ($t=-2.7$, $p=0.02$), 30 ($t=-5.7$, $p<0.01$), 40 ($t=-6.2$, $p<0.01$), 50 ($t=-3.4$, $p=0.02$), 80 ($t=-19.7$, $p<0.01$), and 100 kHz ($t=-3.6$, $p=0.02$). Thresholds were 15–34 dB higher when testing the right ear, except at 100 kHz where it was 6 dB higher. Differences at 10 and 60 kHz were not significantly different. Overall, AEP threshold estimates for the left ear were nearer to behavioral thresholds than were those obtained from the right ear.

Agreement between mean behavioral and AEP estimates (i.e., average of the left and right ears) were best at 60 and 80 kHz when all measurements made in SD Bay were considered (Fig. 7). All AEP predictions of threshold were lower than behavioral thresholds at 50 and 10 kHz, whereas AEP estimates of threshold were both lower and higher than behavioral thresholds at frequencies from 20 to 40 kHz. Con-

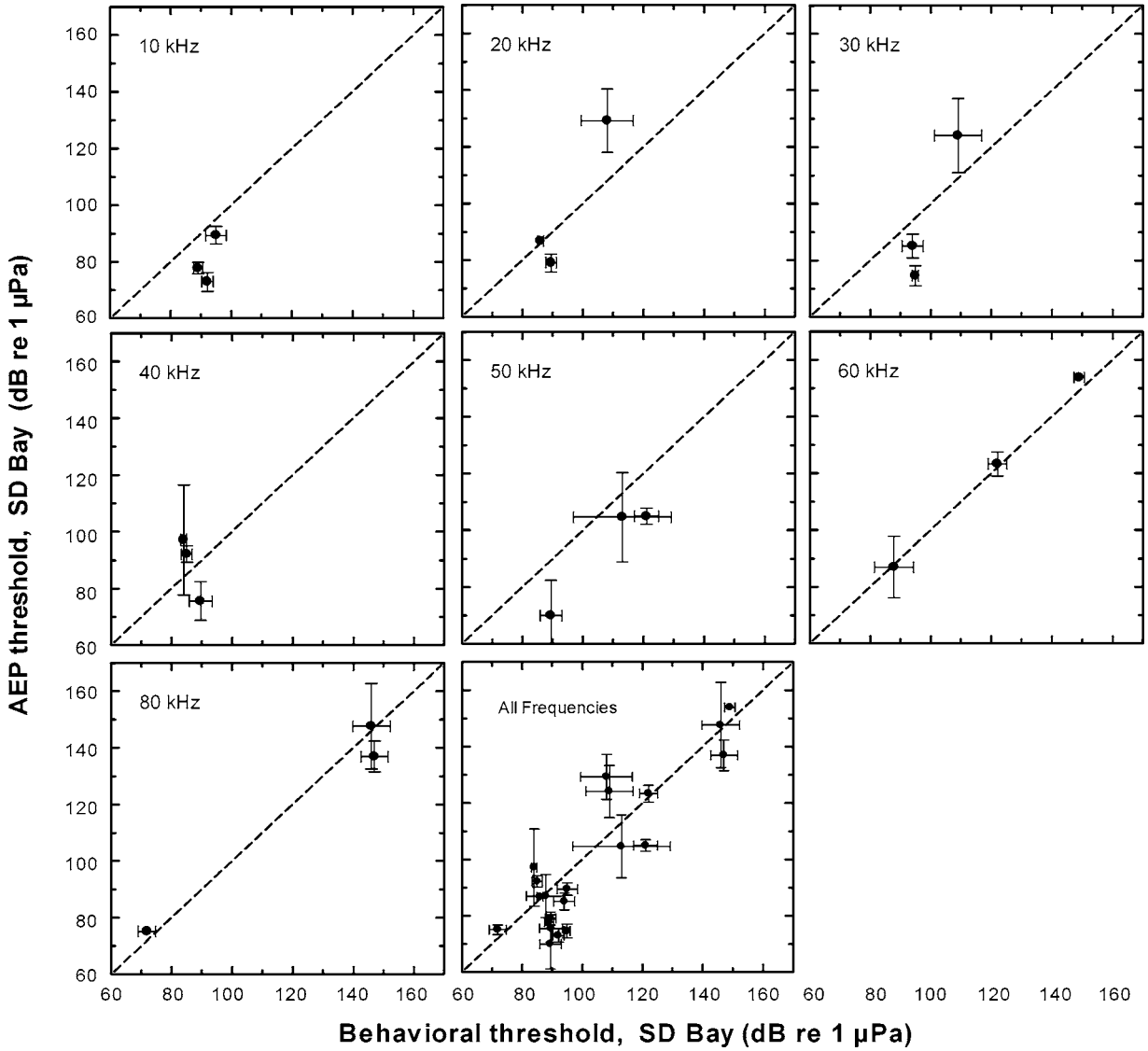


FIG. 7. EFR threshold estimates vs behavioral thresholds for all measurements made in SD Bay. Each panel, except the bottom center panel, shows data from all subjects for a single frequency. The bottom, center panel shows data from all of the subjects and for all of the frequencies tested. The dotted line indicates where perfect agreement (equivalence) between the threshold predictions would occur. Vertical error bars denote the standard deviation of AEP threshold estimates. Horizontal error bars denote the standard deviation of behavioral thresholds.

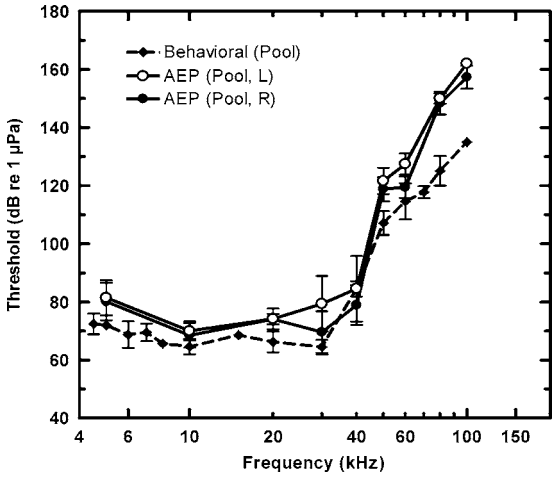


FIG. 8. Comparison of EFR threshold estimates obtained from left (L) and right (R) ears to behavioral thresholds for BLU. The EFR tests and behavioral tests were performed in an above ground pool. Symbols represent mean values; error bars show the mean ± one standard deviation.

sidering data collected from all three of the subjects, the average underestimate of threshold was 11 dB and the average overestimate was 13 dB.

Figure 8 demonstrates the relationship between AEP threshold estimates and behavioral thresholds collected in a pool for BLU. This was the only animal for which pool collections were made, thus providing a single comparison for a quiet environment. Threshold estimates derived from AEP data collected in the pool were consistently higher than behavioral thresholds. Considering each ear independently, differences between estimates ranged from −6 to +27 dB, with the greatest deviations occurring at the highest frequencies tested. The mean underestimate of threshold was 3 dB while the mean overestimate of threshold was 13 dB. When the average threshold was determined for both ears, differences between estimates ranged from −3 to +25 dB. Because equivalent testing for each ear occurred across the range of hearing, the mean overestimate and underestimate for the averaged threshold were the same as that obtained when each ear was considered independently. Differences between the left and right ear AEP audiograms were insignificant across all frequencies tested, except at 60 kHz ($t=3.2$, $p=0.01$). At this frequency the left ear threshold was ~8 dB higher than that obtained for the right ear.

IV. DISCUSSION

A. AEP and behavioral thresholds

1. San Diego Bay

Behavioral thresholds for BLU and WEN in SD Bay, and those for BLU in the pool, have been previously reported and discussed (Finneran and Houser, 2006). When only comparisons in San Diego Bay are considered, differences between underwater behavioral thresholds and underwater AEP thresholds were more variable for WEN than for BLU. The relatively high ambient noise levels in SD Bay likely mask hearing up to at least 40 kHz, as previously suggested (Finneran and Houser, 2006); however, it is unlikely that this alone explains the differences in variability since the masking effect should have been consistent between the animals. The variances in behavioral data were generally small compared to variances observed in the AEP data, suggesting that the variability lies within the AEP thresholds and not the behavioral thresholds. The behavioral thresholds for BLU were obtained via the MFR, whereas those of WEN were obtained in a single interval experiment (Finneran *et al.*, 2005; Finneran and Houser, 2006). Behavioral thresholds at low frequencies (where thresholds were masked by ambient noise) were similar, suggesting that the two methods produced comparable results. Sensitivities measured for BLU and WEN were dramatically different at high frequencies where BLU exhibited hearing loss. Whether reduced sensitivity results in greater consistency between AEP and behavioral thresholds is unknown for dolphins. It is possible that the variability in ambient noise levels, from biological sources and vessel traffic, had a more dramatic effect on WEN's thresholds than BLU's, since BLU possessed higher thresholds and test stimuli would have been larger relative to the background noise.

The audiogram of BEN is atypical in that it contains a notch, or reduction in sensitivity, between 10 and 40 kHz. The etiology of this hearing anomaly is unknown and it cannot be determined whether the underlying cause of the notch is also related to differences between the left and right ears. BEN exhibited large left/right differences, with thresholds for the right ear substantially higher (up to 34 dB) than the behavioral results. In contrast, thresholds were similar for both the left and right ears of BLU and WEN for all but one of the frequencies tested. Asymmetry in hearing sensitivity has previously been observed in another bottlenose dolphin (HEP) whose sensitivity was tested through the use of a jawphone (Brill *et al.*, 2001). Right ear thresholds of this subject were up to ~30 dB greater than those of the left and a bilateral reduction in sensitivity was observed above 50 kHz, similar to that observed in BEN. Both BEN and HEP are older males, >35 yr old, and it is possible that the underlying factors contributing to left and right differences in auditory sensitivity are augmented by presbycusis. Unlike WEN and BLU, whose behavioral thresholds were measured concurrently or within weeks of the AEP thresholds, the behavioral audiogram of BEN was obtained several years prior to his AEP audiogram. Given the age of BEN (>40 yr), it is reasonable to suspect that his hearing sensitivity might have continued to decline with aging, thus contributing to the observed differences in AEP and behavioral thresholds. Alternatively, the behavioral audiogram may be more reflective of the more sensitive of the two ears. In this case, the appropriate comparison would be to the AEP threshold from BEN's left ear (the more sensitive ear). The resulting differences between the behavioral and AEP thresholds from the left ear are not significant at most frequencies. The possibility that the audiogram is reflective of the more sensitive ear deserves more research as it could provide insight into the accommodation of hearing loss in odontocete species.

2. Pool

Because of the likelihood for masking to have occurred in SD Bay measurements, the best comparisons of underwa-

ter AEP and behavioral thresholds come from those obtained in the pool environment for BLU. The range of differences between underwater AEP and behavioral thresholds obtained in the pool was approximately 15 dB greater than that observed when in-air AEP threshold estimates were compared to the pool data (Finneran and Houser, 2006). However, the trend was the same between the two studies with the AEP-estimated thresholds typically overestimating the behavioral thresholds.

B. Methodological issues

Prior estimates of hearing sensitivity in delphinid cetaceans obtained via EFRs utilized different methodologies. For example, Yuen *et al.* (2005) performed EFR assessments of hearing sensitivity with their subject, a false killer whale (*Pseudorca crassidens*), underwater in the direct field of the sound source. Other studies have utilized similar procedures (Nachtigall *et al.*, 2005; Popov *et al.*, 2005). In contrast, Finneran and Houser (2006) used EFRs to estimate hearing sensitivity in four bottlenose dolphins by coupling the stimulus to the lower jaw with a suction cup while the animals rested out of the water on a foam mat. Cook *et al.* (2006) used a similar jawphone technique for underwater EFR measurements in a beaked whale. The different manner of stimulus presentation implemented in these studies might result in the use of different acoustic paths to the ear, thus altering the received characteristics of presumably similar stimuli.

Comparability of AEP results when the jawphone is used for stimulus delivery underwater versus in air is of potential concern because the jawphone is not jacketed with sound absorbing material. Thus, the stimulus would have a more omni-directional radiation underwater and could potentially enter the ear via pathways that are otherwise isolated during in-air testing. Using the average threshold of the two ears and considering only the measurements made in SD Bay, the range of differences between underwater AEP threshold estimates and underwater behavioral thresholds obtained in this study (−20 to 21 dB) compared similarly to those observed between in-air AEP threshold estimates and underwater behavioral thresholds (−26 to 20 dB; Finneran and Houser, 2006). The average difference and standard deviation of the differences were also comparable (3±13 dB vs −2±13 dB, respectively) suggesting that the variations were largely unaffected by the medium in which the animals were tested (i.e., water versus air).

Figure 9 demonstrates the relationship between AEP threshold estimates obtained underwater, in the pool, to AEP threshold estimates previously obtained for BLU in air (Finneran and Houser, 2006). The procedures between the two studies were the same, except that the current study presented stimuli and recorded evoked potentials with the subject underwater, and the durations for intermittent stimuli were extended several tens of milliseconds to compensate for electrical leakage to the water. Differences between the in-air and underwater AEP thresholds, calculated as in-air estimates minus those obtained in the pool, ranged from −10 to 9 dB, with a mean difference of 3 dB. This comparison provides the most direct assessment of how medium differences and the potential excitation of different auditory pathways

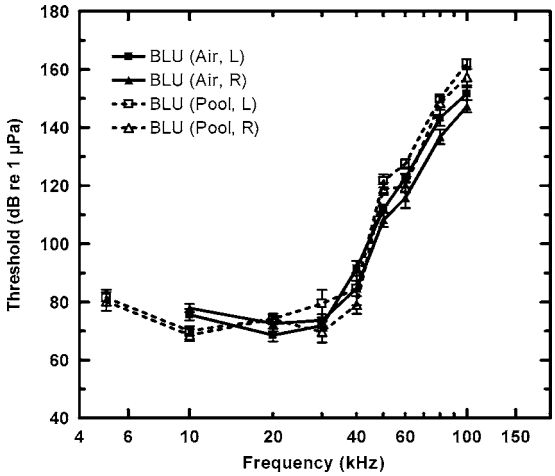


FIG. 9. Comparison of EFR threshold estimates for BLU obtained underwater and in-air. In-air EFR threshold estimates are from Finneran and Houser (2006).

may affect AEP threshold estimates obtained with jawphones. The range of threshold differences between the two data collection efforts and the small mean difference suggests that the two approaches do not result in substantially different threshold estimates. In conjunction with previously discussed comparisons, it appears that greater sound radiation from the jawphone during submersion is not a substantial concern when considering frequencies involved in echolocation and that the primary acoustic pathway at these frequencies is via the lower jaw (Bullock *et al.*, 1968; McCormick *et al.*, 1970, 1980; Brill *et al.*, 1988; Møhl *et al.*, 1999). Similar conclusions may not be valid for lower frequencies, such as those used in whistle communication, because the longer wavelength could be too long to utilize the jaw fat pathway or could diffract around the animal and possibly reach the auditory system through alternative acoustic pathways (Au, 1993).

The suitability of jawphone-measured thresholds as predictors of free field or direct field underwater thresholds is somewhat controversial. There has been no universally accepted technique for calibrating jawphones and the effects of the jawphone attachment on the subject's receiving system are unknown. Calibration measurements for this study were performed at 15 cm distance (beyond the minimum source distance as specified in ANSI S1.20-1998, "Procedures for calibration of underwater electro-acoustic transducers") and the measurements themselves exhibited substantial variance (see Fig. 2). AEP thresholds measured with jawphones were more variable than the direct field behavioral thresholds; however, variability in AEP thresholds seems similar to that observed in direct field measurements of EFR thresholds (Yuen *et al.*, 2005). The degree to which the jawphone calibration measurements match the effective stimulus to the subject is unknown and is likely frequency dependent.

The limitations in the jawphone calibration mean that direct comparisons between sensitivities assessed via jawphone and direct/free field are required to validate the calibrations for each specific jawphone. Inherent calibration problems may limit the utility of absolute threshold values obtained with jawphones. For example, in a previous assess-

ment of hearing sensitivity using jawphones (Brill *et al.*, 2001), the tonal stimulus was calibrated with a hydrophone 40 cm away from the sound source. This distance was 25 cm greater than that used in this study and exceeds the pathway distance from the pan (point of jawphone attachment) to the auditory bulla. However, within the same subject, differences between underwater behavioral thresholds measured with the jawphone and free transducer (Brill *et al.*, 2001) were within the range of differences between direct field behavioral and AEP thresholds observed in this study. Most important, direct field behavioral and jawphone AEP thresholds have agreed closely as to the shape of the audiogram and the upper cutoff frequency. This suggests that jawphone testing may be an acceptable substitute for underwater direct field measurements in many applications (e.g., hearing screening of a rehabilitating animal prior to release) where audiogram shape and cutoff frequency are of primary interest. In many environments, such as in-air testing of stranded cetaceans, jawphones may be the only practical method for testing higher frequencies and obtaining results comparable to those obtained in underwater direct field conditions.

The use of evoked potentials for odontocete audiometry has been practiced for several decades. The use of tone pips for audiometry predates the use of EFR (Popov and Supin, 1987, 1990; Szymanski *et al.*, 1999). The latter technique has been gaining favor in recent years because it produces a stimulus with narrower spectral spreading than does a tone pip. AEP audiograms obtained with tone pips as stimuli have been compared to behavioral audiograms in the same animal in killer whales (*Orcinus orca*; Szymanski *et al.*, 1999). This work demonstrated that electrophysiological estimates of hearing sensitivity obtained with tone pips range from 3 to 41 dB of the behavioral threshold, with average differences of ±8 dB. The procedure was useful in identifying the regions of best sensitivity and the upper limits of hearing. Qualitatively, both AEP audiograms obtained with tone pips and with EFR appear to show similar relationships to behavioral audiograms. Whether the frequency specificity provided by the EFR method provides noticeable gains given the variability inherent in evoked potential measures remains to be determined.

Estimates of hearing sensitivity in delphinids from recordings of the EFR show a similar degree of accuracy and precision relative to behavioral thresholds, regardless of the exact methodology employed for stimulus delivery (this study; Yuen *et al.*, 2005; Finneran and Houser, 2006). More promising to the future application of EFR methods to other marine mammal species is that the variation in accuracy and precision relative to behavioral thresholds is similar to that observed in human clinical studies (Lins *et al.*, 1995; Rance *et al.*, 1995; Aoyagi *et al.*, 1999; Dimitrijevic *et al.*, 2002; Vander Werff and Brown, 2005). The EFR estimation of delphinid hearing seems robust given its results under various modifications of the testing procedures. This characteristic of the EFR estimation technique should make it a useful tool in future efforts of meeting the NRC recommendation (NRC, 1994; 2000) to increase our knowledge base on marine mammal hearing through the testing of stranded or rehabilitating marine mammals.

V. CONCLUSIONS

(1) Underwater AEP measurements with a jawphone transducer produced comparable data to in-air AEP measurements with jawphones.
(2) Underwater or in-air AEP measurements using a jawphone may be a suitable proxy for direct field behavioral/AEP measurements in many circumstances.

ACKNOWLEDGMENTS

We thank R. Dear and C. Schlundt for help in data collection and analysis and L. Green, C. Quintos, L. Lewis, and E. Linberg for animal training. S. Ridgway, D. Carder, and P. Moore provided much helpful advice and technical assistance. Financial support was provided by the U.S. Office of Naval Research, Marine Mammal S&T Program and the SSC San Diego In-house Laboratory, Independent Research (ILIR) program.

Amos, D. E., and Koopmans, L. H. (**1963**). "Tables of the Distribution of the Coefficient of Coherence for Stationary Bivariate Gaussian Processes," Sandia Corporation, Livermore, CA, p. 328.

Andre, M., Supin, A., Delory, E., Kamminga, C., Degollada, E., and Alonso, J. M. (**2003**). "Evidence of deafness in a striped dolphin, Stenella coeruleoalba," Aquat. Mamm. **29**, 3–8.

Aoyagi, M., Suzuki, K., Yokota, M., Furuse, H., Watanabe, T., and Ito, T. (**1999**). "Reliability of 80-Hz amplitude-modulation-following response detected by phase coherence," Audiol. Neuro-Otol. **4**, 28–37.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer, New York).

Au, W. W. L., and Moore, P. W. B. (**1990**). "Critical ratio and critical bandwidth for the Atlantic bottlenosed dolphin," J. Acoust. Soc. Am. **88**, 1635–1638.

Brill, R. L., Moore, P. W. B., and Dankiewicz, L. A. (**2001**). "Assessment of dolphin (*Tursiops truncatus*) auditory sensitivity and hearing loss using jawphones," J. Acoust. Soc. Am. **109**, 1717–1722.

Brill, R. L., Sevenich, M. L., Sullivan, T. J., Sustman, J. D., and Witt, R. E. (**1988**). "Behavioral evidence for hearing through the lower jaw by an echolocating dolphin (*Tursiops truncatus*)," Marine Mammal Sci. **4**, 223–230.

Brillinger, D. R. (**1978**). "A note on the estimation of evoked response," Biol. Cybern. **31**, 141–144.

Bullock, T. H., Grinnell, A. D., Ikezono, E., Kameda, K., Katsuki, K., Nomoto, M., Sato, O., Suga, N., and Yanagisawa, K. (**1968**). "Electrophysiological studies of central auditory mechanisms in cetaceans," Zeitschrift fur Vergleichende Physiologie **59**, 117–156.

Cook, M. L. H., Varela, R. A., Goldstein, J. D., McCulloch, S. D., Bossart, G. D., Finneran, J. J., Houser, D., and Mann, D. A. (**2006**). "Beaked whale auditory evoked potential hearing measurements," J. Comp. Physiol., A **192**, 489–495.

Dimitrijevic, A., John, M. S., Van Roon, P., Purcell, D. W., Adamonis, J., Ostroff, J., Nedzelski, J. M., and Picton, T. W. (**2002**). "Estimating the audiogram using multiple auditory steady-state responses," J. Am. Acad. Audiol **13**, 205–224.

Dobie, R. A. (**1993**). "Objective response detection," Ear Hear. **14**, 31–35.

Dobie, R. A., and Wilson, M. J. (**1989**). "Analysis of auditory evoked potentials by magnitude-squared coherence," Ear Hear. **10**, 2–13.

Dobie, R. A., and Wilson, M. J. (**1996**). "A comparison of t test, F test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids," J. Acoust. Soc. Am. **100**, 2236–2246.

Dolphin, W. F. (**2000**). "Electrophysiological measures of auditory processing in odontocetes," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York), pp. 294–329.

Dolphin, W. F., Au, W. W., Nachtigall, P. E., and Pawloski, J. (**1995**). "Modulation rate transfer functions to low-frequency carriers in three species of cetaceans," J. Comp. Physiol., A **177**, 235–245.

Egan, J. P., Greenberg, G. Z., and Schulman, A. I. (**1961**). "Operating characteristics, signal detectability, and the method of free response," J.

Acoust. Soc. Am. **33**, 993–1007.
Finneran, J. J., Carder, D. A., Schlundt, C. E., and Ridgway, S. H. (**2005**). "Temporary threshold shift (TTS) in bottlenose dolphins (*Tursiops truncatus*) exposed to mid-frequency tones," J. Acoust. Soc. Am. **118**, 2696–2705.
Finneran, J. J., and Houser, D. S. (**2006**). "Comparison of in-air evoked potential and underwater behavioral hearing thresholds in four bottlenose dolphins (*Tursiops truncatus*)," J. Acoust. Soc. Am. **119**, 3181–3192.
Finneran, J. J., Schlundt, C. E., Carder, D. A., and Ridgway, S. H. (**2002**). "Auditory filter shapes for the bottlenose dolphin (*Tursiops truncatus*) and the white whale (*Delphinapterus leucas*) derived with notched noise," J. Acoust. Soc. Am. **112**, 322–328.
Houser, D. S., Finneran, J. J., Carder, D. A., Van Bonn, W., Smith, C. R., Hoh, C., Mattrey, R., and Ridgway, S. H. (**2004**). "Structural and functional imaging of bottlenose dolphin (*Tursiops truncatus*) cranial anatomy," J. Exp. Biol. **207**, 3657–3665.
Klishin, V. O., Popov, V. V., and Supin, A. Y. (**2000**). "Hearing capabilities of a beluga whale, *Delphinapterus leucas*," Aquat. Mamm. **26**, 212–228.
Lins, O. G., Picton, P. E., Picton, T. W., Champagne, S. C., and Durieux-Smith, A. (**1995**). "Auditory steady-state responses to tones amplitude-modulated at 80-110 Hz," J. Acoust. Soc. Am. **97**, 3051–3063.
McCormick, J. G., Wever, E. G., Palin, J., and Ridgway, S. H. (**1970**). "Sound conduction in the dolphin ear," J. Acoust. Soc. Am. **48**, 1418–1428.
McCormick, J. G., Wever, E. G., Ridgway, S. H., and Palin, J. (**1980**). "Sound reception in the porpoise as it relates to echolocation," in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York).
Møhl, B., Au, W. W. L., Pawloski, J., and Nachtigall, P. E. (**1999**). "Dolphin hearing: Relative sensitivity as a function of point of application of a contact sound source in the jaw and head region," J. Acoust. Soc. Am. **105**, 3421–3424.
Mooney, T. A., and Nachtigall, P. E. (**2006**). "Temporal resolution of the Risso's dolphin, *Grampus griseus*, auditory system," J. Comp. Physiol. [A] **192**, 373–380.
Nachtigall, P. E., Yuen, M. M. L., Mooney, T. A., and Taylor, K. A. (**2005**). "Hearing measurements from a stranded infant Risso's dolphin, *Grampus griseus*," J. Exp. Biol. **208**, 4181–4188.
National Research Council (NRC) (**1994**). *Low-Frequency Sound and Marine Mammals: Current Knowledge and Research Needs* (National Academy Press, Washington, DC).
National Research Council (NRC) (**2000**). *Marine Mammals and Low-Frequency Sound: Progress Since 1994* (National Academy Press, Washington, DC).
Popov, V. V., and Supin, A. Y. (**1987**). "Characteristics of hearing in the beluga *Delphinapterus leucas*," Dokl. Akad. Nauk SSSR **294**, 1255–1258.
Popov, V. V., and Supin, A. Y. (**1990**). "Electrophysiological investigation of hearing in the freshwater dolphin *Inia geoffrensis*," Dokl. Biol. Sci. **313**, 238–241.
Popov, V. V., and Supin, A. Y. (**1998**). "Auditory evoked responses to rhythmic sound pulses in dolphins," J. Comp. Physiol. [A] **183**, 519–524.
Popov, V. V., Supin, A. Y., Wang, D., Wank, K., Xiao, J., and Li, S. (**2005**). "Evoked-potential audiogram of the Yangtze finless porpoise *Neophocaena phocaenoides asiaeorientalis* (L)," J. Acoust. Soc. Am. **117**, 2728–2731.
Rance, G., Rickards, F. W., Cohen, L. T., De Vidi, S., and Clark, G. M. (**1995**). "The automated prediction of hearing thresholds in sleeping subjects using auditory steady-state evoked potentials," Ear Hear. **16**, 499–507.
Supin, A. Y., and Popov, V. V. (**1995**). "Envelope-following response and modulation transfer function in the dolphin's auditory system," Hear. Res. **92**, 38–46.
Supin, A. Y., and Popov, V. V. (**2000**). "Frequency-modulation sensitivity in bottlenose dolphins, *Tursiops truncatus*: Evoked-potential study," Aquat. Mamm. **26**, 83–94.
Supin, A. Y., Popov, V. V., and Mass, A. M. (**2001**). *The Sensory Physiology of Aquatic Mammals* (Kluwer Academic, Boston, MA).
Szymanski, M. D., Bain, D. E., Kiehl, K., Pennington, S., Wong, S., and Henry, K. R. (**1999**). "Killer whale (*Orcinus orca*) hearing: Auditory brainstem response and behavioral audiograms," J. Acoust. Soc. Am. **106**, 1134–1141.
Vander Werff, K. R., and Brown, C. J. (**2005**). "Effect of audiometric configuration on threshold and suprathreshold auditory steady-state responses," Ear Hear. **26**, 310–326.
Yuen, M. M. L., Nachtigall, P. E., Breese, M., and Supin, A. Y. (**2005**). "Behavioral and auditory evoked potential audiograms of a false killer whale (*Pseudorca crassidens*)," J. Acoust. Soc. Am. **118**, 2688–2695.

A simulation study on tissue harmonic imaging with a single-element intravascular ultrasound catheter

Martijn E. Frijlink
Biomedical Engineering, Erasmus MC, University Medical Center Rotterdam, the Netherlands

David E. Goertz
Biomedical Engineering, Erasmus MC, University Medical Center Rotterdam, the Netherlands, and Interuniversity Cardiology Institute of the Netherlands, Utrecht, the Netherlands

Ayache Bouakaz
Inserm U619 and Université François Rabelais, Tours, France

Antonius F. W. van der Steen
Biomedical Engineering, Erasmus MC, University Medical Center Rotterdam, the Netherlands, and Interuniversity Cardiology Institute of the Netherlands, Utrecht, the Netherlands

(Received 30 December 2005; revised 21 June 2006; accepted 21 June 2006)

Recently, *in vivo* feasibility of tissue harmonic imaging with a mechanically rotated intravascular ultrasound (IVUS) catheter was experimentally demonstrated. To isolate the second harmonic signal content, a combination of pulse inversion and analog filtering was used. In this paper the development of a simulation tool to investigate nonlinear IVUS beams is reported, and the influence of transducer rotation and axial catheter-to-tissue motion on the efficiency of PI signal processing is evaluated. Nonlinear beams were simulated in homogeneous tissue-mimicking media at a transmit frequency of 20 MHz, which resulted in second harmonic pressure fields at 40 MHz. The competing effects of averaging and decorrelation between neighboring rf lines on the signal-to-noise ratio (SNR) were studied for a single point scatterer. An optimal SNR was achieved when lines were combined over 3°–3.75°. When the transducer was rotated with respect to point scatterers, simulating the acoustic response of tissue, the fundamental frequency suppression using PI degraded rapidly with increasing interpulse angles. The effect of axial catheter-to-tissue motion on the efficiency of pulse inversion seemed to be of less influence for realistic motion values. The results of this study will aid in the optimization of harmonic IVUS imaging systems. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2226069]

PACS number(s): 43.80.Vj [CCC] Pages: 1723–1731

I. INTRODUCTION

Intravascular ultrasound (IVUS) is capable of providing high resolution real time cross-sectional images of coronary arteries and it has therefore become an important clinical tool for the detection and evaluation of coronary artery diseases, as well as for therapy guidance and clinical research (Saijo and van der Steen, 2003). At present, clinicians use both rotating single-element IVUS catheters with center frequencies between 30 and 40 MHz, and 64-element phased-array IVUS catheters with a center frequency of 20 MHz.

At frequencies below 10 MHz, tissue harmonic imaging (THI) has been shown to increase the diagnostic value of medical ultrasound through improvements in image quality. An initial study showed the feasibility of THI at high frequencies (20–40 MHz) (van der Steen *et al.*, 1999). Cherin *et al.* (2002) described the experimental characterization of fundamental and second harmonic beams of a focused ultrasound biomicroscopy (UBM) transducer in the 20 to 40 MHz range. The reduction of stent imaging artifacts was shown when high frequency THI (transmit f_c=20 MHz, receive f_c=40 MHz) was applied with a similar focused UBM transducer (Frijlink *et al.*, 2005). More recently, we demonstrated *in vivo* the feasibility of THI for an IVUS system using a 20–40 MHz mechanically rotated IVUS catheter (Frijlink *et al.*, 2004). In that study, tissue harmonic signals were isolated using a combination of analog filtering and pulse inversion (PI).

The PI technique requires at least one firing of a pulse and its inverted counterpart along each line of sight (Hope Simpson *et al.*, 1999). In the presence of relative motion between the tissue and catheter, the fundamental frequency signal (i.e., transmit bandwidth signal) is not completely canceled, and the harmonic intensity becomes smaller due to signal decorrelation. Motion artifacts of PI-based THI have been studied by Shen and Li (2002), whose results indicated that the performance of tissue harmonic imaging can be significantly affected by tissue motion. In particular, it was found that for axial motion, both the fundamental reduction and the tissue harmonic intensity decreased much more rapidly than with lateral motion. This study was conducted under conditions relevant to low frequency array based scanning. These conditions depart significantly from the case of harmonic IVUS systems, where a rapidly rotating single-element transducer is imaging in the presence of significant cardiac tissue motion.

Nonlinear ultrasound beams and their exploitation in the context of THI have been extensively investigated at lower diagnostic ultrasound frequencies (Duck, 2002; Humphrey, 2000; Averkiou *et al.*, 1997). A number of different approaches have been developed to model nonlinear propagation in the field of an ultrasonic transducer. The most common technique is to solve the Khokhlov-Zabolotskaya-Kuznetsov (KZK) equation, which is a nonlinear parabolic equation that accounts for the combined effects of diffraction, absorption, and nonlinearity for directional sound beams (Hamilton and Blackstock, 1998). Comparisons of nonlinear simulation results and measurements show a very high level of agreement (Duck, 2002; Humphrey, 2000; Averkiou and Hamilton, 1997). For relatively short pulses, as typically used in diagnostic ultrasound, the time-domain implementation of the KZK equation is advantageous because many harmonic components could then be taken into account for the relatively broadband imaging pulses (Hamilton and Blackstock, 1998). At high frequencies, both nonlinear tissue properties and nonlinear ultrasound beams remain relatively unexplored.

In this study we simulated fundamental 20 MHz (F20), second harmonic (H40), and fundamental 40 MHz (F40) fields for an unfocused circular IVUS element using a two-dimensional nonlinear parabolic KZK equation and medium characteristics (attenuation, nonlinearity, scattering) in the range of those of vascular tissue and blood. The pulse-echo responses from a single point-scatterer and a cloud of point-scatterers were calculated for successive inverted pulses as the beam was rotated. The influence of rotation, and the effect of axial tissue-to-catheter motion on the performance of PI was then investigated for different interpulse angles.

II. METHODS

A. Simulation design

1. Nonlinear propagation

The proposed simulation method resembles the method as described by Li and Zagzebski (2000). The simulation of the transmitted nonlinear field by an unfocused circular transducer is based on a time-domain implementation of the KZK equation based on the numerical approach of Lee and Hamilton (1995). A two-dimensional implementation has been written in FORTRAN and has been evaluated (Bouakaz *et al.*, 2003). The attenuation of acoustic waves propagating in a wide variety of lossy media obeys a power law dependence on the frequency of the general form (Szabo, 1994),

$$\alpha = \alpha_0 |2\pi\nu|^{\gamma}, \tag{1}$$

where ν is the frequency, and α_0 and γ are arbitrary real non-negative constants. Most biologic fluids and tissues have power law exponents in the $1 \leqslant \gamma \leqslant 2$ range (Duck, 1990). The algorithm described by Bouakaz *et al.* (2003) was modified to account for a frequency dependent attenuation different than a power law exponent of 2 (which corresponds to that of attenuation in water). This modification was implemented in the frequency domain after a Fast Fourier Transformation (FFT). Frequency dependent attenuation factors were then multiplied with individual frequency components. By means of an inverse FFT, a hybrid time and frequency domain algorithm was created.

2. Pulse echo scattering

The received signal from an individual scatterer is calculated using an analytically derived spatial impulse response $h(r,z)$ for an unfocused circular transducer (Stephanishen, 1971):

$$h(r,z{:}t) = c_0, \quad \text{if } r < a \text{ and } t_0 < t < t_1;$$
$$h(r,z{:}t) = \frac{c_0}{\pi} \cos^{-1}\left(\frac{r^2 - a^2 - z^2 + c_0^2 t^2}{2r\sqrt{c_0^2 t^2 - z^2}} \right), \quad \text{if } t_1 < t < t_2;$$
$$h(r,z{:}t) = 0, \quad \text{elsewhere},$$

in which

$$t_0 = z/c_0$$
$$t_1 = \frac{\sqrt{z^2 + (r-a)^2}}{c_0},$$
$$t_2 = \frac{\sqrt{z^2 + (r+a)^2}}{c_0}, \tag{2}$$

where c_0 is the velocity of propagation within the medium, z is the axial coordinate of the field point, r is the radial coordinate of the field point, a is the radius of the piston transducer, and t is time. Linear field calculations for a circular transducer with this analytically derived impulse response function showed a very high degree of similarity with acoustic fields calculated with the numerical ultrasound simulation program Field II (Jensen and Svendsen, 1992), which uses an approximation based on the summation of responses from small rectangles to calculate pressure fields from arbitrary shaped transducers using linear acoustics. The benefit of the analytically derived expression for the impulse response is that it is computationally efficient.

In order to be able to use this spatial impulse response in combination with propagation in a frequency dependent attenuating medium, the distance from an individual point scatterer to the piston transducer is approximated by a single value, similar to the approach described by Jensen *et al.* (1993), which used the on-axis distance (z). Since signals in attenuating media exponentially decay with distance, the part of the transducer closest to the field point contributes most to the signal amplitude of this field point. Therefore, the shortest distance from a point scatterer to the transducer, z_{att}, is used for attenuation calculations and is given by

$$z_{\text{att}} = c_0 \cdot t_0 = z, \quad \text{if } r < a,$$
$$z_{\text{att}} = c_0 \cdot t_1, \quad \text{if } r \geqslant a, \tag{3}$$

where t_0 and t_1 are defined in (2). Comparisons of this approximation with a numerical spatial impulse response method where frequency dependent attenuation was included showed a very high degree of similarity. For field points where $z \geqslant a/2$, amplitude differences of rf signals were <0.5 dB, calculated for attenuating media with a linear frequency dependency.

For the purposes of simulating ultrasound backscattered signals, tissue can be represented by many point scatterers positioned randomly in three dimensions (Kerr and Hunt, 1992; Hunt *et al.*, 1995). Scatterers were placed at random locations, thus generally in between grid locations where the transmitted field was calculated. All point scatterers were assigned the same scattering strength. Frequency dependent backscatter (μ_b) was taken into account by the approximation (Lockwood *et al.*, 1991)

$$\mu_b = \mu_0 \cdot \nu^{\gamma}, \tag{4}$$

where μ_0 is the backscatter coefficient at 1 MHz, ν is the frequency expressed in MHz, and γ represents the frequency dependence. The individual response from a scatterer was calculated by a convolution of the spatial impulse response with the time-domain wave as calculated with the KZK equation. The backscatter signal from the cloud of scatterers is calculated by a summation of the individual responses from each scatterer. The scatterer density (1250 scatterers/mm^3) was sufficient to produce Rayleigh envelope statistics.

The three-dimensional scatterer volume could be rotated with respect to the transmitted field to simulate catheter rotation of a mechanically scanned IVUS transducer.

For the simulations described in this paper, the following assumptions and simplifications have been made.

(1.) The sound-propagating medium is uniform and lossy, and dispersion is negligible.
(2.) The medium is assumed to be directly adjacent to the transducer surface.
(3.) Tissue is assumed to be a weak scattering medium (Born approximation).
(4.) Since weak scattering is assumed, the backscattered low-amplitude signal propagation is approximated by linear propagation.
(5.) The amplitudes of the backscattered signals are all relative, so no absolute pressures are given.
(6.) The catheter motion was assumed to be negligible during the transmit and receive of a single firing.
(7.) Motion other than catheter rotation was assumed to be not present, unless explicitly mentioned.

B. Simulation parameters

The nonlinear beams were calculated for a circular, unfocused transducer with a diameter of 0.9 mm, similar to elements used in previous IVUS THI studies (Frijlink *et al.*, 2004). Since this element was centered at 30 MHz, the approach to attempt to use the lower and upper parts of the bandwidth for fundamental and harmonic imaging explained the choice of 20 MHz (fundamental) and 40 MHz (fundamental and second harmonic). Due to the circular geometry, the three-dimensional field was calculated by a simulation in cylindrical coordinates. In all calculations, the excitation pulse was a Gaussian enveloped sine wave with a −6-dB fractional bandwidth of 30%.

The propagation medium characteristics were chosen to be in the range of those of vascular tissue and blood. Arterial tissue and blood have an attenuation frequency dependency that can be described by Eq. (1), with power law exponents of $\gamma=1.1$ and $\gamma=1.2$, respectively, in the frequency range 15–60 MHz (Lockwood *et al.*, 1991; Foster *et al.*, 2000). In the 10–50 MHz range, this is approximated by attenuation that has a linear frequency dependency in the range from 0.5 to 1.5 dB/cm/MHz. In this study, two different frequency dependent attenuation values of 0.6 and 1.0 dB/cm/MHz were used. The sound speed of the propagation medium was chosen to be 1560 m/s and the mass density was set at 1060 kg/m^3, both corresponding to published values of human tissue and human blood (Duck, 1990). The nonlinear parameter (B/A) (Hamilton and Blackstock, 1998) of tissue is estimated at six to seven and that of human blood 6.0 (Duck, 1990), therefore the B/A constant was chosen to be 6.0. These values appear not to have been measured at high frequencies (>10 MHz).

Since all scatterers are assigned the same scattering strength, we simulate soft tissue-like media only. Extended analysis would be necessary when locally highly scattering objects (e.g., calcified material, bones, etc.) are taken into account. The backscatter signal from scatterers is calculated by a summation of responses from all scatterers that are within the −20-dB beam width (−40-dB two way). In our case this means that all scatterers within a radial extent of 0.5 mm from the z axis (corresponding to the propagation axis) are taken into account. In the depth direction, scatterers located between $z=0.5$ and $z=6.0$ mm are included in the calculations. There are no scatterers assumed to be close (<0.5 mm) to the transducer, because of the water surrounding the transducer element within the protective sheath. Lockwood *et al.* (1991) showed that the power law exponent of the frequency dependent backscatter in the artery wall ranged from $\gamma=1.1$ to 1.4 at high frequencies, and that this parameter in blood ranged from $\gamma=1.3$ to 1.4 at the same frequency range. A frequency dependence of $\gamma=1.3$ has been chosen in this study.

C. Simulations

1. Nonlinear beam simulations

Two-dimensional beam profiles were calculated for F20 and H40 mode through a medium with an attenuation of 1.0 dB/cm/MHz. For comparison, a beam profile of a fundamental 40 MHz pulse (F40) was also calculated. To study the influence of attenuation and excitation amplitudes on nonlinear generation, 20 MHz input pulses were propagated through two different attenuating media at peak envelope pressures ranging from 250 kPa to 4 MPa. F20 and H40 signal contents were isolated by digital filtering.

Penetration depth is an important issue in diagnostic ultrasound imaging, with attenuation effects limiting the signal strengths with increasing propagation distance. Since higher frequencies are attenuated more, the absolute penetration depth of a F40 field is less than that of a F20 field for identical excitation pressures. The penetration depth for the second harmonic (H40) field is more complicated since harmonic energy is both attenuated and generated during propagation. Since the harmonic generation is highly dependent on the fundamental amplitude, the absolute penetration

depth will be highest for high F20 excitation. The effective penetration depth of a F40 beam is therefore compared to an H40 beam by simulating a high amplitude (5 MPa) F40 beam and an H40 signal from a high amplitude (5 MPa) 20 MHz pulse through an attenuating medium of 1.0 dB/cm/MHz.

2. Effects of rotation on averaging and decorrelation

As previously reported (Frijlink *et al.*, 2004), averaging of neighboring rf lines was used to increase the signal-to-noise ratio (SNR). This assumes that the decorrelation effect on the H40 tissue signal from catheter rotation was small compared to uncorrelated noise. The competing effect of averaging and decorrelation by rotation on the SNR is investigated by pulse-echo simulations with a single point scatterer located 4 mm from the transducer face. The SNR is defined as the integrated rf power of the signal only, divided by the integrated rf power of the noise. White Gaussian noise; characterized by a flat frequency spectrum and random phase, was generated by a random number generator. Both signal and noise were digitally bandpass filtered (32–50 MHz, fifth order Butterworth). The catheter rotation was simulated by rotating the scatterer with respect to the beam over incremental angles of 0.15°, which corresponded to the experimentally employed line density of 2400 rf lines per rotation. Different background noise levels were added to calculate rf lines, and the results were digitally filtered. Next, the SNR difference was calculated as a function of the number of averages. This SNR difference was then normalized for each different background noise level separately.

The point-spread function (PSF) was calculated for a single ideal point target at 4 mm from the transducer through an attenuating medium (1.0 dB/cm/MHz) for an F20, H40, and F40 field. The single-element transducer was rotated around its center (as indicated in Fig. 1) with an angle of 0.15° between each line of sight.

3. Effects of rotation on pulse inversion

The fundamental frequency signal will not be completely canceled with PI when motion between the tissue and catheter is present, which is explicitly shown for IVUS by Goertz *et al.* (2006). Nevertheless, this will be the situation when PI is performed with a single-element continuously rotating IVUS catheter. To study the effect of catheter rotation on PI, two inverted 20 MHz pulses (0° and 180°) were propagated through the attenuating medium as the beam rotated past a three-dimensional (3-D) volume of point scatterers (Fig. 1). The incremental interpulse angle was again 0.15°. Backscatter signals from the 3-D volume of point scatterers have been calculated to simulate rf pulse-echo responses for different angles with respect to the IVUS transducer. A dataset consisting of 32 different realizations of a volume of randomly positioned scatterers was calculated for a F20 2 MPa amplitude pulse pair. The average cross-correlation value between rf lines was calculated as a function of the interpulse angle for raw and filtered rf lines (32–50 MHz, fifth order Butterworth). For comparison, the average cross-correlation for F40 rf lines has also been calculated. Fundamental suppression by PI was estimated at different distances from the transducer as a function of the interpulse angle, assuming that the rotation speed is uniform. The mean fundamental signal content of pulse pairs was calculated by summing the frequency power in the 18–22 MHz range.

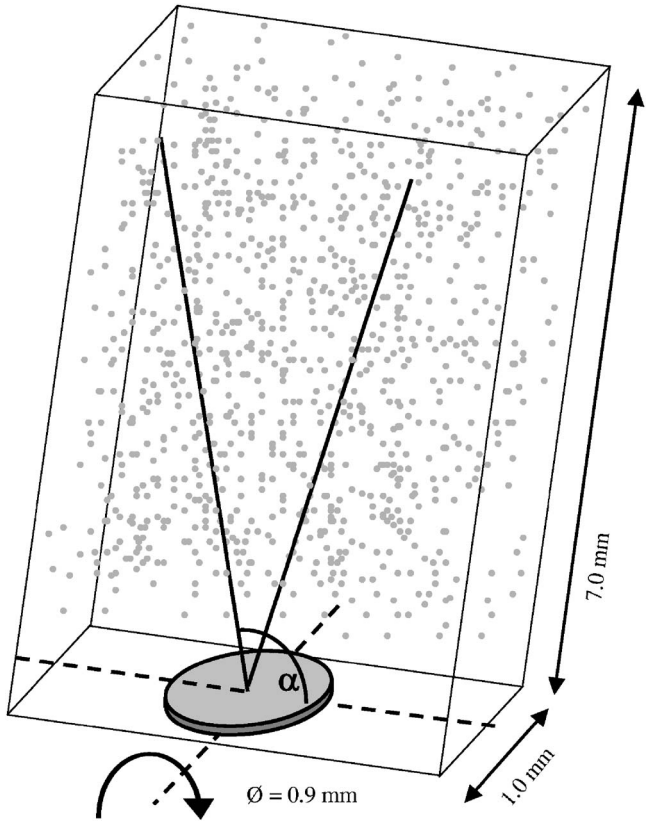


FIG. 1. Schematic simulation set-up, showing randomly distributed point-scatterers in a 3-D volume relative to a rotating transducer.

4. Effects of axial motion on pulse inversion

Relative catheter and tissue motion due to effects other than element rotation is another potential cause for the additional decorrelation between successive rf lines. As investi-

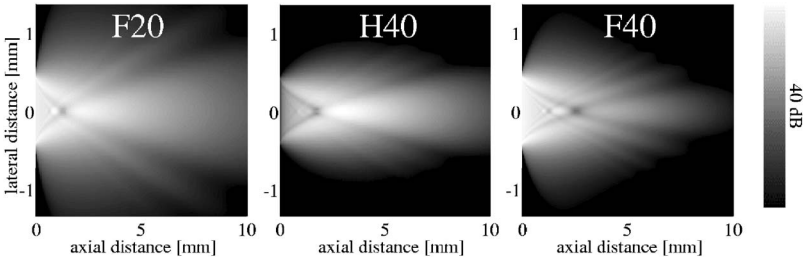


FIG. 2. Two-dimensional beam profiles in the F20, H40, and F40 modes. The beam profiles have been normalized with respect to the maximum signal within individual images.

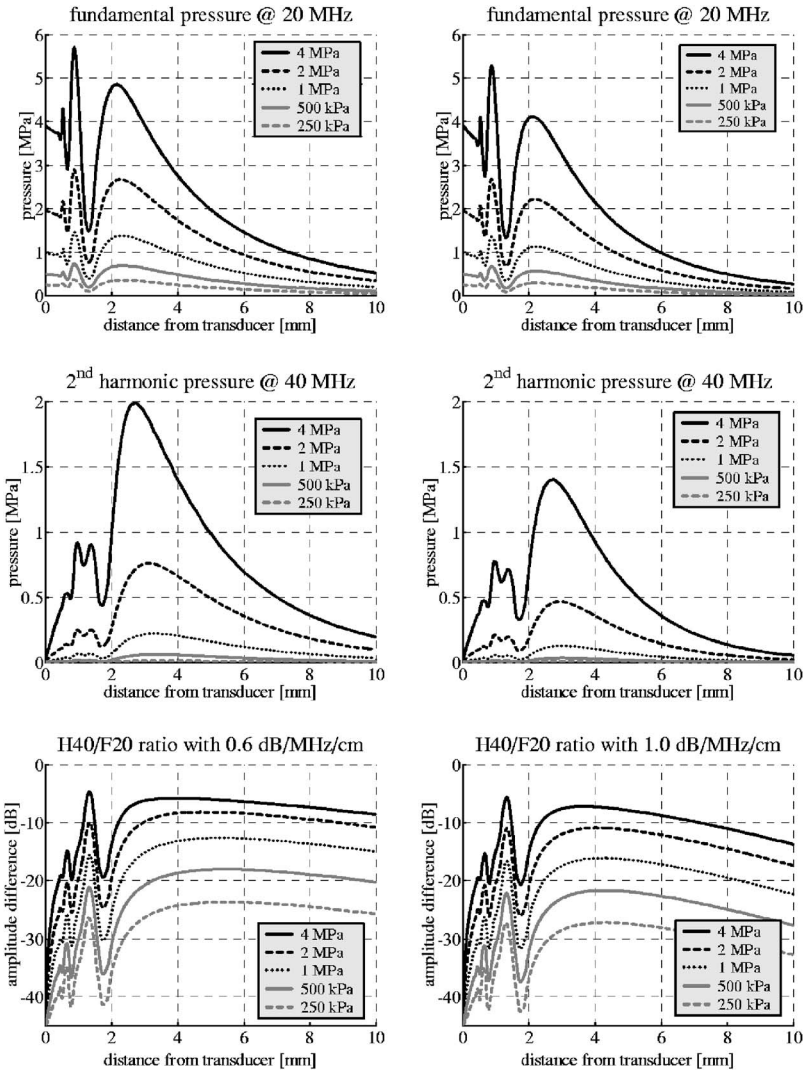


FIG. 3. Fundamental F20 (top) and second harmonic H40 (middle) on-axis pressures and the H40/F20 ratios (bottom) for two different attenuating media (0.6 and 1.0 db/cm/MHz).

gated by Shen and Li (2002), axial motion, resulting in a phase/time shift, is much more severe than lateral motion for degrading the efficiency of PI. Therefore we simulated axial catheter-to-tissue motion (perpendicular to the transducer surface) to investigate the effect on fundamental suppression by PI at fixed interpulse angles. The calculated axial catheter-to-tissue motion is based on a PRF of 12.5 kHz, corresponding to a previously reported PRF (Frijlink *et al.*, 2004). The mean fundamental signal content of pulse pairs was calculated by summing the frequency power in the 18–22 MHz range.

III. RESULTS

1. Nonlinear beam simulations

Example two-dimensional beam profiles for the F20, H40, and F40 modes are plotted in Fig. 2. These images are

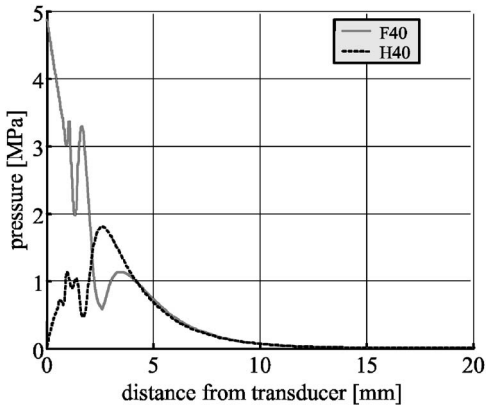


FIG. 4. F40 and H40 (fundamental excitation pressure=5 MPa) on-axis pressures in an attenuating medium (1.0 dB/cm/MHz).

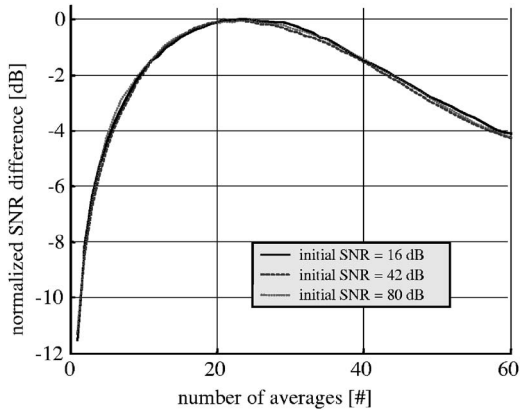


FIG. 5. The normalized SNR effect of averaging neighboring noisy rf-pulse-echo responses from a single scatterer 4 mm from the transducer.

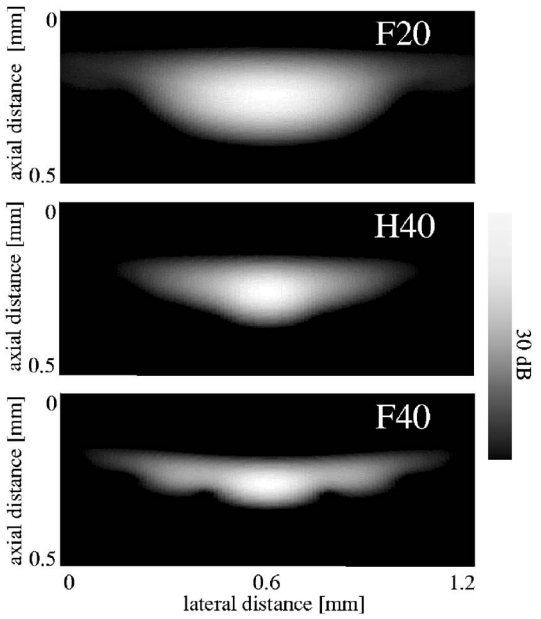


FIG. 6. Point Spread Functions of F20, H40, and F40 of an ideal point scatterer at 4 mm distance from the transducer in an attenuating medium (1.0 dB/cm/MHz).

normalized with respect to the maximum signal within the individual images. The frequency dependent attenuating medium causes the F40 beam to decay faster than both the F20 and H40 beams. Close to the transducer, the H40 field exhibits a low intensity compared to the F20 and F40 field. Figure 3 shows on-axis fundamental and second harmonic pressures of the 20 MHz pulse at five different amplitudes, propagated through two different attenuating media. The frequency dependent attenuation in these simulations was 0.6 (left column) and 1.0 dB/cm/MHz (right column). The second harmonic pressure can be seen to build up progressively to reach a maximum pressure around 3 mm. From these figures the amplitude dependent second harmonic generation is evident. The ratio of the second harmonic at 40 MHz and the fundamental can also be seen in Fig. 3 (bottom). Due to the increased attenuation, the H40/F20 ratio exhibits a faster decrease from approximately 3 mm than in the medium with the lower 0.6 dB/cm/MHz attenuation.

To compare the penetration depth between F40 and H40,

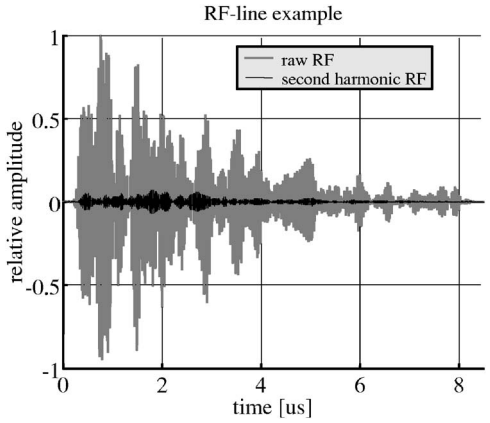


FIG. 7. Typical raw and filtered (H40) rf responses.

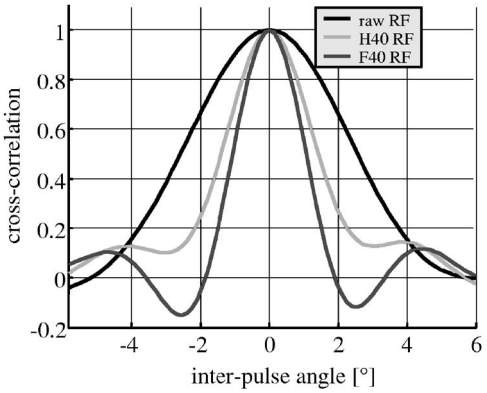


FIG. 8. Cross-correlation between raw, H40, and F40 rf lines at a distance from 2 to 3 mm from the transducer, plotted as a function of the interpulse angle.

the absolute on-axis pressures of both beams are plotted in Fig. 4. For both pulses the fundamental excitation amplitude was 5 MPa. From this graph it is clear that between 2 and 4 mm (imaging depth), which corresponds to the near field of the F40 beam, the absolute on-axis pressure of the second harmonic is higher than the F40. Above 4 mm the absolute pressures are similar. This simulation suggests that the H40 mode could result in a similar effective penetration depth compared to the F40 mode.

2. Effects of rotation on averaging and decorrelation

For a single scatterer located 4 mm from the transducer, the normalized SNR difference is plotted logarithmically as a function of averaging in Fig. 5 for different noise levels (the SNR is 16, 42, and 80 dB). These curves show the effect of averaging neighboring rf lines around a center rf line receiving the highest amplitude signal from the scatterer. The competing effects of averaging and decorrelation produce a local maximum in SNR when 20–25 neighboring lines (corresponding to 3°–3.75°) were combined.

Point spread functions for a single scatterer at 4 mm from the transducer are depicted in Fig. 6. These PSFs are normalized to the maximum signal level in each individual image. The different appearance can be partly attributed to the diffraction pattern of the different fields (see Fig. 2). The −6 dB lateral beamwidths of the F20, H40, and F40 PSFs are

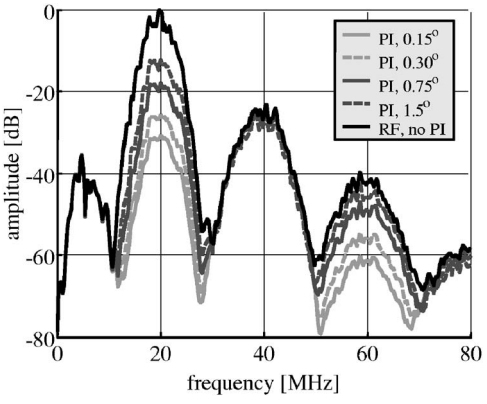


FIG. 9. Average frequency spectra of a single rf line (solid black) and of PI pairs with different interpulse angles.

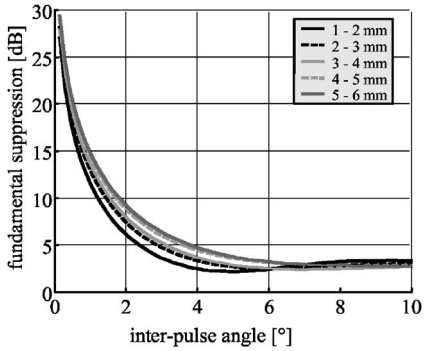


FIG. 10. The fundamental suppression with PI as a function of the interpulse angle as calculated for a 2 MPa pulse pair calculated for different backscatter distances from the transducer.

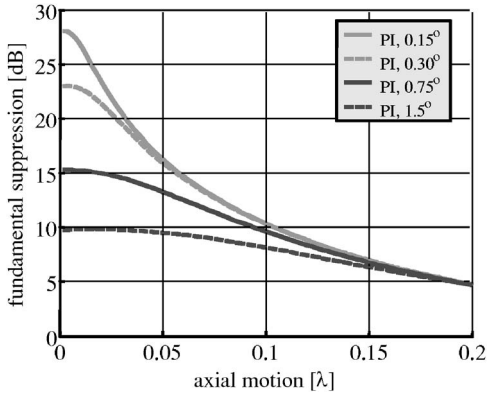


FIG. 12. The fundamental suppression with PI as a function of axial catheter-to-tissue motion at a distance from 2 to 3 mm from the transducer for different interpulse angles.

0.39, 0.23, and 0.23 mm, respectively. PSFs at different distances from the transducer will differ considerably, though Fig. 6 gives a clarifying example of the PSF differences between F20, H40, and F40 modes. Note that the number of averages of neighboring lines will increase the effective beam width and therefore will degrade lateral resolution.

3. Effects of rotation on pulse inversion

Nonlinear backscatter signals have been calculated to simulate the rf pulse-echo responses from different angles with respect to the IVUS transducer. A typical pulse-echo response can be seen in Fig. 7. The digitally bandpassed second harmonic signal of this rf line is indicated in the same figure. The average interpulse cross-correlation estimates for the raw, H40, and F40 rf lines, are given in Fig. 8 as a function of the interpulse angle.

These curves show the mean of 100 cross-correlation curves for windowed rf lines corresponding to backscatter from 2 to 3 mm from the transducer surface. It can be seen that the cross-correlation peak is narrower for H40 and narrowest for F40, where cross-correlation is >0.5 within an angle of 1.0°. The width of these cross-correlation peaks is attributed to the beamwidth at this distance (see Fig. 2). These curves indicate the efficacy of averaging neighboring rf lines to improve the SNR for small interpulse angles. For example, the decorrelation between H40 rf lines (from 2–3 mm) is less than 0.1 within an angle of 0.5°. Figure 9 shows the average frequency spectra of PI pairs (2–4 mm) for different interpulse angles for 2 MPa amplitude pulse pairs (0° and 180°). The fundamental suppression through PI at different distances is expressed as a function of the interpulse angle in Fig. 10. From this figure it is clear that the efficiency of PI decreases more rapidly with increasing interpulse angles. This figure also shows that the interpulse angle dependent fundamental reduction decreases quicker for scatterers that are located closer to the transducer. A fundamental suppression of approximately 16 and 10 dB is obtained when PI is applied with 512 (interpulse angle ~0.70°) and 256 (interpulse angle ~1.4°) rf lines per rotation, respectively. These line densities correspond to those used in current clinical mechanically scanned IVUS systems.

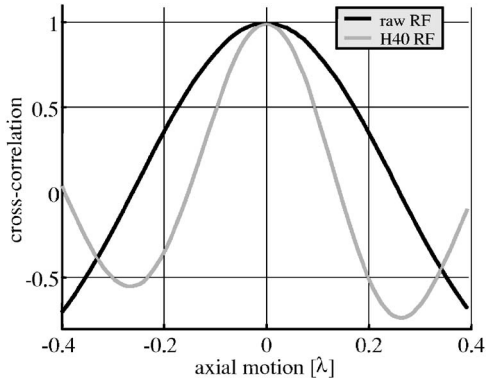


FIG. 11. Cross-correlation of raw and H40 rf lines at a distance of 2 to 3 mm from the transducer, plotted as a function of axial catheter-to-tissue motion expressed in fundamental wavelengths.

4. Effects of axial motion on pulse inversion

The effect of axial catheter-to-tissue motion on the average cross-correlation values between raw and H40 filtered rf signals from a depth of 2 to 3 mm has been plotted in Fig. 11. For a transmit amplitude of 2 MPa PI pairs at a depth of 2 to 3 mm and for interpulse angles of 0.15°, 0.30°, 0.75°, and 1.5°, the average fundamental suppression is plotted as a function of catheter-to-tissue motion expressed in fundamental wavelengths between successive pulses (λ) in Fig. 12. This graph illustrates that axial motion begins to degrade fundamental suppression for an interpulse angle of 0.15° from $\lambda > 0.005$, corresponding to >5 mm/s, based on a fundamental frequency of 20 MHz and a PRF of 12.5 kHz.

IV. CONCLUSION AND DISCUSSION

Nonlinear fields at 20–40 MHz have been simulated for circular IVUS transducers through media with frequency dependent attenuation values in the range of those of vascular tissue and blood. The influence of catheter rotation on the fundamental suppression of pulse inversion has been studied for a range of interpulse angles. The effect of axial catheter-to-tissue motion on the fundamental suppression of pulse inversion has also been investigated.

The nonlinear beam simulations were carried out in a lossy and nonlinear medium, where dispersion was assumed to be negligible. Dispersion, the frequency dependent phase

speed, would lead to smoothing of the wave profile, thereby suppressing nonlinear distortions (Naugolnykh and Ostrovsky, 1998). Limited experimental evidence shows that biologic tissue is dispersive but that it is only 0.7% per frequency decade for an attenuation of 1 dB/cm/MHz (Gurumurthy and Arthur, 1982). For many practical purposes this variation in the phase velocity can be assumed constant (Angelsen, 2000). The frequencies of interest in this study cover only a limited frequency range and the relatively high frequency dependent attenuation of blood and tissue prevents shock waves from being generated. Hence, dispersion was considered negligible. Thorough studies on the effect of dispersion under nonlinear propagation in media with frequency power law dependent attenuation have been performed (Szabo, 1995; He, 1998; Wallace *et al.*, 2001).

In the nonlinear beam simulations where absolute penetration depths of F40 and H40 fields were compared, the assumption was made that the maximum pressure at the transducer surface was the same for the 20 MHz as well as for the 40 MHz excitation pulse. The maximum excitation pressure for a single-element IVUS catheter was approximated by 5 MPa. Moran *et al.* reported a peak negative pressure of 3.27 MPa at a distance of 2 mm from a commercial 30 MHz single-element catheter, corresponding to a Mechanical Index (MI) of 0.59 (Moran *et al.*, 2002). The 20 and 40 MHz pulses with a peak negative pressure of 5 MPa, which were used in this study, would lead to maximum MI values of 1.1 and 0.79, respectively, which is below the MI limit (of 1.9) of current clinical ultrasound equipment. Another potential restriction on the maximum peak pressure is acoustic saturation, which is a physical limit. Acoustic saturation, defined as the condition where incremental transmission loss closely approximates the incremental increase in source intensity, limits the maximum acoustic pressure that may be reached in a soft tissue. According to Duck (1999), saturation pressure in tissue may tend toward the saturation limits given by weak-shock theory. These theoretical saturation limits were far from exceeded by the 5 MPa amplitude pulses at 20 and 40 MHz, as used in this simulation.

The effect of rotation on averaging, decorrelation, and pulse inversion was simulated by rotating the transducer with respect to the scatterer volume. The minimal rotation angle between successive lines of sight was chosen to be 0.15°, corresponding to the experimentally used line density of 2400 lines per rotation (Frijlink *et al.*, 2004). Using this line density, the commercially and clinically used rotational speed of 30 rotations per second will then result in a pulse-repetition-frequency (PRF) of approximately 75 kHz, which is still lower than the maximum PRF (~100 kHz) of a rotating single-element IVUS system, as limited by sound propagation speed assuming a penetration depth of 7.5 mm. So in spite of the lower line density (e.g., 256 lines per rotation) at 30 rotations per second of current commercially available IVUS systems, no physical limitations exist to increase to a line density of 2400 lines per rotation.

Experimental data on the effect of catheter rotation on PI with a prototype harmonic IVUS system has been published by Goertz *et al.* (2006). In this study we reported that the average fundamental frequency spectra from tissue mimicking material showed a pronounced peak around 20 MHz, which was reduced by approximately 20 dB with the application of PI (with a PRF of 12.5 kHz and 2400 lines per rotation). These experimental results correspond reasonably well to the simulation results, as presented in this study, indicating the validity of the simulation approach.

The effect of axial tissue-to-catheter motion on the average fundamental suppression by PI was expressed as a function of relative axial motion between successive inverted pulses. Shen and Li (2002) found that axial motion between two successive pulses of less than 0.125λ (where λ is the wavelength at the fundamental frequency) partly cancels the second harmonic intensity and increases fundamental leakage. An estimation of absolute axial motion as present in arteries is needed for the useful interpretation of Fig. 12. The maximum axial tissue (vessel wall) motion with respect to the centerline of a vessel can be obtained from published data (Tortoli *et al.*, 2001; Hoeks *et al.*, 1990), which show common carotid artery wall peak velocities in the range of 2–8 mm/s. Derived from coronary artery pressure-area curves (Williams *et al.*, 1999), maximum vessel wall velocities in the radial direction are 2 mm/s. If no catheter translation is assumed, an upper estimate of axial vessel wall motion between successive pulses at the experimentally used PRF of 12.5 kHz (Frijlink *et al.*, 2004) would be 0.008λ. For the more realistic situation where catheter translation is assumed, the maximum axial tissue motion of the coronary vessel wall with respect to an IVUS catheter was estimated at approximately 3.5 mm/s, derived from recent IVUS palpography data (Leung *et al.*, in press). Since this would correspond to a movement of 0.004λ (when the PRF is 12.5 kHz), axial catheter-to-tissue motion between consecutive pulses could be considered to be of low influence on fundamental suppression (<1 dB) and on the second harmonic intensity. Nevertheless, the significance of axial catheter-to-tissue motion will increase with increasing fundamental frequencies and for decreased PRFs.

The results from this simulation study are important to guide future optimization of system design for harmonic IVUS applications (THI and contrast harmonic imaging) with mechanically scanned single-element catheters. The PRF, line-density, and transducer size and geometry could be altered to optimize the fundamental suppression with PI. This simulation tool could also be used to investigate different pulse schemes (coded excitation) for isolating harmonic signals. Further, such simulations may also be useful in the context of guiding the implementation and optimization of nonlinear contrast imaging systems (Goertz *et al.*, 2006). For example, in second harmonic intravascular contrast imaging applications, simulations could help minimize the effect of unwanted tissue harmonics.

ACKNOWLEDGMENTS

This work is financially supported by the Dutch Technology Foundation (STW). We thank Charles Lancée for his implementation of a numerical spatial impulse response method where arbitrary frequency dependent attenuation could be included.

Angelsen, B. A. J. (**2000**). *Ultrasound Imaging, Waves, Signals and Signal Processing* (Emantec As, Trondheim), Vol. **1**.
Averkiou, M. A., and Hamilton, M. F. (**1997**). "Nonlinear distortion of short pulses radiated by plane and focused circular pistons," J. Acoust. Soc. Am. **102**, 2539–2548.
Averkiou, M. A., Roundhill, D. N., and Powers, J. E. (**1997**). "A new imaging technique based on the nonlinear properties of tissues," Proceedings of the 1997 IEEE Ultrasononics Symposium **2**, 1561–1566.
Bouakaz, A., Lancee, C. T., and de Jong, N. (**2003**). "Harmonic ultrasonic field of medical phased arrays: Simulations and measurements," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 730–735.
Cherin, E. W., Poulsen, J. K., van der Steen, A. F. W., Lum, P., and Foster, F. S. (**2002**). "Experimental characterization of fundamental and second harmonic beams for a high-frequency ultrasound transducer," Ultrasound Med. Biol. **28**, 635–646.
Duck, F. A. (**1990**). *Physical Properties of Tissue* (Academic, London).
Duck, F. A. (**1999**). "Acoustic saturation and output regulation," Ultrasound Med. Biol. **25**, 1009–1018.
Duck, F. A. (**2002**). "Nonlinear acoustics in diagnostic ultrasound," Ultrasound Med. Biol. **28**, 1–18.
Foster, F. S., Harasiewicz, E. A., and Sherar, M. D. (**2000**). "A history of medical and biological imaging with polyvinylidene fluoride (PVDF) transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1363–1371.
Frijlink, M. E., Goertz, D. E., van Damme, L. C. A., Krams, R., and van der Steen, A. F. W. (**2004**). "Intravascular ultrasound tissue harmonic imaging in vivo," Proceedings of the 2004 IEEE Ultrasononics Symposium **2**, 1118–1121.
Frijlink, M. E., Goertz, D. E., and van der Steen, A. F. W. (**2005**). "Reduction of stent artifacts using high-frequency harmonic ultrasound imaging," Ultrasound Med. Biol. **31**, 1335–1342.
Goertz, D. E., Frijlink, M. E., de Jong, N., and van der Steen, A. F. W. (**2006**). "Nonlinear intravascular ultrasound contrast imaging," Ultrasound Med. Biol. **32**, 491–502.
Gurumurthy, K. V., and Arthur, R. M. (**1982**). "A dispersive model for the propagation of ultrasound in soft tissue," Ultrason. Imaging **4**, 355–377.
Hamilton, M. F., and Blackstock, D. T. (**1998**). *Nonlinear Acoustics* (Academic, San Diego).
He, P. (**1998**). "Simulation of ultrasound pulse propagation in lossy media obeying a frequency power law," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 114–125.
Hoeks, A. P., Brands, P. J., Smeets, F. A., and Reneman, R. S. (**1990**). "Assessment of the distensibility of superficial arteries," Ultrasound Med. Biol. **16**, 121–128.
Hope Simpson, D. H., Chin, C. T., and Burns, P. N. (**1999**). "Pulse inversion doppler: A new method for detecting nonlinear echoes from microbubble contrast agents," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 372–382.
Humphrey, V. F. (**2000**). "Nonlinear propagation in ultrasonic fields: measurements, modelling and harmonic imaging," Ultrasonics **38**, 267–272.
Hunt, J. W., Worthington, A. E., and Kerr, A. T. (**1995**). "The subtleties of ultrasound images of an ensemble of cells: simulation from regular and more random distributions of scatterers," Ultrasound Med. Biol. **21**, 329–341.
Jensen, J. A., Gandhi, D., and O'Brien, W. D. (**1993**). "Ultrasound fields in an attenuating medium," Proceedings of the 1993 IEEE Ultrasononics Symposium **1**, 943–946.
Jensen, J. A., and Svendsen, N. B. (**1992**). "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 262–267.
Kerr, A. T., and Hunt, J. W. (**1992**). "A method for computer simulation of ultrasound doppler color flow images-i. theory and numerical method," Ultrasound Med. Biol. **18**, 861–872.
Lee, Y. S., and Hamilton, M. F. (**1995**). "Time-domain modeling of pulsed finite-amplitude sound beams," J. Acoust. Soc. Am. **97**, 906–917.
Leung, K. Y. E., Baldewsing, R. A., Mastik, F., Schaar, J. A., Gisolf, A., and Van der Steen, A. F. W. (**2006**). "Motion compensation for intravascular ultrasound palpography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 1269–1280.
Li, Y. D., and Zagzebski, J. A. (**2000**). "Computer model for harmonic ultrasound imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1000–1013.
Lockwood, G. R., Ryan, L. K., Hunt, J. W., and Foster, F. S. (**1991**). "Measurement of the ultrasonic properties of vascular tissues and blood from 35–65 mhz," Ultrasound Med. Biol. **17**, 653–666.
Moran, C. M., Watson, R. J., Fox, K. A., and McDicken, W. N. (**2002**). "In vitro acoustic characterisation of four intravenous ultrasonic contrast agents at 30 MHz," Ultrasound Med. Biol. **28**, 785–791.
Naugolnykh, K., and Ostrovsky, L. (**1998**). *Nonlinear Wave Processes in Acoustics* (Cambridge University Press, Cambridge).
Saijo, Y., and van der Steen, A. F. W. (**2003**). *Vascular Ultrasound* (Springer-Verlag, Tokyo).
Shen, C. C., and Li, P. C. (**2002**). "Motion artifacts of pulse inversion-based tissue harmonic imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1203–1211.
Stephanishen, P. R. (**1971**). "Transient radiation from pistons in an infinite planar baffle," J. Acoust. Soc. Am. **49**, 1629–1638.
Szabo, T. L. (**1994**). "Time-domain wave-equations for lossy media obeying a frequency power-law," J. Acoust. Soc. Am. **96**, 491–500.
Szabo, T. L. (**1995**). "Causal theories and data for acoustic attenuation obeying a frequency power-law," J. Acoust. Soc. Am. **97**, 14–24.
Tortoli, P., Bettarini, R., Guidi, F., Andreuccetti, F., and Righi, D. (**2001**). "A simplified approach for real-time detection of arterial wall velocity and distension," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1005–1012.
van der Steen, A. F. W., Poulsen, J. K., Cherin, E., and Foster, F. S. (**1999**). "Harmonic imaging at high frequencies for ivus," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **2**, 1537–1540.
Wallace, K. D., Holland, M. R., and Miller, J. G. (**2001**). "Improved description of shock wave evolution in media with frequency power law dependent attenuation," J. Acoust. Soc. Am. **109**, 2263–2265.
Williams, M. J., Stewart, R. A., Low, C. J., and Wilkins, G. T. (**1999**). "Assessment of the mechanical properties of coronary arteries using intravascular ultrasound: an in vivo study," Int. J. Card. Imaging **15**, 287–294.

Erratum: "The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception" [J. Acoust. Soc. Am. 118, 1730–1741 (2005)]

Catherine Mayo[a] and Alice Turk
Linguistics and English Language,University of Edinburgh, Adam Ferguson Building, 40 George Square, Edinburgh, EH8 9LL, United Kingdom

(Received 18 May 2006; accepted 20 June 2006)
[DOI: 10.1121/1.2225761]

PACS number(s): 43.10.Vx, 43.71.Ft [MSS]

The first paragraph of Section B(3) of the Method (p. 1735) should begin:

"Nine different VOT values were synthesized, varying in 5–ms steps from 40 ms of aspiration (as generated by the AH parameter of Sensimetrics Corporation's Klatt synthesizer, Sensyn) (most /t/-like) to 0 ms of aspiration (most /d/-like). Voicing (as generated by the AV parameter of the Sensyn synthesizer), began immediately at VOT."

The final paragraph of Section B(3) of the Method (p. 1735) should begin:

"The total duration of each syllable was 355 ms, with total voiced vowel duration ranging from 315 ms for the stimuli with the most /t/-like VOT values, to 355 ms for the stimuli with the most /d/-like VOT values."

[a]Electronic mail: catherin@ling.ed.ac.uk

Erratum: "Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions" [J. Acoust. Soc. Am. 115, 3184–3194 (2004)]

Catherine Mayo[a] and Alice Turk
Linguistics and English Language,University of Edinburgh, Adam Ferguson Building, 40 George Square, Edinburgh, EH8 9LL, United Kingdom

(Received 18 May 2006; accepted 20 June 2006)
[DOI: 10.1121/1.2225779]

PACS number(s): 43.10.Vx, 43.71.Ft [MSS]

The first paragraph of Section B(3) of the Method (p. 3187) should begin:

"Nine different VOT values were synthesized, varying in 5-ms steps from 40 ms of aspiration (as generated by the AH parameter of Sensimetrics Corporation's Klatt synthesizer, Sensyn) (most /t/-like) to 0 ms of aspiration (most /d/-like). Voicing (as generated by the AV parameter of the Sensyn synthesizer), began immediately at VOT."

The first paragraph of Section B(4) of the Method (p. 3188) should begin:

"The VOT values used for the /ti/-/di/ contrast were the same as those used for the /ta/-/da/ contrast. These varied in 5-ms steps from 40 ms of aspiration (most /t/-like) to 0 ms of aspiration (most /d/-like). Voicing began immediately at VOT."

The final paragraphs of Section B(3), and Section B(4) of the Method (both p. 3188) should begin:

"The total duration of each syllable was 355 ms, with total voiced vowel duration ranging from 315 ms for the stimuli with the most /t/-like VOT values, to 355 ms for the stimuli with the most /d/-like VOT values."

[a]Electronic mail: catherin@ling.ed.ac.uk

 0001-4966/2006/120(3)/1733/1/$22.50