CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 122, No. 2 August 2007

(Continued)

CONTENTS—*Continued from preceding page*

TRANSDUCTION [38]

STRUCTURAL ACOUSTICS AND VIBRATION [40]

NOISE: ITS EFFECTS AND CONTROL [50]

ARCHITECTURAL ACOUSTICS [55]

ACOUSTIC SIGNAL PROCESSING [60]

PHYSIOLOGICAL ACOUSTICS [64]

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

BIOACOUSTICS [80]

JASA EXPRESS LETTERS

CUMULATIVE AUTHOR INDEX

Statistical analysis of sound transmission results obtained on the New Jersey continental shelf

Simona M. Dediu and William L. Siegmann
Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, New York, 12180
dedius3@rpi.edu, siegmw@rpi.edu

William M. Carey
Department of Aerospace and Mechanical Engineering, Boston University, Boston, Massachusetts 02215
wcarey@bu.edu

Abstract: Experiments have been conducted near the site of AMCOR Borehole 6010 on the New Jersey Shelf to evaluate propagation predictability in sandy shallow-water environments. The influence of a nonlinear frequency dependence of the sediment volume attenuation in the uppermost sediment layer at this location is examined. Previously it was determined that a frequency power-law exponent of 1.5 was required for the best modeling of experimental results over the band 50–1000 Hz. The approach here references the attenuation to an accepted value at 1 kHz and makes extensive comparisons between measurements and calculations, to determine a power-law exponent of 1.85 ± 0.15.

PACS numbers: 43.30.Ma, 43.30.Zk, 43.30.Dr [JFL]
Date Received: March 14, 2007 **Date Accepted:** May 25, 2007

1. Introduction

The focus of this work is an analysis of the frequency dependence of sediment volume attenuation. Experimental evidence from many investigators, summarized in Ref. 1, shows that the accurate calculation of transmission loss in shallow-water waveguides with sandy-silty bottoms requires a nonlinear frequency dependence for the intrinsic attenuation in the upper sediment layer for frequencies between 100 Hz and 1 kHz. The previous results show that a nonlinear dependence in such environments is not a new finding. For instance, experiments have been conducted on the coastal margins and seas, with bottom sediment layers often formed by deposition of sand and silt, at locations including the West Coast of Florida, the New Jersey Shelf and near the Hudson Canyon, the Korean Strait, and the East and South China Sea. These experiments, over many mid-range frequencies and in areas with known geophysical sediment properties as functions of depth, confirm the requirement of nonlinear frequency dependence of the effective upper sediment attenuation.

This paper is motivated by transmission loss (TL) results obtained on the New Jersey Shelf near AMCOR Borehole 6010. These experiments, during October 1988 and September 1993, were designed to study acoustic influences of environmental parameters such as range-dependent bathymetry, sub-bottom structure, sound speed variability, and sea state.[2] Several 50–1000 Hz continuous-wave transmissions were performed, both parallel and perpendicular to the New Jersey Shelf break, close to the region of a major experiment in August 2006. This site has relatively well studied geoacoustic properties, and both experiments were conducted under similar downward-refracting conditions, with supporting measurements of salinity, temperature, and sound speed. Here we analyze the acoustic measurements from the longer (26 km) of the two runs parallel to the shelf break, with slowly varying depth (70–74 m in 1988, and 71–77 m in 1993).[3] Other than the small differences in bathymetry, the geometries for the two experiments were essentially the same. The source depths were 36 m in 1988 and 30 m in 1993, and the vertical receiver arrays were located in the bottom two thirds of the water

column (with data used from 42.5, 57.5, and 73 m in 1988, and from 43.1, 52.6, and 69.7 m in 1993). A detailed description of the experiments is contained in Carey *et al.*[2]

In the first extensive analysis of these experiments by Evans and Carey,[3] parameters that characterize the nonlinear dependence of attenuation are referenced to the value at a frequency of 50 Hz, with a power-law exponent increase to a frequency of 1 kHz. In contrast, we use an estimate of the attenuation at 1 kHz based on the work of Hamilton[4] and determine the power-law decrease to lower frequencies. This eliminates some environmental uncertainty, since the value of the actual attenuation at 50 Hz is small. The sound speed profiles were reexamined to obtained representative profiles.

In Sec. II we present our hypotheses. The comparison of measured and calculated TL is described in Sec. III, and conclusions are summarized in Sec. IV.

2. Hypotheses

We examine the consequences of frequency dependence of the sediment volume attenuation in the uppermost sediment layer. The analysis uses a power-law form of the frequency dependence, with exponent n, based on many previous investigations that are described by Holmes *et al.*[1] The nonlinear frequency dependence of attenuation is assumed to occur only in the upper (typically 5 m) sediment. We will describe the excellent agreement between measured and calculated sound transmissions that can be obtained when the value of n is between 1.7 and 2. The metric used for the influence of the nonlinear frequency dependence of attenuation on the range decrease of TL is an effective attenuation coefficient (EAC). This quantity is derived from range- and depth-averaged TL for both measurements and calculations. The TL calculations use measured geophysical properties and range-dependent bathymetry, along with water sound speed profiles derived from measurements. The frequency power-law exponent is allowed to vary until a comparison between measured and calculated EACs is achieved within acceptable bounds.

One significant change from the procedure used by Evans and Carey[3] is that the intrinsic sediment attenuation model is modified for a reference frequency at 1 kHz instead of 50 Hz. The surface value of the attenuation profile is taken to be consistent with Hamilton's results at 1 kHz[4] and other recent results.[5] Our principal hypothesis is that for sandy-silty bottoms in the frequency range 100 Hz to 1 kHz, the effective sediment attenuation follows a nonlinear frequency dependence with power-law exponent within the interval 1.5 and 2.0, and the near-surface attenuation at 1 kHz is within the range 0.3–0.4 dB/m.

Figures 1 and 2 include all measured water sound speed profiles for the 1988 and 1993 experiments. The representative profiles used in previous calculations[3] were found by considering the mean of the collected profiles and examining variations from that mean, based on known information about other profiles in the area. Our approach seeks an effective representative sound speed profile for each track that is considered successful if averaged TL results from parabolic equation calculations give good agreement between measured and computed EACs over a band of frequencies and parameter cases. We investigated variations of the 1988 profiles given in Fig. 1 and found that the profile employed by Evans and Carey[3] (thick curve) gives the best agreement between measured and calculated EACs. The profile used for the 1993 experiment calculations is based on all the measured profiles and is indicated by the thick curve in Fig. 2, which has different thermocline characteristics than that used in Ref. 3. We analyze frequencies greater than 400 Hz because at lower frequencies, attenuation values are small and the field has relatively few modes.

An implicit hypothesis in our analysis is that for an incremental sediment volume (sides small compared to a wavelength), the sediment can be regarded as homogeneous and isotropic. Consequently, the Lame constants λ and μ along with a density ρ_m describe the homogenized sediment-water mixture. This implies that the sediment moduli are the constants $B_m=\lambda+(2/3)\mu$ and $G_m=\mu$. It follows that the compressional and shear wave speeds can be calculated using the formulas

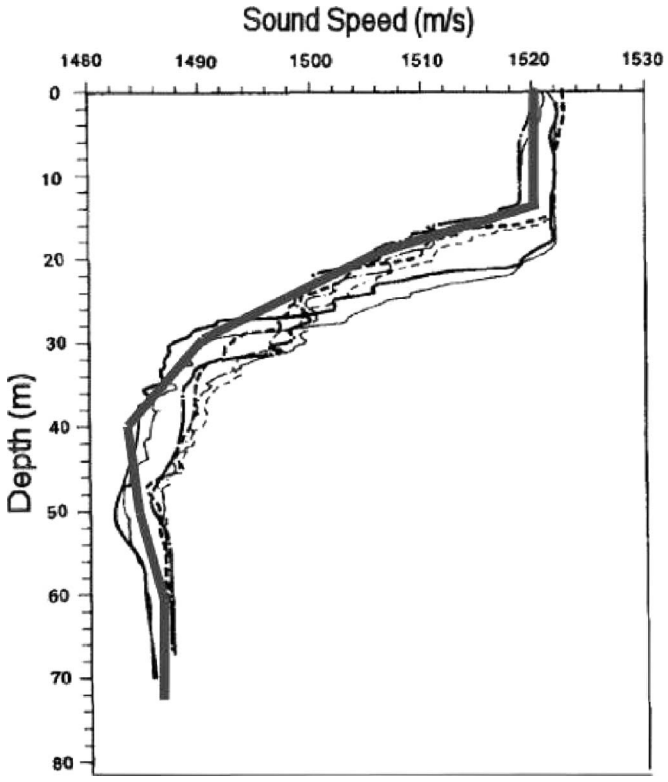


Fig. 1. 1988 experiment: measured water sound speed profiles and effective profile (thick curve) used in calculations.

$$C_{mc} = \sqrt{\frac{B_m}{\rho_m}}, \quad C_{ms} = \sqrt{\frac{G_m}{\rho_m}}. \tag{1}$$

Therefore, the variations with depth z for the compressional and shear wave speeds should be proportional

$$\frac{C_{mc}(z)}{C_{ms}(z)} = \sqrt{\frac{B_m}{G_m}} = (\text{constant}) \tag{2}$$

In addition we have

$$B_m = B_{om}(1 - i\beta), \; G_m = G_{om}(1 - i\zeta), \tag{3}$$

where $-\beta$ and $-\zeta$ are loss factors and B_{om} and G_{om} are mean values of B_m and G_m. For the plane wave approximation, we write the wave number k in $\exp(ikr)$ as

$$k = \frac{\omega}{C_{mc}} = \omega \Big/ \sqrt{\frac{B_{om}(1 - i\beta)}{\rho_m}} = \omega\left(1 + i\frac{\beta}{2}\right) \Big/ \sqrt{\frac{B_{om}}{\rho_m}} = k_o + i\frac{\omega\beta}{2C_{omc}(z)}. \tag{4}$$

This implies that $\exp(ikr) = \exp(ik_o r - \alpha r)$, where the intrinsic attenuation $\alpha(z)$ is

$$\alpha(z) = \frac{\omega\beta}{2} C_{omc}^{-1}(z). \tag{5}$$

Since B_{om} and $\rho_m(z)$ are determined from geophysical measurements and empirical relationships, we can find the mean compressional sound speed profile $C_{omc}(z)$ and then the attenuation

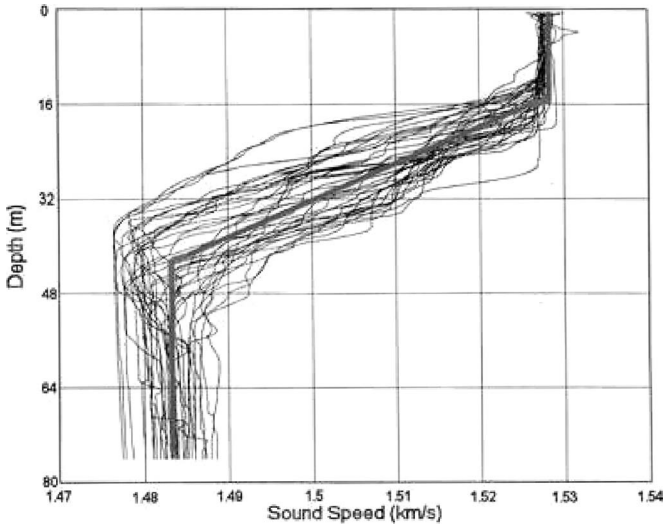


Fig. 2. 1993 experiment: measured water sound speed profiles and effective profile (thick curve) used in calculations.

profile from Eq. (5). Note that the factor β in Eq. (5) is specified by one value of the attenuation, conveniently taken at the water-sediment interface.

3. Data analysis

Transmission loss calculations are performed using the parabolic approximation method, which produces accurate one-way propagation solutions for range-dependent environments in shallow-water channels. The source and receiver depths are interchanged by the principle of reciprocity to simplify the computations. In addition, the geoacoustics is simplified by neglecting shear properties, since previous work[3] concluded that shear effects are evidently not important at the experimental frequencies in this waveguide. From Sec. II the nonlinear frequency dependence of the attenuation is taken as

$$\alpha(z,f) = \alpha(z,f_0)\left(\frac{f}{f_0}\right)^n, \tag{6}$$

where f_0 is the reference frequency 1 kHz, n is the frequency power-law exponent, and $\alpha(z,f_0)$ is the intrinsic attenuation profile (in dB/m) at 1 kHz. The objective is to determine the parameters n and $\alpha(z,f_0)$, using the known range of values for $\alpha(z,f_0)$ from Hamilton.[4] From Eq. (5) it is only n and $\alpha(H,f_0)$ that need to be found, where H is the water depth. These two are the only free parameters in the analysis, because other geophysical properties, geoacoustic profiles, sound speed profiles, and bathymetry are all specified independently prior to the TL calculations.

Comparisons between measured and calculated TL follow generally the procedure in Ref. 3. The measured and calculated EACs are the comparison metrics and are defined as follows. For each receiver the measured and calculated TL data are first range averaged to minimize effects of noise and modal interference over the interval 3–21 km using a 1 km window. Calculated sample points are 50 m apart, and measured sampled points are separated by between 35 and 230 m. The window-averaged measured and calculated TL are each fit using the expression

$$TL \approx \alpha_{\text{eff}}\, r + b + 10\log(r), \tag{7}$$

where r is range in m, α_{eff} is the EAC in dB/m, the second term b on the right is a mean level, and the third term is cylindrical spreading. Thus, EACs are slopes of the least-squares fit of the

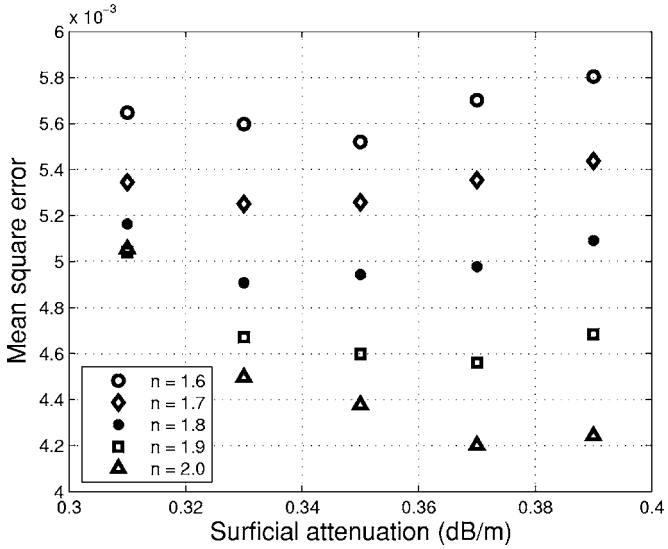


Fig. 3. Mean square errors obtained for different values of $\alpha(73,1\ \text{kHz})$ and exponents from 1.5 to 2.0.

range-averaged reduced TL. We obtain measured-computed EAC pairs for each frequency and parameter value, and average these pairs over the three receivers to minimize the depth variability. Appropriate ranges for the values of surficial attenuation $\alpha(H,1\ \text{kHz})$ (with $H=73$) and n are based on agreement between measurements and calculations. The agreement is based on three steps. First, the minimum mean square error (MSE) about a line with slope one is found. Then, unlike Ref. 3, unweighted least-squares fits are used to compare measured and computed EAC pairs. Finally, the goodness of fit of the measured to calculated data is assessed.

First, the minimum MSEs are obtained by minimizing over $\alpha(73,1\ \text{kHz})$ and n

$$\frac{1}{N}\sum_{i=1}^{N}(\alpha_{\text{eff}}(f)_{\text{calc},i}-\alpha_{\text{eff}}(f)_{\text{meas},i})^2, \tag{8}$$

where $N=8$ is the number of frequencies available and α_{eff} for measurements and calculations are given in Eq. (7) and depend on frequency. The MSEs for different values of $\alpha(73,1\ \text{kHz})$ and n are plotted in Fig. 3. As the exponent increases, the MSEs decrease. Also, each set of MSE symbols has a minimum for certain $\alpha(73,1\ \text{kHz})$ values: for exponents $n=1.5$ and 1.6, the minimum is at 0.35 dB/m; for $n=1.7$ and 1.8, at 0.33 dB/m; and for $n=1.9$ and 2.0, at 0.37 dB/m. Consequently, we estimate the range of $\alpha(73,1\ \text{kHz})$ from 0.33 to 0.37 dB/m and the range of n from 1.7 to 2.0. The minimum MSE for these parameter values varies between 0.0053 and 0.0042, for a maximum of about 25%.

Next, we obtain the linear least-squares fit of the measured and computed EAC pairs. If the agreement is perfect, then the slope of the line would be one. Figure 4 displays an EAC scatter plot for the experimental frequencies used, for values $\alpha(73,1\ \text{kHz})-0.35$ dB/m and $n=1.8$. The dashed line has slope one and intercept zero, and the least-squares line is solid, with slope 0.845 and intercept 0.101. If the procedure implicit in this figure is repeated for surficial attenuation values in the range specified by Hamilton, 0.3–0.4 dB/m, and for exponents between 1.5 and 2.0, a broad maximum occurs in the neighborhood of 0.35 dB/m. Moreover, variation of the least-squares slopes for different exponent values is quite small.

Finally, we estimate the goodness of fit of measured EACs to the calculated data, to determine how well the EAC pairs are fit by a straight line with slope one. The hypothesis tested is that the slope is one, and it is rejected if the slope is more than two standard deviations from one. The regression analysis is performed using a 95% level of confidence. For the example

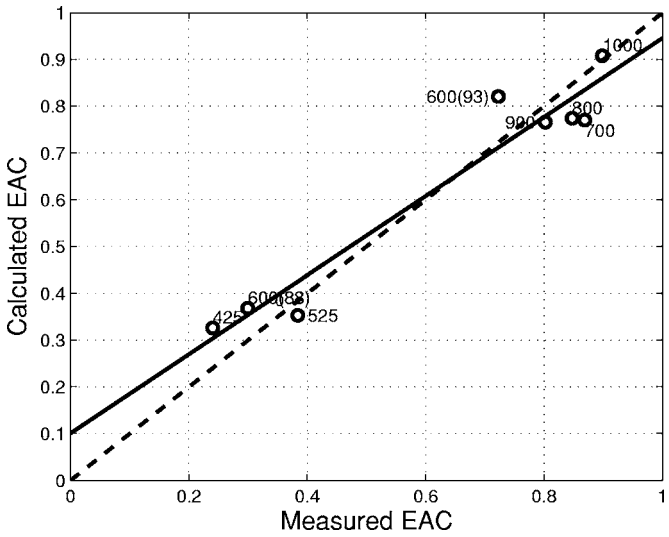


Fig. 4. Scatter plot for EACs over eight experimental frequencies for surficial attenuation 0.35 dB/m and exponent 1.8. The dashed line is ideal with slope one, and the least-squares line is solid.

shown in Fig. 4, the slope of the least-squares fit is 0.85 with a standard deviation of 0.09, so the slope is within two standard deviations from one. The assessment of the goodness of fit of the slope regressions leads to a range for surface attenuation values from 0.33 to 0.35 dB/m and for exponents n from 1.7 to 2.0.

4. Conclusion

Previous studies by many investigators,[1] including those of two experiments at a New Jersey Shelf location, show that accurate calculations of shallow-water sound transmission in waveguides with sandy-silty bottoms require a nonlinear frequency dependent attenuation in the near-surface sediment layer between 100 Hz and 1 kHz. The principal goal of this paper is to focus on measurements from two New Jersey Shelf experiments and to estimate the two principal parameters of the intrinsic attenuation profile, using an attenuation value at 1 kHz as a reference. The analysis shows that, in order to account for the frequency behavior of measured transmission loss, significant nonlinear frequency dependence is necessary in the attenuation of the uppermost sediment. We found that the site-specific surficial attenuation range of 0.33–0.35 dB/m and the exponent range of 1.7–2.0 achieve the near optimal agreement between measurements and calculations. These values are consistent with many previous results for sandy-silty sediments.[1] Note: since the acceptance of this paper, we have become aware of recently published results[6] that are evidently at variance with those contained in this paper, as well as with previous work of others. Based on transmission loss results in the same area as Ref. 6, we report a value of the attenuation coefficient that is consistent with Refs. 1, 4, 5, and 7.

References and links

[1]J. D. Holmes, W. M. Carey, S. M. Dediu, and W. L. Siegmann, "Nonlinear frequency dependent attenuation in sandy sediments," J. Acoust. Soc. Am. **121**, EL218–EL222 (2007).
[2]W. M. Carey, J. Doutt, R. B. Evans, and L. M. Dillman, "Shallow-water sound transmission measurements on the New Jersey Continental Shelf," IEEE J. Ocean. Eng. **20**, 321–336 (1995).
[3]R. B. Evans and W. M. Carey, "Frequency dependence of sediment attenuation in two low-frequency shallow-water acoustic experimental data sets," IEEE J. Ocean. Eng. **23**, 439–447 (1998).
[4]E. L. Hamilton, "Geoacoustic modeling of the sea floor," J. Acoust. Soc. Am. **68**, 1313–1340 (1980).
[5]J.-X. Zhou and X.-Z. Zhang, "Nonlinear frequency dependence of the effective seabottom acoustic attenuation from low-frequency field measurements in shallow water (A)," J. Acoust. Soc. Am. **117**, 2494 (2005).
[6]Y.-M. Jiang, N. R. Chapman, and M. Badiey, "Quantifying the uncertainty of geoacoustic parameter estimates for the New Jersey shelf by inverting air gun data," J. Acoust. Soc. Am. **121**, 1879–1894 (2007).
[7]R. Stoll, *Sediment Acoustics* (Springer-Verlag, New York, 1989).

Effect of filter spacing on melody recognition: Acoustic and electric hearing

Kalyan Kasturi and Philipos C. Loizou
Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75083-0688
loizou@utdallas.edu

Abstract: This paper assesses the effect of filter spacing on melody recognition by normal-hearing (NH) and cochlear implant (CI) subjects. A new semitone filter spacing is proposed for music. The quality of melodies processed by the various filter spacings is also evaluated. Results from NH listeners showed nearly perfect melody recognition with only four channels of stimulation, and results from CI users indicated significantly higher scores with a 12-channel semitone spacing compared to the spacing used in their daily processor. The quality of melodies processed by the semitone filter spacing was preferred over melodies processed by the conventional logarithmic filter spacing.

PACS numbers: 43.66.Ts, 43.66.Hg [QJF]
Date Received: March 12, 2007 **Date Accepted:** May 2, 2007

1. Introduction

Central to any speech coding strategy used in multi-channel cochlear implants is the decomposition of the acoustic signal into frequency bands. Given the large number (12–22) of electrodes available in commercial implant devices, it is becoming more important to find the best mapping (or, equivalently, spacing) of frequency bands to electrodes. The majority of implant processors use logarithmic (or semilog) filter spacing and that has worked well so far, at least for speech recognition.

A number of studies evaluated alternative filter spacings for vowel recognition and F0 discrimination (e.g., Fourakis *et al.*, 2004; Geurts and Wouters, 2004; Laneau *et al.*, 2004). Fourakis *et al.* (2004) advocated the placement of more filters in the F1/F2 region for better representation of the first two formants. Small but significant improvements were noted on vowel recognition with an experimental map which included one additional electrode in the F2 region. Similar outcome was reported in Skinner *et al.* (1995) and Loizou (2006). Other studies also considered the possibility of allocating more filters in the low frequencies for better place coding of individual harmonics and consequently better pitch perception. A new filter bank was proposed by Geurts and Wouters (2004) based on a simple loudness model used in acoustic hearing. The new filter bank, which allocated more filters in the low frequencies, was tested on an F0 detection task in the absence of temporal cues and yielded lower detection thresholds to F0 for synthetic vowel stimuli compared to a conventional filter bank based on log spacing.

The above studies demonstrated that the filter spacing can have a positive effect on vowel recognition and can in some cases reduce F0 difference limens, at least for steady-state vowels with a steady F0 contour. Little is known, however, about the effect of filter spacing on music signals which have a dynamic F0 contour. This paper investigates the hypothesis that a filter-bank spaced according to a musical scale would provide better place coding of individual harmonics and consequently improve melody recognition. The present experiments investigate the effect of semitone frequency spacing on melody recognition by normal-hearing and cochlear implant users.

Table 1. The 3 dB frequency boundaries of the semitone filterbank. Lower (L), upper (UP) and center (C) frequencies are given for each band in Hz.

	2 channels			4 channels			6 channels			12 channels		
	L	U	C	L	U	C	L	U	C	L	U	C
1	300	424	362	300	357	328	300	337	318	300	318	309
2	424	600	512	357	424	391	337	378	357	318	337	327
3				424	505	464	378	424	401	337	357	347
4				505	600	552	424	476	450	357	378	367
5							476	535	505	378	400	389
6							535	600	567	400	424	412
7										424	449	437
8										449	476	463
9										476	505	490
10										505	535	520
11										535	566	550
12										566	600	583

2. Experiment design

2.1 Subjects and material

Six Clarion CII (Advanced Bionics Corporation) cochlear implant users participated in this experiment. All subjects were postlingually deafened adults wearing the cochlear implant (CI) for a minimum of 2–3 years (no consideration was given to their musical training experience). For comparative purposes, we also tested ten normal-hearing (NH) subjects listening to stimuli processed via acoustic simulations of cochlear implants. A set of 34 simple melodies (e.g., "Twinkle Twinkle," "O1d McDonald") with all rhythm information removed was used (Hartmann and Johnson, 1991) as test material. These same melodies were used in the study by Smith *et al.* (2002). Melodies consisted of 16 equal-duration notes synthesized using samples of a grand piano. The mean of all 16 note frequencies of each tune was concert A (440 Hz) plus or minus a semitone. The largest difference between the highest and lowest notes was 12 semitones.

2.2 Signal processing

For the NH listeners, the test material was first bandpass filtered (sixth order Butterworth) into 2–12 channels according to a semitone filter spacing that spanned an octave (300–600 Hz). This frequency range was chosen as it encompasses the mean note frequency (440 Hz) of the test stimuli. For the 12-channel condition, each filter had a bandwidth of 1 semitone (see Table 1). For the 6-channel condition, each filter had a bandwidth of 2 semitones, and for the 4- and 2-channel conditions the filters had a bandwidth of 3 and 6 semitones, respectively. In addition to the semitone spacing, a 16-channel logarithmic spacing (225 Hz–4.5 kHz) was used as control. Following the bandpass filtering, the channel envelopes are computed using a half-wave rectifier followed by a second order Butterworth low-pass filter with a cutoff frequency of 120 Hz. The resulting envelopes of each channel were modulated with white noise and refiltered with the same analysis filters. The melodies were finally synthesized by summing up the outputs of all the channels.

For the CI users, the test material was processed through the continuous interleaved sampling strategy used in the subject's daily processor, and implemented with different frequency spacings. Two different filter spacings were considered. The first one was based on the semitone scale mentioned above. Four semitone (4SM), six semitone (6SM) and 12 semitone (12SM) based filter banks were considered. In the 4SM condition, only the four most apical

electrodes were used for stimulation. Similarly, in the 6SM and 12SM conditions, only the 6 and 12 most apical electrodes were used for stimulation. The remaining electrodes were not stimulated.

The second filter spacing considered involved a combination of semitone and log spacings. This was done to account for a more realistic scenario in which the melodies might contain sung lyrics. In the 4SM condition, we considered utilizing a log spacing for the remaining 12 channels in the high frequencies. Similarly, a log spacing was used for the remaining ten channels in the 6SM condition and the remaining four channels in the 12SM condition. We refer to these hybrid frequency spacings as 4SM+LOG, 6SM+LOG and 12SM+LOG, respectively. All hybrid frequency spacings used 16 channels of stimulation. For comparative purposes, we tested subjects with the 16-channel logarithmic spacing (16LOG) used in their daily processor.

2.3 Procedure

The experiments with NH listeners were performed using a PC equipped with a Creative Labs SoundBlaster 16 soundcard. Stimuli were played to the listeners monaurally through Sennheiser HD 250 Linear II circumaural headphones. The names of the melodies were displayed on a computer monitor, and a graphical user interface enabled the subjects to indicate their response. Prior to the test, each subject was asked to select ten familiar melodies from the list of 34 melodies (with a few exceptions, most subjects selected the same melodies). A training session (lasting about 10–15 min) with the ten selected melodies was performed using the original unprocessed melodies. Subjects were required to score above 90% with unprocessed melodies before participating in the experiment. After the training session, the subjects were tested with the melodies processed through the various number of channels. The order of test conditions was randomized across subjects.

The cochlear implant subjects were tested using the Clarion research interface-II (Advanced Bionics Corporation). Prior to the test, the subjects were asked to select ten known melodies from a list of 34 melodies. The subjects were given a practice session that lasted for about 10–15 min. Following the practice session, the subjects were tested on the ten selected melodies using the logarithmic spacing, semitone spacing, and hybrid spacings. The names of the melodies were displayed on a computer monitor, and a graphical user interface enabled the subjects to indicate their response. The subjects were tested for a total of seven different filter spacings. Each spacing was tested in two blocks of three repetitions each. The order of the various spacings tested was randomized across subjects.

Following the melody recognition test, the CI users participated in an AB paired preference test. In one condition, the task was to evaluate and compare the quality of melodies processed by the 16LOG and 6SM spacings. In another condition, the task was to compare the 16LOG and 6SM+LOG spacings. In each trial, the subjects listened to two stimuli each processed using a different filter spacing. The preference test included ten melody pairs composed of five different melodies. Five of the ten melody pairs were presented as filter spacing A followed by spacing B, while the other five were presented as spacing B followed by spacing A. The subjects were instructed to make a preference as to which stimulus sounded more "musical" (i.e., sounding like a melody with "natural" melodic contour) and more pleasant. In addition, they were asked to make a confidence rating on each comparison at six distinct scales: slightly better (or slightly worse), better (or worse), and much better (or much worse). A numeric score was assigned to each rating ranging in values from +3 (much better) to −3 (much worse). A total of six (signed) confidence ratings were assigned and a distance measure was computed. The percentage preference was computed as the percentage of the number of times stimulus B was preferred over stimulus A. The distance measure was computed to assess quantitatively how much stimulus B sounded better than stimulus A. Since the distance measure is computed over ten test pairs, it ranged in values from −30 to 30, with a positive value indicating that the strategy B is preferred, and a negative value indicating otherwise.

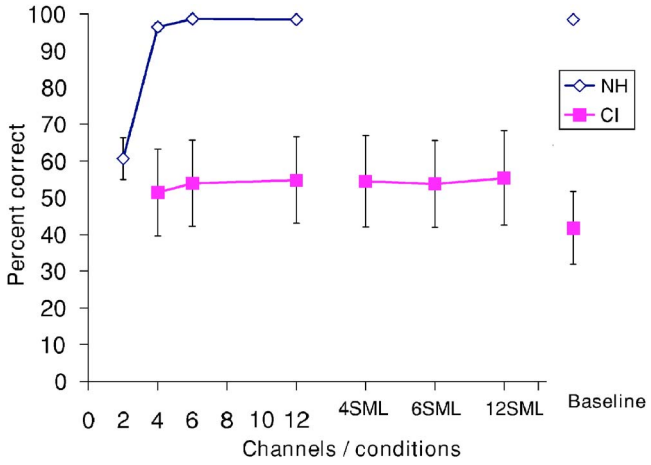


Fig. 1. (Color online) Mean percent correct scores for melody recognition by CI users (filled symbols) and normal hearing listeners (open symbols) as a function of number of semitone-spaced channels. The scores obtained by CI users in the hybrid spacing conditions (4SML, 6SML and 12SML) are also included. The baseline condition corresponds to the spacing (16 channels) used in the subjects' daily processor. Error bars indicate standard errors of the mean.

3. Results and discussion

The mean percent correct scores for melody recognition by normal hearing listeners are depicted in Fig. 1 (open symbols). Analysis of variance (ANOVA) with repeated measures showed a significant effect ($F[4,16]=59.4$, $p<0.0005$) of number of channels on melody recognition. Nearly perfect melody recognition was achieved with only four channels of stimulation spaced according to a semitone scale. The three-semitone frequency resolution does not allow individual harmonics (spaced a semitone apart) to be resolved, yet it was found sufficient for accurate melody recognition, at least by normal-hearing listeners who receive acoustic envelope information at the correct place in the cochlea and have good frequency selectivity.

The mean percent correct scores for melody recognition by CI users are depicted in Fig. 1 (filled symbols). ANOVA (with repeated measures) indicated a significant effect ($F[6,30]=2.8$, $p=0.026$) of frequency spacing on melody recognition. Post-hoc tests indicated that the scores obtained with the 12SM spacing were significantly higher ($p=0.021$) than the scores obtained with the conventional filter spacing (16LOG) used by CI users in their daily processor. Scores obtained with the other semitone spacings were not significantly higher, but approached ($p\approx0.06$) the significance level.

The preference judgments for each subject are given in Table 2. Results indicated that the quality of melodies processed by the semitone filter spacing was preferred over melodies

Table 2. Subject preference scores (ranging from 0 to 100) indicating the number of times the semitone spacing (6SM) (or hybrid spacing, 6SM+LOG) was preferred over the conventional logarithmic filter spacing (16LOG). The distance scores in parentheses (ranging from −30 to 30) are positive if the semitone (or hybrid) spacings were preferred and negative if the log spacing was preferred. Large positive distance scores indicate stronger preference of the semitone-based filter spacing over the log spacing.

	Subjects						
Filter spacing comparisons	S1	S2	S3	S4	S5	S6	Mean
16LOG vs. 6SM	100 (25)	100 (20)	100 (20)	90 (14)	90 (14)	100 (20)	96.7 (18)
16LOG vs. 6SM+LOG	100 (12)	80 (8)	0 (−19)	20 (−10)	100 (25)	50 (0)	58.3 (3)

processed by the conventional filter spacing (16LOG). The quality of melodies processed by the 6SM strategy was preferred 97% of the time over the CI user's daily strategy (16LOG). The preference of the hybrid spacing (6SM+LOG) was not as strong (58%). This could be attributed to the fact that subjects were perhaps perceiving conflicting or noncoherent pitch cues in the low- (semitone spaced) and high-frequency (log spaced) channels. We cannot exclude the possibility that the hybrid spacing might yield higher preference scores when tested with music containing sung lyrics.

The above analyses indicate that the filter spacing can have a significant effect on music perception both in terms of melody recognition and subjective quality. These results suggest that the semitone filter spacing enhanced access to place (spectral) cues resulting in better F0 discrimination. The magnitude of the improvement, however, was not as large as that observed by normal-hearing listeners receiving the same number of channels of frequency information. Two factors could have contributed to that. First, in cochlear implants the acoustic information is rarely presented in the correct place in the cochlea due to the shallow insertion depth. As a result, the frequency-to-place mapping is somewhat compressed or expanded. There is evidence (Oxenham *et al.*, 2004) to suggest that a correct (i.e., matched) frequency-to-place mapping is necessary for complex pitch perception and consequently melody recognition. We cannot exclude, however, the possibility that if the subjects were given more time to adapt to the new frequency spacing, their scores might improve even further and this warrants further investigation. Second, the place-coding resolution in cochlear implants is limited and constrained by several factors including the electrode spacing, location of electrodes in terms of their proximity to excitable neuron elements and electrode configuration (monopolar vs. bipolar). All these factors limit the frequency specificity needed for complex pitch perception. If we assume that the mismatch in frequency-to-place mapping can be compensated over time with learning, then based on the outcome by NH listeners (Fig. 1), a place-coding resolution of 3 semitones (or better) would be required for accurate melody recognition by CI users.

The data from the present experiment demonstrate that the channel density in the low-frequency range plays a critical role in melody recognition. In cochlear implants, this channel density is influenced by the signal bandwidth and number of electrodes available. In a follow up experiment we investigated the effect of signal bandwidth on melody recognition using acoustic simulations and NH listeners. Test material was bandpass filtered into $N(\mathrm{N}=2,4,6,12,40)$ frequency bands using sixth-order Butterworth filters. The N bands were uniformly spaced on a logarithmic scale and spanned either a 5 kHz (225 Hz–4.5 kHz range) or an 11 kHz (225 Hz–10.5 kHz range) signal bandwidth. Following the envelope detection (120 Hz) and white noise modulation, the signals were re-filtered through the same analysis filters and summed up for reconstruction. A new group of ten listeners participated in this experiment using the same procedure and test material. The mean results are shown in Fig. 2. Two-way ANOVA (with repeated measures) indicated a significant effect ($F[1,4]=10.5$, $p=0.031$) of signal bandwidth, a significant effect ($F[4,16]=81.6$, $p<0.005$) of spectral resolution (number of channels) and a significant interaction ($F[4,16]=5.8$, $p=0.004$). For the small-bandwidth condition, post-hoc tests (Fisher's LSD) showed that the performance asymptoted with 6 channels, while for the large-bandwidth condition performance asymptoted with 12 channels. Near perfect melody identification was achieved with 12 (or more) channels in both conditions.

The results shown in Fig. 2 clearly demonstrate that the signal bandwidth, which in turn affects the filter spacing (for a fixed number of channels), is extremely important for melody recognition. Higher performance was achieved with the small signal bandwidth, as more filters were allocated in the low-frequency range. In the 6-channel condition (based on large-bandwidth allocation), only one filter was allocated in the 300–600 Hz range, while in the corresponding 6-channel condition, based on small-bandwidth allocation, two filters were allocated within the same range. This small difference in the number of filters in the low-frequency range produced a difference of 34 percentage points in melody recognition (Fig. 2).

It is important to note that the proposed filter spacings were only tested with melodies and not with speech. Further tests are needed to assess the effects of the proposed filter-bank

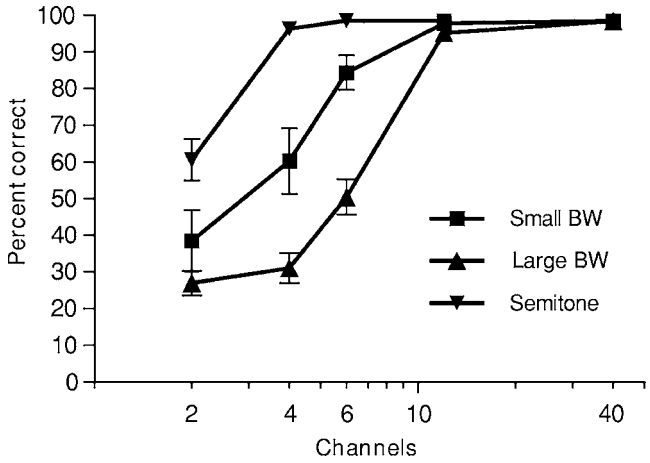


Fig. 2. Mean scores (percent correct) on melody recognition as a function of number of channels and signal bandwidth (small BW=5 kHz, large BW=11 kHz). Scores from the semitone spacing (300–600 Hz) are also plotted for comparison. Error bars indicate standard errors of the mean.

manipulations on speech recognition. The hybrid filter spacings (4SM+LOG,6SM +LOG,12SM+LOG) would clearly be more appropriate for speech recognition as they span the speech bandwidth. These spacings produced comparable performance on melody recognition as the semitone spacings (see Fig. 1). Alternatively, the semitone filter spacing could be programmed as a separate "music map" which CI users can switch to when wanting to listen to (instrumental) music.

Acknowledgment

This research was supported by Grant No. R01 DC007527 from the National Institute of Deafness and Other Communication Disorders, NIH.

References and links

Fourakis, M., Hawks, J., Holden, L., Skinner, M., and Holden, T. (**2004**). "Effect of frequency boundary assignment on vowel recognition with the Nucleus 24 ACE speech coding strategy," J. Am. Acad. Audiol **15**, 281–289.

Geurts, L., and Wouters, J. (**2004**). "Better place coding of the fundamental frequency in cochlear implants," J. Acoust. Soc. Am. **115**(2), 844–852.

Hartmann, W. M., and Johnson, D. (**1991**). "Stream segregation and peripheral channeling," Music Percept. **9**(2), 155–184.

Laneau, L., Moonen, M., and Wouters, J., (**2004**). "Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees," J. Acoust. Soc. Am. **106**(6), 3606–3619.

Loizou, P. (**2006**). "Speech processing in vocoder-centric cochlear implants," *Cochlear and Brainstem Implants*, edited by A. Moller, (Karger, Basel) Vol. **64**, pp. 109–143.

Oxenham, A. J., Bernstein, J. G. W., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**(5), 1421–1425.

Skinner, M., Holden, L., and Holden, T. (**1995**). "Effect of frequency boundary assignment on speech recognition with the SPEAK speech-coding strategy," Ann. Otol. Rhinol. Laryngol. **104**(Suppl. 166), 307–311.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.

Relationship between fundamental and formant frequencies in voice preference

Peter F. Assmann
School of Behavioral and Brain Sciences, The University of Texas at Dallas, Box 830688, Richardson, Texas 75083
assmann@utdallas.edu

Terrance M. Nearey
Department of Linguistics, University of Alberta, Edmonton, AB T6G 2E2, Canada
t.nearey@ualberta.ca

Abstract: Covariation in the size of laryngeal and vocal tract structures leads to a moderate correlation between fundamental frequency (*F*0) and formant frequencies (FFs) in natural speech. A method of adjustment procedure was used to test whether listeners prefer combinations of *F*0 and FFs that reflect this covariation. Vowel sequences spoken by two men and two women were processed by the STRAIGHT vocoder to construct three sets of frequency-shifted continua. The distributions of "best choice" responses in all three experiments confirm that listeners prefer coordinated patterns of *F*0 and FF similar to those of natural speech.

PACS numbers: 43.71.Bp, 43.71.An, 43.70.Mn, 43.70.Fq [JH]
Date Received: January 12, 2007 **Date Accepted:** March 4, 2007

1. Introduction

In speech production, source and filter properties are largely independent (Fant, 1970). Thus, vowels with the same formant frequencies (FFs) can be produced on different fundamentals (*F*0s). However, across talkers there is a moderate correlation of mean *F*0 and mean FF that stems from covariation in the size of laryngeal and supra-laryngeal structures (Fant, 1970; Titze, 1989; Fitch and Giedd, 1999). Intelligibility is preserved across a fairly wide range when *F*0 and FFs are scaled up or down in proportionate amounts (e.g., Chiba and Kajiyama, 1941; Daniloff *et al.*, 1968). However, *F*0 and FFs do not scale proportionately in natural speech. For example, the *F*0 difference between adult male and adult female voices is about 80%–90%, while FFs increase only 12%–15% (Peterson and Barney, 1952). Scaling of *F*0 and FFs that deviates from this observed pattern can produce "mismatched" combinations that evince both lower naturalness ratings (Assmann *et al.*, 2006) and reduced intelligibility (Assmann *et al.*, 2002). In this study we ask whether listeners apply an internalized knowledge of the relationship between mean *F*0 and FFs in their assessment of the naturalness of voices.

In Assmann *et al.* (2006) listeners judged the naturalness of frequency-shifted sentences by adjusting the position of a graphical slider. Two sentences, each spoken by two men and two women, were processed using the STRAIGHT vocoder (Kawahara, 1997; Kawahara *et al.*, 1999). STRAIGHT is a high-quality vocoder that uses instantaneous frequency-based *F*0 extraction and *F*0-adaptive smoothing to reconstruct the spectrum envelope. STRAIGHT provides separate scale factors for manipulating *F*0 and spectrum envelope. The former changes the average voice pitch, while the latter shifts the frequencies of all the formants (and other features of the spectrum envelope) by a constant fraction to simulate changes in vocal tract size.[1] Scale factors were chosen to provide specific combinations of mean *F*0 and mean FF for each sentence, spanning a 10×10 grid with equal logarithmic steps in mean *F*0 versus mean FF.

The results of Assmann *et al.* (2006) are summarized in Fig. 1, along with acoustic measurements from a sample of approximately 3000 vowels spoken in /hVd/ words by ten men, ten women, and 30 children from the Dallas area (Assmann and Katz, 2000; Katz and Assmann,

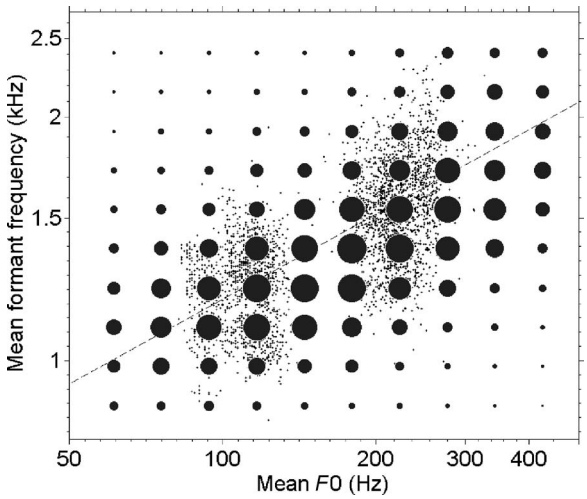


Fig. 1. Naturalness ratings (indicated by the size of the circles) for frequency-shifted sentences as a function of mean *F*0 and mean FF. For comparison, the dots show measurements of *individual* vowels from the reference database along the same dimensions. Mean *F*0 was computed as the geometric mean across all voiced frames; mean FF was computed as the geometric mean of the formant frequencies (*F*1, *F*2, and *F*3) sampled near the onset of the vowel.

2001; henceforth referred to as the *reference database*). Frequency-shifted sentences were judged more natural when synthesized with *F*0 and FF combinations that are similar to those of acoustic measurements of natural vowels. Mismatched combinations (e.g., low *F*0s combined with high FFs) were rated as less natural than matched combinations.

The present study explored to what extent listeners show a preference for natural combinations of mean *F*0 and mean FF when they are given independent control over these dimensions using a method of adjustment task. One dimension (either mean *F*0 or mean FF) was held constant while the other was adjusted by the listener over values spanning the range observed in natural adult voices. Listeners were instructed to search for the "best voice" in the series.

2. Experiments

2.1 Stimuli

The stimuli were vowels (/i/, / ɑ/, / u /) extracted from /hVd/ words ("heed," "hod," "who'd") spoken by two men and two women from the reference database. Vowel onset was defined as the first clearly voiced pitch period; vowel offset was defined by the last pitch period before the stop closure. Vowel triads were formed by concatenating the three vowels, with each vowel separated from its neighbor by a 50-ms silent interval. Frequency-scaled versions were constructed using the STRAIGHT vocoder (Kawahara, 1997). The vowels were recorded and synthesized at a sample rate of 48 kHz and were scaled to the maximum peak-to-peak amplitude of the 16-bit quantization range.

Experiment 1: In experiment 1 vowel triads were processed using the STRAIGHT vocoder to generate two 25-step logarithmically spaced continua: one with the spectrum envelope (mean FF) scale factor fixed at 1.0, and *F*0 varied over ±2 oct; the other with *F*0 scale factor fixed at 1.0, and a range of FF scale factors spanning ±0.66 oct. The endpoints of the continua were chosen to cover (and extend somewhat beyond) the natural ranges found in adult male and female voices. In each continuum, step 13 corresponded to a scale factor of 1.0 (i.e., the original voice).

Example sound files from the *F*0 and FF continua are included for one of the male voices.

Mm.1. [*F*0 continuum steps 7, 13, and 19 (*F*0 scale factor 0.5, 1.0, and 2.0; FF scale factor 1.0)]. (21 Kb). This is a file of type "wav."

Mm.2. [FF continuum steps 7, 13, and 19 (FF scale factors 0.80, 1.0, and 1.26; *F*0 scale factor 1.0)]. (21 Kb). This is a file of type "wav."

Two further experiments were conducted to ensure that listeners' preferences were not based solely on differences in synthesis quality associated with large versus small frequency shifts.

Experiment 2: Two continua were constructed for the four voices as described in experiment 1, except that each voice was "gender transposed" on the property that was not being adjusted. Thus, male voices were assigned the mean female *F*0 for the FF adjustment task, while they were assigned the mean female FF for the *F*0 adjustment task. Similarly, female voices were assigned male values on the fixed property. The assigned scale factors were determined by acoustic measurements from the reference database.[2]

Experiment 3: Experiment 3 transformed each voice to a "gender-neutral" position, midway (on a log frequency scale) between the average male and average female in the vowel database. Specifically, the male *F*0 and FF were scaled up by a factor of 1.4 and 1.1071, respectively, while the female *F*0 and FF were scaled down by a factor of 1/1.4 and 1/1.1071.

2.2 Listeners

Three separate groups of 10–14 young adult listeners participated in the three experiments. All had normal hearing and received research participation credit in the Psychology program at the University of Texas at Dallas. They were monolingual speakers of American English with no reported history of speech, hearing, or language disorders.

2.3 Procedure

Stimuli were presented on-line at 48 kHz, low-pass filtered at 10 kHz (Tucker-Davis Technologies FT5), attenuated (TDT PA4), amplified (TDT HB5), and presented diotically via headphones (Sennheiser HD-414) at a comfortable listening level. Listeners were tested individually in a sound-treated booth. In separate conditions (with order counterbalanced) listeners adjusted either the mean *F*0 or mean FF of the vowel triad: by moving the computer mouse to the left for stimuli with lower *F*0 (or FF), and moving it to the right for higher *F*0 (or FF). Each condition included 40 sets of vowel triads (4 talkers $\times$ 10 repetitions) presented in randomized order. When the mean *F*0 was adjusted, the time-varying formant pattern of the original utterance (male or female) was unmodified; when the mean FF was adjusted, the original *F*0 contour was preserved. Vowel triads were presented at intervals of 0.3 s.

Listeners were instructed to search for the most natural-sounding voice from the continuum. They were told to listen to a wide range of different voices by moving the mouse back and forth, and when satisfied that they had found the "best voice," to press the ⟨Enter⟩ key to record their response and continue to the next item. The experiment was self-paced, with an approximate 4-s delay between the best-choice response and the presentation of the next vowel. Experimental sessions (including a few familiarization trials) were completed in about 50 min. The computer monitor was turned off to eliminate any visual cues to stimulus location in the continuum. The starting position (i.e., the step number along the continuum) varied randomly from trial to trial to encourage listeners to adopt a careful search for the best voice. To dissuade listeners from selecting the middle item from the continuum, either the five lowest or five highest items from the continuum were omitted, again using a random sequence. Listeners were encouraged to "bracket" their responses, i.e., to explore a range of higher and lower values before making their final choice, and to sample the entire continuum on each trial.

2.4 Results

The results of each experiment were summarized in terms of the mean *F*0 or mean FF for listeners' best-choice responses. Mean *F*0 was estimated using the algorithm provided by

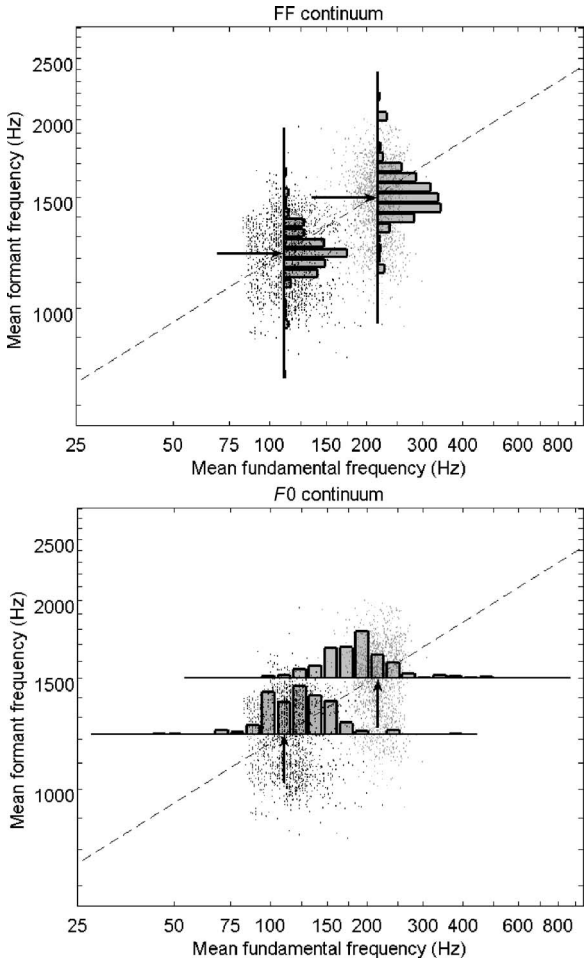


Fig. 2. Response histograms for experiment 1. Upper Panel: FF continuum. Left bars: male voices; right bars: female voices. Left/right arrows: measured FF means from the reference database for male and female voices, respectively. Lower Panel: *F*0 continuum. Upper bars: female voices; lower bars: male voices. Upper/lower arrows: measured *F*0 means for female and male voices, respectively.

STRAIGHT (and averaged across all voiced frames, at a 1-ms frame rate). Mean FF was estimated using the geometric mean of the formants F1, F2, F3 sampled at the 20% point in the voiced portion of each vowel using the algorithm developed by Nearey (Nearey *et al.*, 2002), and averaged across the three vowels in the triad. For some of the stimuli (e.g., those with high *F*0) it was difficult to obtain accurate FF measurements. To eliminate the additional variability associated with measurement error, the measurements reported below were obtained from the baseline (unshifted) stimuli and scaled up or down according to the *F*0 or FF scale factor. However, analyses carried out using the measurements of the frequency-scaled stimuli yielded the same statistical pattern of results. The data from each experiment were subjected to a one-way repeated measures analysis of variance, using the mean FF (or *F*0) of the best-choice response as the dependent variable (averaged across the ten replications per condition per listener).

Experiment 1: Figure 2 displays the histograms of best-choice responses from 13 listeners[3] superimposed on acoustic measurements from the reference database. The response distribution for the FF continuum (upper panel) was centered near the FF mean for males (left bars/arrow) and females (right bars/arrow) in the reference database. Table 1 lists the means and

Table 1. Summary of best-choice responses in experiment 1. Means and standard deviations (std) across the *N* listeners are shown together with measured mean FF and mean *F*0 for the individual talkers (columns 4 and 7) and vowel database means (columns 5 and 8) for males (rows 1 and 2) and females (rows 3 and 4) in the Assmann and Katz (2000) database.

		Best-choice FF (listeners)	Mean FF (measured)		Best-choice *F*0 (listeners)	Mean *F*0 (measured)	
Talker	*N*	Mean (std)	Talker Mean	Gender Mean	Mean (std)	Talker mean	Gender Mean
M1	13	1131 (34)	1140	1158	109 (13)	93	111
M2	13	1206 (42)	1200	1158	129 (16)	123	111
F1	13	1440 (48)	1439	1441	189 (28)	222	219
F2	13	1424 (54)	1434	1441	192 (29)	236	219

standard deviations (across listeners) separately for each talker. For comparison, the measured means for each talker's productions of /i/, /ɑ/, and /u/ are included, along with male or female averages for these vowels in the reference database. For each of the four voices, the mean best-choice response was close to the mean FF of the original. Analysis of variance revealed a significant effect of talker [$F(3,36)=228.85; p<0.01$]. *Post-hoc* (Scheffé) tests indicated a significant difference between males and females, a significant difference between the two male talkers ($p<0.05$), but no difference between the two female talkers.

For the *F*0 continuum (lower panel) the best-choice *F*0 was on average slightly higher than the original *F*0 for the male voices and lower for the female voices (notably, however, the rank ordering of the original *F*0 values was preserved across the four voices). There was a significant effect of talker [$F(3,36)=55.43; p<0.01$]. *Post-hoc* (Scheffé) tests indicated a significant difference between males and females ($p<0.05$) but no difference between the two talkers of the same sex. For the two male voices, the mean best-choice responses were close to the corresponding male talker's mean *F*0; for the two female voices the choices were somewhat lower than both the corresponding female talkers' measured *F*0.

Overall, listeners selected scale factors that closely matched the natural voices, suggesting that their judgments were based on an internalized representation of the natural covariation of *F*0 and FF. To provide a more stringent test of this idea, experiment 2 applied a "gender-swapping" transformation, in which the male voices were transposed to the mean *F*0 or mean FF of average female voices and female voices were transposed to male values.

Experiment 2: Figure 3 shows response histograms across 14 listeners for "gender-swapped" voices. If listeners rely on the natural covariation of mean *F*0 and mean FF to select their best-choice responses, then voices that were originally male should be assigned female FF

Table 2. Best-choice responses in experiment 2. Means and standard deviations (std) across the *N* listeners are shown together with vowel database mean FF (column 4) and *F*0 (column 6) for the *opposite* gender (i.e., females in rows 1 and 2, males in rows 3 and 4).

		Best-choice FF (listeners)	Mean FF (measured)	Best-choice *F*0 (listeners)	Mean *F*0 (measured)
Talker	*N*	Mean (std)	*Opposite* gender mean	Mean (std)	*Opposite* gender mean
M1	14	1473 (119)	1441	209 (32)	219
M2	14	1451 (79)	1441	207 (41)	219
F1	14	1231 (141)	1158	138 (25)	111
F2	14	1193 (103)	1158	121 (22)	111

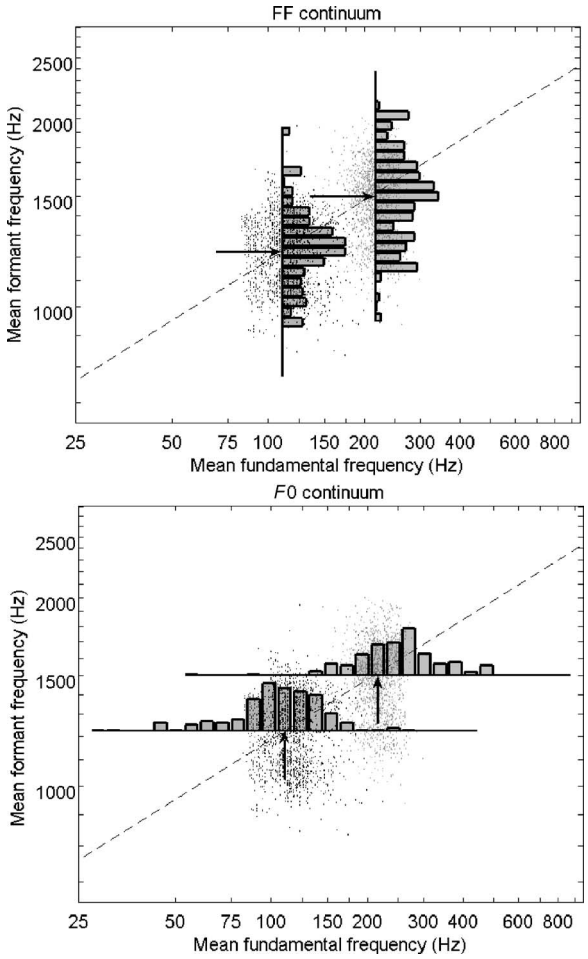


Fig. 3. Response histograms for experiment 2. Upper Panel: FF continuum. Left bars: original female voices with *F*0 gender-swapped to average male value; right bars: original male voices with *F*0 gender-swapped to average female value. Left/right arrows: measured FF means from the reference database for male and female voices, respectively. Lower Panel: *F*0 continuum. Upper bars: original male voices with FF gender-swapped to average female value; lower bars: original female voices with FF gender-swapped to average male value. Upper/lower arrows: measured *F*0 means for female and male voices, respectively.

and *F*0 values and *vice versa*. The results show that this is indeed the case. The histograms show peaks close to the measured means for the opposite gender, although response distributions for mean FF are somewhat broader and less smooth compared to experiment 1.

Similar to experiment 1, there was a significant effect of talker for the FF continuum $[F(3,39)=35.65; p<0.01]$. *Post-hoc* (Scheffé) tests indicated a significant difference between (gender-swapped) male and female stimuli, but no difference between the two males or two females. Table 2 shows that for each of the four voices, the mean best-choice response was within 75 Hz of the (gender-swapped) mean FF, and more than 200 Hz from that of the original voice. A similar pattern was observed for *F*0, with a significant effect of talker $[F(3,39)=70.94; p<0.01]$. *Post-hoc* (Scheffé) tests indicated a significant difference between males and females, but no difference between the two male or two female talkers. For each of the four voices, the mean best-choice *F*0 was within 30 Hz of the (gender-swapped) mean *F*0, and more than 80 Hz from that of the original voice.

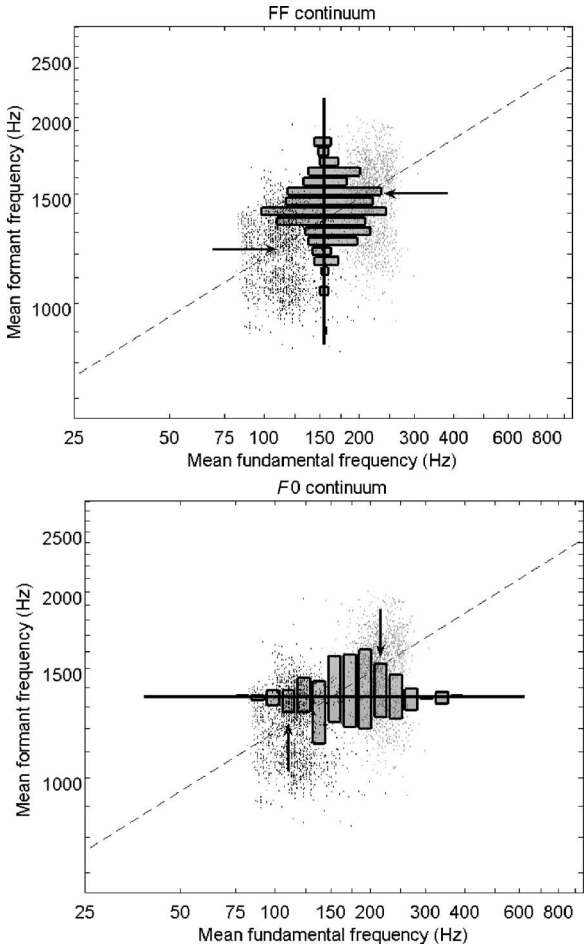


Fig. 4. Response histograms for experiment 3. Upper Panel: FF continuum (*F*0 fixed at gender-neutral mean). Left bars: male voices; right bars: female voices. Left/right arrows: measured FF means from the reference database for male and female voices, respectively. Lower Panel: *F*0 continuum (FF fixed at gender-neutral mean). Upper bars: female voices; lower bars: male voices. Upper/lower arrows: measured *F*0 means for female and male voices, respectively.

Experiment 3: Figure 4 shows that response distributions for "gender-neutral" voices are shifted to an intermediate position, midway between the database measurements for males and females. Since the male and female vowel triads were shifted to the same intermediate values, each one was predicted to be assigned the same best-choice mean FF or *F*0. An analysis of variance did not reveal a significant effect of talker for either mean FF or mean *F*0. For each of the four voices, the best-choice mean FF was close to the midpoint between the database averages for males and females. The best-choice mean *F*0 for each voice was closer to the female average *F*0, suggesting that the gender-neutral formant pattern was assimilated to the female *F*0 range to some extent.

3. General discussion

The results of the experiments indicate that listeners' judgments of the "best voice" in each series produce combinations of mean *F*0 and mean FF that match the covariance pattern observed in acoustic measurements of natural vowels. In experiment 1, listeners selected *F*0 and FF scale factors close to those of the original voice. This outcome suggests that their judgments

reflect the statistical distribution of *F*0 and FF in natural voices. However, a less interesting explanation might be that stimuli with scale factors close to 1.0 are synthesized more accurately than stimuli with large frequency shifts.

Two further experiments were conducted to ensure that listeners' preferences were not based solely on differences in synthesis quality associated with large versus small frequency shifts. Experiment 2 showed that listeners selected natural combinations of mean *F*0 and mean FF in "gender-swapped" voices. Experiment 3 extended the findings to "gender-neutral" voices positioned midway between adult male and female ranges. Both experiments confirmed the predictions; however, the response distributions for "gender-swapped" voices in experiment 2 were broader and somewhat uneven compared to those in experiment 1. Careful listening to these stimuli and acoustic analysis failed to reveal issues related to synthesis quality. An alternative explanation may be that factors other than mean *F*0 and mean FF influence listeners' judgments of voice gender. This interpretation is supported by recent findings (Assmann *et al.*, 2006) that sentences originally spoken by male talkers received higher ratings of masculinity, and sentences spoken originally spoken by females received higher ratings of femininity, even when both were frequency shifted to the same mean *F*0 and mean FF. This raises the possibility that the "gender-swapped" stimuli in experiment 2 may have provided conflicting cues concerning voice gender, and that factors other than mean *F*0 and mean FF contribute to the perception of voice gender (see also Hillenbrand, 2005).

Overall the results support the view that listeners' judgments of voice quality are influenced by the statistical distribution of mean *F*0 and FF in human voices (Nearey and Assmann, 2007). Consistent with this interpretation is the informal observation that listeners can readily identify voices with atypical combinations of mean *F*0 and mean FF as distinctive (Nearey, 1989). It is likely that such judgments include a component of perceived size (Smith *et al.*, 2005) as well as voice gender (Smith and Patterson, 2005; Assmann *et al.*, 2006).

Acknowledgments

This work was supported by National Science Foundation Grant No. 0318451 (PFA) and Social Sciences and Humanities Research Council Regular Research Grant 410-2005-1329 (TMN). Thanks to Jim Hillenbrand and two anonymous reviewers for thoughtful comments on the manuscript, Hideki Kawahara for providing the STRAIGHT software, and Derrick Chen for help in carrying out the experiments. Portions of this work were reported at the 4th joint meeting of the Acoustical Society of America and the Acoustical Society of Japan [J. Acoust. Soc. Am. **120**(5), 3248 (2006)].

References and links

[1]STRAIGHT applies the envelope scale factor to the frequency axis of a nonparametrically smoothed spectrum. Since formant frequencies are used throughout the paper to summarize key properties of the spectrum envelope believed to be relevant to perception, we refer to changes in the envelope scaling in STRAIGHT as FF scaling.

[2]To obtain the scale factors, the formant frequencies (F1, F2, and F3) were measured near the onset of each vowel in the reference database. The geometric mean was calculated across the three formants and across all vowel tokens, separately for the male talkers (1221Hz) and female talkers (1497Hz). The geometric mean of F1, F2, and F3 was then calculated for each of the vowel triads used in the experiment. The spectrum envelope scale factor was obtained for each vowel triad by dividing its geometric mean by the reference database mean for the opposite gender. F0 scale factors were obtained in the same way, except that F0 measurements were averaged across all voiced frames in the vowel.

[3]The data of three additional listeners were excluded from the analysis because they had difficulty with the instructions and/or failed to complete the experiment.

Assmann, P. F., and Katz, W. F. (**2000**). "Time-varying spectral change in the vowels of children and adults," J. Acoust. Soc. Am. **108**, 1856–1866.

Assmann, P. F., Nearey, T. M., and Scott, J. M. (**2002**). "Modeling the perception of frequency-shifted vowels," *Proceedings of the 7th International Conference on Spoken Language Processing*, Denver, CO, pp. 425–428.

Assmann, P. F., Dembling, S., and Nearey, T. M. (**2006**). "Effects of frequency shifts on perceived naturalness and gender information in speech," *Proceedings of the 9th International Conference on Spoken Language Processing*,

Pittsburgh, PA, pp. 889–892.
Chiba, T., and Kajiyama, M. (**1941**). *The Vowel: Its Nature and Structure* (Tokyo-Kaiseikan, Tokyo).
Daniloff, R. G., Shriner, T. H., and Zemlin, W. R. (**1968**). "Intelligibility of vowels altered in duration and frequency," J. Acoust. Soc. Am. **44**, 700–707.
Fant, G. (**1970**). *Acoustic Theory of Speech Production*, 2nd ed. (Mouton, Paris).
Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511–1522.
Hillenbrand, J. M. (**2005**). "The role of fundamental frequency and formants in the perception of speaker sex," J. Acoust. Soc. Am. **118**, 1932–1933(A).
Katz, W. F., and Assmann, P. F. (**2001**). "Identification of children's and adults' vowels: Intrinsic fundamental frequency, fundamental frequency dynamics, and presence of voicing," J. Phonetics **29**, 23–51.
Kawahara, H. (**1997**). "Speech representation and transformation using adaptive interpolation of weighted spectrum: vocoder revisited," *Proc. IEEE Int. Conf. on Acoustics, Speech & Signal Processing* (ICASSP '97, Munich, Germany), Vol. **2**, pp. 1303–1306.
Kawahara, H., Masuda-Katsuse, I., and de Cheveigné, A. (**1999**). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based *F*0 extraction," Speech Commun. **27**, 187–207.
Nearey, T. M. (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.
Nearey, T. M., and Assmann, P. F. (**2007**). "Probabilistic 'sliding-template' models for indirect vowel normalization," in *Experimental Approaches to Phonology*, edited by M. J. Solé, P. S. Beddor, and M. Ohala (Oxford University Press, Oxford, UK), pp. 246–269.
Nearey, T. M., Hillenbrand, J. M., and Assmann, P. F. (**2002**). "Evaluation of a strategy for automatic formant tracking," J. Acoust. Soc. Am. **112**, 2323(A).
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of vowels," J. Acoust. Soc. Am. **24**, 175–184.
Smith, D. R., and Patterson, R. D. (**2005**). "The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age," J. Acoust. Soc. Am. **118**, 3177–3186.
Smith, D. R., Patterson, R. D., Turner, R., Kawahara, H., and Irino, T. (**2005**). "The processing and perception of size information in speech sounds," J. Acoust. Soc. Am. **117**, 305–318.
Titze, I. R. (**1989**). "Physiologic and acoustic differences between male and female voices," J. Acoust. Soc. Am. **85**, 1699–1707.

Adaptive microphone array for unknown desired speaker's transfer function

Istvan I. Papp
Faculty of Technical Sciences, University of Novi Sad, 21000 Novi Sad, Serbia
istvan.papp@micronasnit.com

Zoran M. Saric
MicronasNIT, Fruškogorska 11a, 21000 Novi Sad, Serbia
zoran.saric@micronasnit.com

Slobodan T. Jovicic
School of Electrical Engineering, University of Belgrade, 11000 Belgrade, Serbia
jovicic@etf.bg.ac.yu

Nikola Dj. Teslic
Faculty of Technical Sciences, University of Novi Sad, 21000 Novi Sad, Serbia
nikola.teslic@micronasnit.com

Abstract: The main drawback of minimum variance distortionless response (MVDR) beamformer is the cancellation of the desired speech signal and its degradation in multi-path wave propagation environment. To make the adaptive algorithm robust against room reverberation and to prevent desired signal cancellation an estimation of unknown desired speaker's transfer function was proposed. The estimation is based on the signal and the interference covariance matrices. The estimated transfer function is then applied to the MVDR beamformer. The proposed algorithm was tested on a simulated room with reverberation. The results showed better quality of the restored speech compared to some typical adaptive algorithms.

PACS numbers: 43.60.Fg, 43.60.Mn, 43.72.Dv [JC]
Date Received: April 7, 2007 **Date Accepted:** April 30, 2007

1. Introduction

The problem of high quality speech recording in a room with reverberation and the cocktail-party interference has been long under consideration. It has been established that microphone arrays, compared to a single microphone, render a better quality of speech recording. The commonly used minimum variance distortionless response (MVDR) beamformer is the optimal estimator for the Gaussian process and for the known desired signal transfer function.[1–3] In the reverberant room, actual transfer function is not known. The incomplete knowledge of the transfer function causes desired speaker cancellation.[4,6] In order to reduce this cancellation, some linear and quadratic constraints have to be applied.[3] Unfortunately, these constraints can degrade the interference suppression performance significantly. The alternative methods exploit the nonstationary nature of the speech signal[4–7] by estimating array weights during the pauses in the speech. In this case there is no desired speech cancellation.[6]

In this paper a two step minimum variance beamforming algorithm, denoted by TS-MV, is proposed. In the first step the unknown transfer function of the desired speaker is estimated using both estimates of the signal and interference covariance matrices. In the second step the estimated transfer function is used for MVDR weights calculation to prevent desired signal cancellation. In addition, it is shown that the proposed estimation of the transfer function is robust against imperfect signal covariance matrix. The proposed estimation algorithm is experimentally tested in a simulated room with reverberation. Experimental results showed the improvement in restored speech quality compared to some similar algorithms.

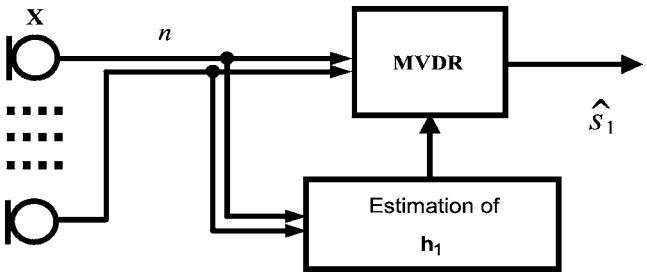


Fig. 1. General structure of the adaptive beamformer.

2. Baseline approach

Let us assume a reverberant room with an array of n microphones, desired signal s_1, and m acoustical interferences $s_2, \ldots, s_{m+1}$. The microphone signals are processed in discrete Fourier transform (DFT) domain. All signals are represented by the complex DFT coefficients with central frequency f. For the sake of simplicity the index f will be omitted, i.e., $x=x(f)$. Column vector $\mathbf{X}$ of the n microphone signals can be expressed by

$$\mathbf{X} = \mathbf{S} + \mathbf{U}, \quad \mathbf{S} = \mathbf{h}_1 s_1, \tag{1}$$

where n-column vector $\mathbf{S}$ is the room response to the desired signal s_1 excitation, and n-column vector $\mathbf{h}_1$ is its transfer function containing both direct path and reflections. The vector $\mathbf{U}$ is the sum of responses to the interference signal vector $\mathbf{S}_I$, $\mathbf{S}_I=[s_2, \ldots, s_{m+1}]'$ and to the uncorrelated microphone noise $\mathbf{N}$, $\mathbf{N}=[n_1 \ldots n_n]'$ expressed by

$$\mathbf{U} = \mathbf{H}_I \mathbf{S}_I + \mathbf{N}, \tag{2}$$

where $\mathbf{H}_I$ is $n \times m$ interference transfer matrix. In the rest of the paper superscript ^{H} denotes a complex conjugate transpose, * denotes complex conjugation, $'$ denotes matrix/vector transposition, and $E\{\cdot\}$ denotes the statistical expectation operator. Microphone signals are processed by adaptive algorithm displayed in Fig. 1. Output signal $\hat{s}_1$ is the weighted sum of the microphone signals, $\hat{s}_1=\mathbf{W}^H\mathbf{X}$, where $\mathbf{W}$ is weight vector of the MVDR beamformer expressed by[1,2]

$$\mathbf{W} = \frac{\mathbf{\Phi}_{U,U}^{-1}\mathbf{h}_1}{\mathbf{h}_1^H\mathbf{\Phi}_{U,U}^{-1}\mathbf{h}_1}. \tag{3}$$

The interference cross-spectral matrix $\mathbf{\Phi}_{U,U}$, $\mathbf{\Phi}_{U,U}=E\{\mathbf{U}\mathbf{U}^H\}$ has to be estimated from available measurements $\mathbf{X}$ during the absence of desired speech.[6] The problem is that transfer vector $\mathbf{h}_1$ is not known in reverberant environment. The use of the direct path transfer vector instead of $\mathbf{h}_1$ causes unwanted desired speech cancellation.[4,6] To prevent this, the actual $\mathbf{h}_1$ has to be estimated.

3. Transfer function estimation

Transfer vector $\mathbf{h}_1$ can be estimated from the signal covariance matrix $\mathbf{\Phi}_{S,S}$ that is defined by $\mathbf{\Phi}_{S,S}=E\{\mathbf{X}\mathbf{X}^H\}$ under the assumption that only desired signal s_1 and noise $\mathbf{N}$ are present. From Eqs. (1) and (2), it follows

$$\mathbf{\Phi}_{S,S} = \Phi_{s,s}\mathbf{h}_1\mathbf{h}_1^H + \Phi_{N,N}\mathbf{I}, \tag{4}$$

where $\Phi_{s,s}$ is desired signal power, and $\Phi_{N,N}$ is uncorrelated noise power. Using the principal eigenvector $\mathbf{v}_p$ of $\mathbf{\Phi}_{S,S}$, the estimate of $\mathbf{h}_1$ is

$$\hat{\mathbf{h}}_1 = C_\varphi \mathbf{v}_p, \quad C_\varphi = \exp(-j\varphi), \tag{5}$$

where C_φ is a unit magnitude complex multiplier that influences only the signal delay.[7] The phase compensation will be defined in Sec. 4. There is a problem in signal covariance matrix estimation because at least one of the interferences is almost always present.[8] Hence, we must take into account that $\mathbf{\Phi}_{S,S}$ is contaminated with $\mathbf{\Phi}_{U,U}$ by

$$\hat{\mathbf{\Phi}}_{S,S} = \alpha \mathbf{\Phi}_{S,S} + (1-\alpha)\mathbf{\Phi}_{U,U} \quad 0.5 < \alpha < 1, \tag{6}$$

where α is a positive scalar. The second term in Eq. (6) significantly degrades the estimation of $\mathbf{h}_1$. The improved estimate of $\mathbf{h}_1$ can be defined under the following assumptions:

(A1)
The estimate of the interference covariance matrix $\mathbf{\Phi}_{U,U}$ is available.
(A2)
The number of the interference signals is less than the number of microphones ($m < n$).
(A3)
Uncorrelated noise power $\Phi_{N,N}$ is much less than the desired signal power $\Phi_{N,N} \ll \Phi_{s,s}$.

Let us define auxiliary matrix $\mathbf{\Phi}$, $\mathbf{\Phi} = \hat{\mathbf{\Phi}}_{S,S}\mathbf{\Phi}_{U,U}^{-1}\mathbf{\Phi}_{U,U}^{-1}\hat{\mathbf{\Phi}}_{S,S}$. Under assumptions A1, A2, A3, the principal eigenvector of $\mathbf{\Phi}$ can be used as an approximation of the principal eigenvector of $\mathbf{\Phi}_{S,S}$, required in Eq. (5). The proof is given in Appendix A 1.

4. Proposed algorithm

Finally, the proposed two step minimum variance (TS-MV) algorithm can be described by:

Step 1: Estimate of $\mathbf{h}_1$

(i) Estimate $\mathbf{\Phi}_{U,U}$ on pause intervals of desired speech signal, and $\mathbf{\Phi}_{S,S}$ on intervals with high speech to interference ratio. Calculate $\hat{\mathbf{\Phi}}_{S,S}$ by Eq. (6).

(ii) Calculate auxiliary matrix $\mathbf{\Phi}$, $\mathbf{\Phi} = \hat{\mathbf{\Phi}}_{S,S}\mathbf{\Phi}_{U,U}^{-1}\mathbf{\Phi}_{U,U}^{-1}\hat{\mathbf{\Phi}}_{S,S}$, calculate principal eigenvector $\mathbf{v}_p$ of the matrix $\mathbf{\Phi}$, and estimate transfer vector $\mathbf{h}_1$ by Eq. (5).

Step 2: Apply MVDR

(iii) Apply diagonal loading to the interference covariance matrix $\mathbf{\Phi}_{U,U}$ by $\widetilde{\mathbf{\Phi}}_{U,U} = (\mathbf{\Phi}_{U,U} + \beta\mathbf{I})$, to make the MVDR beamformer robust against steering error and room reverberation.[3,6] Scalar $\beta, \beta > 0$, makes a compromise between stability and high interference suppression.

(iv) Calculate MVDR weight vector $\mathbf{W}$ as $\mathbf{W} = \widetilde{\mathbf{\Phi}}_{U,U}^{-1}\hat{\mathbf{h}}_1 \big/ \hat{\mathbf{h}}_1^H \widetilde{\mathbf{\Phi}}_{U,U}^{-1}\hat{\mathbf{h}}_1$.

(v) As the estimated $\mathbf{h}_1$ has random phase shift factor C_φ (5), apply phase compensation by[7] $\bar{\mathbf{W}} = \left(\mathbf{W}^H\mathbf{h}_d / \mathbf{W}^H\mathbf{h}_d\right)\mathbf{W}$, $\mathbf{h}_d = [1 \; e^{-j2\pi f\tau} \ldots e^{-j2\pi f(n-1)\tau}]$, τ delay on adjacent microphones, where $\mathbf{h}_d$ is the direct path transfer vector.

5. Experimental results

The proposed TS-MV algorithm has been examined in a room with reverberation simulated by Allen's image method.[8] The room reverberation time was $T_{60} = 270$ ms. The number of sources was 2: source s_1 was the desired speaker and source s_2 was the interference (Fig. 2.). In the experiment 1 the interference s_2 was at position s_2' (easier to suppress) while in the experiment 2 it was at position s_2'' (harder to suppress). Critical distance boundary was calculated from the room model for which the direct path power is equal to the reverberant power. The microphone array consisted of eight microphones with equidistant spacing of 6 cm. The sampling rate of the speech signals was 10 kHz, while the length of data processing block was 2048 points. Signal of the microphone 1 for position s_2' is in Mm. 1. The following algorithms were compared: (1) The conventional beamformer (CBF) (Mm. 2), (2) Generalized sidelobe canceller (GSC) (Mm. 3),

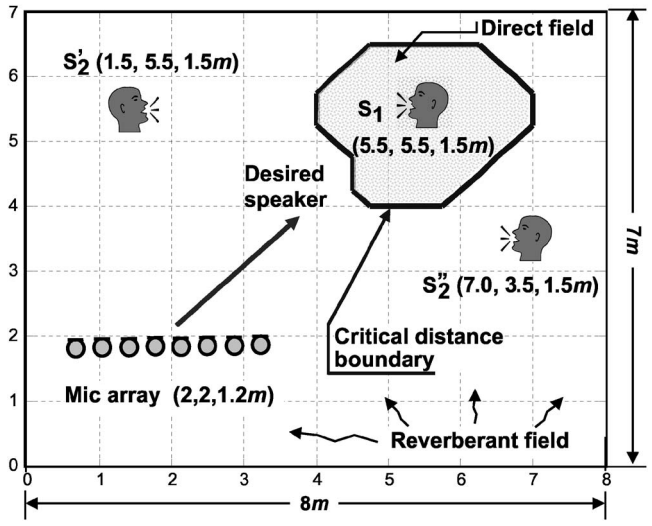


Fig. 2. Simulated room with a reverberation time of 270 ms and a microphone array with eight microphones.

(3) GSC with weights estimated within hand labeled pause intervals,[6] (4) GSC with weights estimated under an ideal scenario where only interference is present,[6] (5) GEVBF with hand-labeled signal and pause intervals,[7] (6) GEVBF with signal and interference covariance matrices estimated under an ideal scenario where either a desired speech or interference is exclusively present, (7) Proposed TS-MV algorithm with covariance matrices $\mathbf{\Phi}_{S,S}$ and $\mathbf{\Phi}_{U,U}$ estimated within hand labeled time intervals of speech and pause, respectively (Mm. 4), (8) Proposed TS-MV algorithm with covariance matrices $\mathbf{\Phi}_{S,S}$ and $\mathbf{\Phi}_{U,U}$ estimated under an ideal scenario, where either a speech signal or interference is exclusively present.

In all algorithms, except CBF, the diagonal loading of the interference covariance matrix is applied to reduce desired signal cancellation.[3] The quality of the speech signal restoration was evaluated by the cepstral distortion measure, and the results are presented in Table 1. As was expected, the worst result is obtained with CBF algorithm. A better result is obtained by the full adaptation GSC, but the restored signal is obviously degraded due to signal cancellation. Further improvement is obtained by GSC weights estimated within the hand labeled pauses (Table 1, row 3). It should be pointed out that the best achievable quality by the MVDR criterion is under the ideal scenario where the desired signal is muted and only interference is present (Table 1, row 4). The additional improvement is obtained by the GEVBF algorithm that maximizes signal to noise ratio.[7] The best results are obtained by the proposed TS-MV algorithm.

Table 1. Cepstral distortion measures of restored signal.

	Cepstral distortion measure	
Estimation algorithms	Experiment 1	Experiment 2
1. CBF	0.860	1.134
2. Ordinary GSC with diagonal loading	0.758	0.984
3. GSC – hand-labeled pauses	0.607	0.800
4. GSC – ideal scenario	0.524	0.638
5. GEVBF – hand-labeled intervals	0.479	0.545
6. GEVBF – ideal scenario	0.453	0.506
7. TS-MV – hand-labeled intervals	0.414	0.427
8. TS-MV – ideal scenario	0.362	0.369

Mm. 1 Microphone 1 signal for position s_2' on Fig. 2 (201 kb). This is a file of type "wav."

Mm. 2 CBF output (209 kb). This is a file of type "wav."

Mm. 3 Output of the MVDR (GSC) with diagonal loading (209 kb). This is a file of type "wav."

Mm. 4 Output of the proposed TS-MV algorithm (209 kb). This is a file of type "wav."

6. Conclusions

In this paper, a two step minimum variance (TS-MV) algorithm for acoustical interference suppression in reverberant environment is proposed. In the first step the unknown desired speaker's transfer function is estimated while this estimate is then used for MVDR beamformer in the second step. This estimate reduces cancellation of the desired speaker signal, while at the same time preserves high noise suppression.

An improved estimate of the unknown transfer function is obtained using both signal and interference covariance matrices. The proposed estimation algorithm is robust against imperfect signal covariance matrix estimation. An additional robustness of the algorithm is obtained by diagonal loading of the interference covariance matrix. Tests in the simulations of the room with reverberation proved the superior performance of the algorithm.

Acknowledgments

This work was partially supported by Ministry of Science and Environment Protection of the Republic Serbia under Grant No. OI-1784.

Appendix A1 Approximation of the principal eigenvector of $\hat{\mathbf{\Phi}}_{S,S}$

Taking into account Eq. (6) the auxiliary matrix $\mathbf{\Phi}$ can be expressed by

$$\mathbf{\Phi} = \hat{\mathbf{\Phi}}_{S,S}\mathbf{\Phi}_{U,U}^{-1}\mathbf{\Phi}_{U,U}^{-1}\hat{\mathbf{\Phi}}_{S,S} = [\alpha\mathbf{\Phi}_{S,S}\mathbf{\Phi}_{U,U}^{-1} + (1+\alpha)\mathbf{I}][\alpha\mathbf{\Phi}_{U,U}^{-1}\mathbf{\Phi}_{S,S} + (1-\alpha)\mathbf{I}]. \quad \text{(A1)}$$

Inverse matrix $\mathbf{\Phi}_{U,U}^{-1}$ can be decomposed by its eigenvectors $\mathbf{u}_i, i=1,n$

$$\mathbf{\Phi}_{U,U}^{-1} = \sum_{i=1}^{m} \frac{1}{\lambda_i}\mathbf{u}_i\mathbf{u}_i^H + \sum_{i=m+1}^{n} \frac{1}{\sigma_N^2}\mathbf{u}_i\mathbf{u}_i^H, \quad \text{(A2)}$$

where $\mathbf{u}_i, i=1,m$, are eigenvectors of the interference signals subspace, and $\mathbf{u}_i, i=m+1,n$ are eigenvectors of the noise subspace; $\lambda_i, i=1,m$ are corresponding eigenvalues of the interference signals subspace, and $\sigma_N^2, \sigma_N^2=\Phi_{N,N}$ is common eigenvalue for eigenvectors of the noise subspace. Substituting (A2) and Eq. (4) into (A1) and taking into account $\mathbf{h}_1^H\mathbf{h}_1=1$ and $\sigma_N^2/\Phi_{s,s}\mathbf{I} \rightarrow \mathbf{0}$, ($\Phi_{s,s}$, is signal power), the auxiliary matrix $\mathbf{\Phi}$ can be approximated by

$$\mathbf{\Phi} \approx \alpha^2\Phi_{s,s}^2\gamma\mathbf{v}_p\mathbf{v}_p^H, \quad \text{(A3)}$$

where γ is a real positive constant defined by $\gamma=\mathbf{v}_p^H\mathbf{\Phi}_{U,U}^{-1}\mathbf{\Phi}_{U,U}^{-1}\mathbf{v}_p$. From Eq. (A3) it is clear that the principal eigenvector of $\mathbf{\Phi}$ is approximately equal to $\mathbf{v}_p$, e.g., the principal eigenvector of $\hat{\mathbf{\Phi}}_{S,S}$.

References and links

[1]R. A. Monzingo and T. W. Miller, *Introduction to Adaptive arrays*, (Wiley, New York, 1980).
[2]K. U. Simmer *et al.*, "Post-filtering techniques," in *Microphone Arrays: Signal Processing Techniques and Applications*, edited by M. S. Brandstein and D. B. Ward (Springer, Berlin, 2001), Chap. 3, pp. 39–60.
[3]H. L. Van Trees, *Optimum Array Processing* (Wiley-Interscience, New York, 2002).
[4]J. E. Greenberg *et al.*, "Evaluation of an adaptive beamforming method for hearing aids," J. Acoust. Soc. Am. **91**, 1662–1676 (1992).
[5]O. Hoshuyama *et al.*, "A realtime robust adaptive microphone array controlled by a SNR estimate," *Proc. ICASSP98*, 3605–3608 (1998).

[6]Z. M. Saric and S. T. Jovicic, "Adaptive microphone array based on pause detection," ARLO **5**(2), 68–74 (2004).
[7]S. T. Jovicic *et al.*, "Application of the maximum signal to interference ratio criterion to the adaptive microphone array," ARLO **6**(4), 232–237 (2005).
[8]J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**(4), 943–950 (1979).

ACOUSTICAL NEWS—USA

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

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

Call for Business Meeting of the Society

Notice. A business meeting of the Acoustical Society of America (ASA) will be held starting at 3:30 p.m. on 29 November 2007 at the Sheraton New Orleans Hotel in New Orleans, Louisiana. All ASA Fellows and Members are urged to attend for the purpose of voting on proposed amendments to the ASA Bylaws, namely Article IX Election of Officers and Executive Council and Article XI Meetings.

Motions to be presented at the business meeting. The motions concern two clauses of the Bylaws, namely Section 2 (Nomination) of Article IX Election of Officers and Executive Council and Article XI Meetings. A proposed amendment may be adopted by a two-thirds vote of the Members and Fellows present and voting in person. Additions to the wording of Article IX and Article XI are shown in bold type and deletions to the wording by means of a strike through.

Motion. To approve the following revised version of Article IX Election of Officers and Executive Council, Section 2 Nomination:

The President, with the approval of the Executive Council, shall appoint a Nominating Committee of six Members or Fellows of the Society, at least one of whom shall be a past President of the Society. At least ninety days prior to the date fixed by the Executive Council for the annual election, the Executive Director shall publish in the Journal **or another official publication of the Society** an announcement of the election and the committee's nominations for the offices to be filled. Additional candidates for such offices may be proposed by any Member or Fellow in good standing by letter received by the Executive Director not less than sixty days prior to the election date, and the name of any eligible candidate so proposed by ~~twenty~~ **fifty** Members or Fellows shall be entered on the ballot.

Rationale for the proposed amendments to Article IX, Section 2. 1) Adding another publication in which proposed bylaws amendments may be published, for example, ASA's new magazine, *Acoustics Today*, provides the Society with greater flexibility for notifying members of proposed changes. 2) The number of signatures required to add a candidate to a ballot was set at 20 when the ASA had 1000 members. The Executive Council felt that the number should be raised to 50 as ASA now has over 7000 members.

Motion. To approve the following revised version of Article XI Meetings:

Meetings of the Society and of the Executive Council shall be held at such times and places and upon such notice as the Executive Council may from time to time determine. ~~Twenty~~ **One hundred** Members or Fellows present in person shall constitute a quorum at meetings of the Society, and a majority of the elected members of the Executive Council shall constitute a quorum at meetings of the Executive Council, but a less number may in each case adjourn the respective meetings from time to time. The Executive Council shall determine the order of business at meetings of the Society.

Rationale for the proposed amendment to Article XI Meetings. ASA's attorney has advised the Society that the lawful number for a quorum is now 100 for organizations of ASA's size.

Charles E. Schmid
Executive Director
Acoustical Society of America

Preliminary notice: 154th Meeting of the Acoustical Society of America

The 154th Meeting of the Acoustical Society of America will be held Tuesday through Saturday, 27 November–1 December 2007 at the Sheraton New Orleans Hotel, New Orleans, Louisiana, USA. A block of rooms has been reserved at the Sheraton New Orleans Hotel. Information about the meeting also appears on the ASA Home Page at ⟨http://asa.aip.org/meetings.html⟩.

Charles E. Schmid
Executive Director

Technical Program

The technical program will consist of lecture and poster sessions. Technical sessions will be scheduled Tuesday through Saturday, 27 November–1 December.

The special sessions described below will be organized by the ASA Technical Committees.

Special Sessions

Acoustical Oceanography (AO)

Deep and shallow seismic sensing of geological structure in the ocean bottom
(Joint with Underwater Acoustics)
Applications of seismic methods to investigate geophysical processes and geological structure in the ocean bottom
Storms and intense air–sea interactions
Passive and active acoustic remote sensing of the physical processes in storms and intense air–sea interactions

Animal Bioacoustics (AB)

Noise and wildlife: Advances in effects research
(Joint with Noise)
Effects of noise on wildlife
Sound source localization
Physiological, behavioral, and anatomical studies of sound source localization

Architectural Acoustics (AA)

Acoustics of modular construction
(Joint with Noise and ASA Committee on Standards)
Acoustic issues concerning modular building construction
Acoustics of rehearsal facilities
Room acoustic qualities and special challenges of rooms designed for rehearsal of music, speech, dance
Even better than the real thing—Rock, pop, and all that jazz!
(Joint with Musical Acoustics, Signal Processing in Acoustics, and Noise)
Design strategies, processing approaches, performance gestures, historical motivations, and architecture that affect music enjoyed in clubs, cars, homes, and bars
Impact and footfall noise
(Joint with Noise and Structural Acoustics and Vibration)
Evaluating footfall noise, relationship to impact noise, and remedies
Sound systems in large rooms and stadia
Case studies and issues concerning sound systems in large rooms and stadia

Biomedical Ultrasound/Bioresponse to Vibration (BB)

Biological effects and medical applications of stable cavitation
(Joint with Physical Acoustics)
Biological effects created by stable cavitation and medical applications of stable cavitation
Biomedical applications of acoustic radiation force
Broad array of ways acoustic radiation force is used in medicine and biology
Topical meeting on tissue response to acoustics and vibrations
Biological response of tissue to acoustics and vibration, including sonoporation, acoustic hemostasis, and stimulation of cell signal pathways

 0001-4966/2007/122(2)/681/6/$23.00

Engineering Acoustics (EA)

Infrasonic instrumentation
(Joint with ASA Committee on Standards)
Infrasonic sources, receivers, calibration, wind screens, and software

Education in Acoustics (ED)

Hands-on experiments for high school students
Experiments for high school students
Professional development programs for K-12 teachers of science
Planned to bring together successful models for promoting acoustics education. Invited speakers will bring ideas for working with teachers, and time for panel discussions will be allowed. Members are encouraged to contribute papers with their suggested ways to work with teachers and schools

Musical Acoustics (MU)

Musical pitch tracking and sound source separation leading to automatic music transcription
Various aspects of automatic music transcription including F0 detection (monophonic or polyphonic), instrument spectrum matching, and instrument/voice separation
Session in honor of Max Mathews
(Joint with Speech Communication)
Recognizing the work of Max Mathews in computer music and other areas of acoustics

Noise (NS)

Lawn, yard, and portable noise in the U.S.
Recent research and developments
Measurement of noise and noise effects on animals and humans
(Joint with Animal Bioacoustics)
Anthropogenic and non-anthropogenic factors
Rain noise
Rain noise on buildings and structures
Soundscape developments: Case studies and best practices
Recent research and projects which illustrate advances in soundscape technique

Physical Acoustics (PA)

Acoustic applications for hurricane and storm preparedness, and response
(Joint with Acoustical Oceanography)
Use of sonar to survey waterways; use of acoustics to track storms; how wind and water affect structural integrity of bridges and roadways; acoustic and vibration detection of early stages of levee failure
Ultrasound, quantum criticality, and magnetic fields
Emerging methods for the measurement of the acoustic properties of solids. Topics include non-contact methods, new approaches to imaging variation of elastic moduli, and non-destructive testing

Signal Processing in Acoustics (SP)

Distributed networks signal processors
Signal processing approaches for integrated distributed sensing systems with emphasis on data fusion for limited-bandwidth networks and autonomous communication, detection, localization, tracking and classification
Session honoring Leon Sibul
(Joint with Acoustical Oceanography and Underwater Acoustics)
Incorporation of the physics of acoustics into signal processing methods

Speech Communication (SC)

Auditory and somatosensory feedback in speech production
Recent advances in probing the role of feedback in speech production
Role of attention in speech perception
Exploring the importance of selective, sustained, and divided attention, and related cognitive mechanisms in speech perception and the acquisition of speech sounds
Speech intelligibility and the vowel space
Exploring the relations between vowel acoustics and speech intelligibility for various talkers (normal and disordered), speaking styles, speech materials, and listener populations

Structural Acoustics and Vibration (SA)

Ground vibration impact on buildings
(Joint with Noise)
Impact of ground vibrations generated by surface transportation on buildings
Modeling of vibration and radiation in complex structures
Analytic, statistical and numerical methods for modeling vibration and acoustic radiation from complex structural systems

Underwater Acoustics (UW)

Design of distributed surveillance and oceanographic monitoring systems
(Joint with Acoustical Oceanography)
Integrated sensing, processing and control concepts for distributed acoustic sensing systems with low-bandwidth, high-latency communication infrastructure. Exploitation of environmental and situational adaptivity, and collaborative processing.
Underwater reverberation measurements and modeling
Reverberation modeling, including benchmark problems/solutions and datasets appropriate for model validation

Other Technical Events

Technical Tour

A technical tour is planned with the U.S. Army Corps of Engineers to view the New Orleans levee system and "ground zero" of the levees which failed during Hurricane Katrina. Army engineers will describe the levee failure mechanisms and show work in progress to protect the city from future hurricanes.

Hot Topics

A "Hot Topics" session sponsored by the Tutorials Committee is scheduled covering the fields of Architectural Acoustics, Speech Communication, and Structural Acoustics and Vibration.

Exhibit

The meeting will be highlighted by an exhibit which will feature displays with instruments, materials, and services for the acoustical and vibration community. The exhibit, which will be conveniently located near the registration area and meeting rooms, will open at the Sheraton with a reception on Tuesday evening, 27 November, and will close Thursday, 29 November, at noon. Morning and afternoon refreshments will be available in the exhibit area.

The exhibit will include computer-based instrumentation, sound level meters, sound intensity systems, signal processing systems, devices for noise control, sound prediction software, acoustical materials, passive and active noise control systems and other exhibits on vibrations and acoustics. For further information, please contact: Robert Finnegan, American Inst. of Physics, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747; (516) 576-2433; rfinneg@aip.org.

Online Meeting Papers

The ASA provides the "Meeting Papers Online" website where authors of papers to be presented at meetings will be able to post their full papers or presentation materials for others who are interested in obtaining detailed information about meeting presentations. The online site will be open for author submissions in September. Submission procedures and password information will be mailed to authors with the acceptance notices.

Those interested in obtaining copies of submitted papers for this meeting may access the service anytime. No password is needed.

The url is ⟨http://scitation.aip.org/asameetingpapers/⟩.

Meeting Program

An advance meeting program summary will be published in the October issue of JASA and a complete meeting program will be mailed as Part 2 of the November issue. Abstracts will be available on the ASA Home Page ⟨http://asa.aip.org⟩ in October.

ASA Student Council Grants and Fellowship Workshop

The ASA Student Council is pleased to announce a workshop on fellowships and grants for students and post-doctoral members of the ASA to be offered during the New Orleans meeting.

During the workshop, representatives from the National Science Foundation (NSF), Office of Naval Research (ONR), National Institutes of Health/National Institute of Deafness and Other Communication Disorders (NIH/NIDCD), and the Acoustical Society of America Prizes and Special Fellowships committee will give short presentations on the following topics:

Eligibility for specific grants/fellowships at each level of education and post-doctoral training

- An overview of the application process, including criteria selection, review process, timelines
- Allowances that are covered under the award (e.g., stipend, travel expenses, research expenses, health insurance)
- General tips and guidelines for application submittal and essay writing

The workshop is planned for Thursday, 29 November, 5:30 p.m. to 7:00 p.m. Look for details and any changes in schedule on the ASA Student website (http://www.acosoc.org/student/), the student E-zine, and on the Student Council bulletin board at the meeting.

Tutorial lecture on weather and acoustics

A tutorial presentation on "Weather and Acoustics" will be given by Alfred Bedard of the National Oceanic and Atmospheric Administration on Tuesday, 27 November, at 7:00 p.m.

The relationships between sound and weather can be fascinating, frightening, useful and at times mystifying. This tutorial lecture explores the range of intersections between weather and acoustics. Weather can affect acoustic environments causing noise increases, noise reductions, and sound focusing. One aspect of this tutorial review results from propagation modeling, indicating that under some conditions the atmosphere can produce vertical wave guides. Conversely, sound can be used to actively interrogate the atmosphere and provide information valuable for weather prediction and warning. Probing capabilities reviewed, with examples, show that wind profiles, temperature profiles, wind shears, gravity waves and inversions can be defined acoustically. There are also possibilities for monitoring other difficult-to-observe parameters such as humidity profiles. In addition, weather processes can generate sound, detectable at long ranges using lower frequencies. Specifically, observing networks have observed infrasound from a growing number of meteorological events (e.g., severe weather, tornadoes, funnels aloft, atmospheric turbulence, hurricanes, and avalanches). Efforts to develop an infrasonic tornado detection system are described in some detail. Results indicate promise to help improve tornado detection and warning lead times, while reducing false alarms. Clear opportunities exist for infrasonic systems to provide operational weather data.

To partially defray the cost of the lecture a registration fee is charged. The fee is $15.00 USD for registration received by 29 October and $25.00 USD at the meeting. The fee for students with current ID cards is $7.00 USD for registration received by 29 October and $12.00 USD at the meeting. To register, use the registration form in the call for papers or register online at ⟨http://asa.aip.org⟩.

Short Course on Bayesian Signal Processing: Classical, modern and particle filtering methods

Signal processing methods capable of extracting the desired signal from hostile environments require approaches that capture all of the "a priori" information available and incorporate them into a processing scheme. This approach is typically model-based employing mathematical representations of the component processes involved. In this short course we develop the Bayesian approach to statistical signal processing in a tutorial fashion including the "next generation" of processors that have recently been enabled with the advent of high speed/high throughput computers. The course commences with an overview of Bayesian inference from batch to sequential processors. Once the evolving Bayesian paradigm is established, simulation-based methods using sampling theory and Monte Carlo realizations are discussed. Here the usual limitations of nonlinear approximations and non-Gaussian processes prevalent in classical nonlinear processing algorithms (e.g., Kalman filters) are no longer a restriction to perform Bayesian inference. Next, importance sampling methods are discussed and shown how they can be extended to sequential solutions. With this in mind, the concept of a particle filter, a discrete nonparametric representation of a probability distribution, is developed and shown how it can be implemented using sequential importance sampling/resampling methods to perform statistical inferences yielding a suite of popular estimators such as the conditional expectation, maximum a-posteriori and median filters. Finally, a set of applications are discussed comparing the performance of the particle filter designs with classical implementations (Kalman filters). Participants will be introduced to a variety of statistical signal processing techniques coupled with applications to demonstrate their capability.

The objective of the course is to provide an introduction to the Bayesian approach to model-based signal processors and compare their performance to classical approaches in terms of applications. We present a detailed overview of the basic Bayesian model-based processors enabling the participant to construct simple processors for further investigations.

Participants should have taken basic courses in random processes, statistics and linear systems theory.

The instructor, James V. Candy, is the Chief Scientist for Engineering and former Director of the Center for Advanced Signal & Image Sciences at the University of California, Lawrence Livermore National Laboratory as well as an Adjunct Professor at the University of California, Santa Barbara. Dr. Candy is a Fellow of the IEEE and a Fellow of the Acoustical Society of America (ASA) and Lifetime Member (Fellow) at the University of Cambridge (Clare Hall College). Dr. Candy received the IEEE Distinguished Technical Achievement Award for the "development of model-based signal processing in ocean acoustics." Dr. Candy was also recently selected as an IEEE Distinguished Lecturer for oceanic signal processing as well as presenting an IEEE tutorial on advanced signal processing available through their video website courses. He has authored three texts on signal processing, the most recent being, "Model-Based Signal Processing."

The course will be held on Saturday, 1 December, from 8:30 a.m. to 5:00 p.m.

The registration fee is $300.00 USD and covers attendance, instructional materials and coffee breaks. The number of attendees will be limited so please register early to avoid disappointment. Only those who have registered by 29 October will be guaranteed receipt of instructional materials. There will be a $50.00 USD discount for registration made prior to 29 October. Full refunds will be made for cancellations prior to 29 October. Any cancellation after 29 October will be charged a $25.00 USD processing fee. To register, use the form in the call for papers or register online at ⟨http://asa.aip.org⟩.

Student Transportation Subsidies

A student transportation subsidies fund has been established to provide limited funds to students to partially defray transportation expenses to meetings. Students presenting papers who propose to travel in groups using economical ground transportation will be given first priority to receive subsidies, although these conditions are not mandatory. No reimbursement is intended for the cost of food or housing. The amount granted each student depends on the number of requests received. To apply for a subsidy, submit a proposal (e-mail preferred) to be received by 17 October to: Elaine Moran, ASA, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, Tel: 516-576-2359, Fax: 516-576-2377, E-mail: asa@aip.org. The proposal should include your status as a student; whether you have submitted an abstract; whether you are a member of ASA; method of travel; if traveling by auto; whether you will travel alone or with other students; names of those traveling with you; and approximate cost of transportation.

Young Investigator Travel Grant

The Committee on Women in Acoustics (WIA) is sponsoring a Young Investigator Travel Grant to help with travel costs associated with presenting

a paper at the New Orleans meeting. Young professionals who have completed their doctorate in the past five years are eligible to apply if they plan to present a paper at the New Orleans meeting, are not currently students, and have not previously received the award. Each award will be of the order of $300 with three awards anticipated. Awards will be presented by check at the WIA luncheon at the meeting. Both men and women may apply. Applicants should submit a request for support, a copy of the abstract for their presentation at the meeting, and a current resume/vita which includes information on their involvement in the field of acoustics and in the ASA. Submission by e-mail is preferred. Send applications to Dr. Helen Hanson at helen.hanson@alum.mit.edu. Deadline for receipt of applications is 17 October.

Students Meet Members for Lunch

The ASA Education Committee provides a way for a student to meet one-on-one with a member of the Acoustical Society over lunch. The purpose is to make it easier for students to meet and interact with members at ASA meetings. Each lunch pairing is arranged separately. Students who wish to participate should contact David Blackstock, University of Texas at Austin, by e-mail ⟨dtb@mail.utexas.edu⟩. Please provide your name, university, department, degree you are seeking (BS, MS, or Ph.D.), research field, acoustical interests, and days you are free for lunch. The sign-up deadline is ten days before the start of the meeting, but an earlier sign-up is strongly encouraged. Each participant pays for his/her own meal.

Plenary Sessions, Awards Ceremony, Fellows' Luncheon and Social Events

Buffet socials with cash bar will be held on Wednesday and Friday evenings.

The ASA Plenary session will be held on Thursday afternoon, 29 November, at the Sheraton where Society awards will be presented and recognition of newly elected Fellows will be announced. There will also be a Business Meeting held during the Plenary Session.

A Fellows Luncheon will be held on Friday, 30 November, at 12:00 noon at the Sheraton. This luncheon is open to all attendees and their guests. To register, use the form in the call for papers or register online at ⟨http://asa.aip.org⟩.

Women in Acoustics Luncheon

The Women in Acoustics luncheon will be held on Thursday, 29 November. Those who wish to attend this luncheon must register using the form in the call for papers or online at ⟨http://asa.aip.org⟩. The fee is $15 USD (students $5 USD) for pre-registration by 29 October and $20 USD (students $5 USD) at the meeting.

Transportation and Hotel Accommodations

Air Transportation

The Louis Armstrong New Orleans International Airport, (Airport Code MSY) is served by the following airlines: AirTran, American Airlines, Continental Airlines, Delta Airlines, jetBlue, Northwest Airlines, Southwest Airlines, United Airlines, and U. S. Airways. For further information see ⟨http://www.flymsy.com⟩.

Ground Transportation

The Airport is approximately 11 miles from the Central Business District.

Taxicabs: A cab ride costs $28.00 USD from the airport to the Central Business District (CBD) for one or two persons and $12.00 USD (per passenger) for three or more passengers. Pick-up is on the lower level, outside the baggage claim area. There may be an additional charge for extra baggage. $1 fuel surcharge added to total fare.

Airport Shuttle: Shuttle service is available from the airport to the hotels in the CBD for $13.00 USD (per person, one way) or $26.00 USD (per person, round trip). Three bags per person. Call 1-866-596-2699 or (504) 522-3500 for more details or to make a reservation. Advance reservations are required 48 hours prior to travel for all ADA accessible transfers. Ticket booths are located on the lower level in the baggage claim area. $2 fuel surcharge added to total fare.

Airport-Downtown Express (E-2) Route: The Airport-Downtown Express (E-2) provides service from the Louis Armstrong New Orleans International Airport in Kenner, down Airline Drive into New Orleans. The Airport bus stop is on the second level of the Airport, near the Delta counter, in the median (look for the sign and bench). At Carrollton at Tulane it connects with RTA's 27-Louisiana and 39-Tulane buses. (Visit the RTA website to check their current schedules.) The Airport-Downtown Express (E-2) Bus picks up outside airport Entrance #7 on the upper level. The fare for the Airport-Downtown Express (E-2) is $1.10 USD. The fare boxes will accept $1, $5, $10, $20 dollar bills and all U.S. coins. The fare boxes will provide change in the form of a value card that can be used for future fares.

Automobile Rental: There are seven rental agencies with offices on the lower level of the airport.

Driving Information: From Louis Armstrong International Airport: Follow I-10 East to Poydras Street, Exit #234B. Turn left on Camp Street and proceed 3 blocks to Canal Street. The hotel is located on the right corner of Canal and Camp Streets. **From the East**: Follow I-10 West to Canal Street, Exit #235B. Turn right on Canal Street and proceed 10 blocks to Camp Street. The hotel is located on the right corner of Canal and Camp Streets. **From the South**: Follow the West Bank Expressway across the Mississippi River into downtown New Orleans and exit at Camp Street. Continue on Camp Street 4 blocks to Canal Street. The hotel is located on the right corner of Canal and Camp Streets.

Valet parking service is available on a first come first serve basis (spaces are limited). Vehicles are secured in a covered garage adjacent to the hotel. Overnight parking rate for cars is $26.95 plus tax.

Hotel Accommodations and Reservations

The meeting will be held at the Sheraton New Orleans Hotel. A block of guest rooms at discounted rates has been reserved for meeting participants at the Sheraton New Orleans Hotel.

Early reservations are strongly recommended. The reservation cutoff date for the special discounted ASA rates is 25 October 2007; after this date, the conference rate will no longer be available. You must mention the Acoustical Society of America when making your reservations to obtain the special ASA meeting rates. Please make your reservations directly with the hotel and ask for one of the rooms being held for the Acoustical Society of America (ASA). Alternatively, reservations can be made directly online at ⟨http://www.starwoodmeeting.com/StarGroupsWeb/res?id=0705236872&key=37743⟩. This site has been set up specifically for the Acoustical Society of America, and has the conference rates and all applicable information incorporated into it.

The Sheraton New Orleans Hotel is located on historic Canal Street, overlooking the Mississippi River and the French Quarter and is a short walk from Bourbon Street, the Aquarium of the Americas, IMAX Theater, Riverwalk Marketplace, all the world-famous restaurants and live music clubs of the Vieux Carré.

The hotel features a pool and fitness center. All rooms are equipped with coffee makers with complimentary coffee, color cable TV with in-room movies and video games (fee), Starwood Turbo Net High Speed Internet Access (fee), in-room safes, hair dryers, irons and ironing boards, telephone voice mail, video checkout, room service and dual-line telephones.

Sheraton New Orleans Hotel 500 Canal Street
New Orleans, LA 70130
Tel.: 504-525-2500; Toll Free: 1-888-627-7033
FAX: 504-595-5552
Online: http://www.starwoodmeeting.com/StarGroupsWeb/res?id=0705236872&key=37743
Rates (excluding taxes)
Single/Double: $159.00 USD
Club Level: $189.00 USD
Taxes: 13% + $3.00 USD occupancy tax

Room Sharing

ASA will compile a list of those who wish to share a hotel room and its cost. To be listed, send your name, telephone number, e-mail address, gender, smoker or nonsmoker preference, not later than 15 October to the Acoustical Society of America, preferably by e-mail: asa@aip.org or by postal mail to Acoustical Society of America, Attn.: Room Sharing, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. The responsibility for completing any arrangements for room sharing rests solely with the participating individuals.

Weather

New Orleans has a subtropical climate with pleasant year-round temperatures. Rainfall is common in New Orleans, with a monthly average of about five inches of precipitation. Carrying a small foldable umbrella may be useful for showers. Average temperatures are between 70 deg F and 50 deg F in November with average rainfall of 4.1 inches.

New Orleans City Information

For the latest information about the city of New Orleans, please visit the New Orleans Convention and Visitors Bureau website at http://www.neworleanscvb.com/.

Assistive Listening Devices

Anyone planning to attend the meeting who will require the use of an assistive listening device is requested to advise the Society in advance of the meeting: Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, asa@aip.org.

Child Care During Meetings

Information concerning child care will be added to the meeting information on the ASA web site when details become available.

Accompanying Persons Program

Spouses and other visitors are welcome at the New Orleans meeting. The registration fee for accompanying persons is $50.00 for preregistration by 29 October and $75.00 at the meeting. A hospitality room for accompanying persons will be open at the Sheraton New Orleans Hotel from 8:00 a.m. to 11:00 a.m., Tuesday through Friday. To "Pass a Good Time" in "New Awlins," programs including lectures and tour information are being planned for the ASA members, spouses, and other visitors accompanying them.

New Orleans is "Food City," so the first speaker scheduled for Tuesday at 9:00 a.m. is restaurant critic Tom Fitzmorris. He's a well known popular host of a daily radio show—The Food Show—and author of 4 cookbooks. Check out his website, ⟨http://www.nomenu.com/tfbio.html⟩. He'll be bringing his latest cookbook and will be available for questions after his talk. Notably, more restaurants are open now than there were before Hurricane Katrina!

Because there is so much to see and do in and around the Crescent City, on Wednesday at 9:00 a.m., Ann Leonard, a free lance tour guide and elder hostel lecturer, will present a slide show and talk entitled "It's a New Orleans Thing," a fun talk about the city and its idiocyncracies.

Ann will be able to tell you about the different tours available in the city. Her knowledge of the industry will give you the insight necessary to pick and choose what you want to do and where you want to go in the time available to you. For the tourist, New Orleans has recovered mightily from Hurricane Katrina. It caused widespread flooding when some of the levees collapsed from the high wind-driven water. The uptown and downtown areas, scenic St. Charles Avenue, the parks, universities and French Quarter are completely restored, so you will not see the devastating effects of the storm unless you take a guided tour of the neighborhoods that have been unable to rebuild.

There will be maps and tour information at the Hospitality Room desk where greeters will be able to assist you.

Founded in 1718, New Orleans is truly a place apart, settled by the French and Spanish, and remaining so until after the Louisiana Purchase, when the "New Americans" moved south. It has its own food, its own culture, and its own rhythm (read music)—all a potpourri of influences mixed into one gumbo pot and emerging with its unique flavor. The ASA will offer a lecture on the history of the area.

Also rich in history is the Mardi Gras—not just a day, but a season, and for many a way of life. "Laissez Les Bon Temp Roule," (Let the Good Times Roll). You'll enjoy hearing about its traditions and how it influences the character of the city.

Please check the ASA website at ⟨http://asa.aip.org/meetings.html⟩ for updates about the accompanying persons program.

Registration Information

The registration desk at the meeting will open on Tuesday, 27 November, at the Sheraton Hotel. To register use the form in the call for papers or register online at ⟨http://asa.aip.org⟩. **If your registration is not received at the ASA headquarters by 29 October you must register on site.**

Registration fees are as follows:

Category	Preregistration by 29 October	Onsite Registration
Acoustical Society Members	$350	$425
Acoustical Society Members One-Day Attendance*	$175	$215
Nonmembers	$400	$475
Nonmembers One-Day Attendance*	$200	$240
Nonmember Invited Speakers One-Day Attendance*	Fee waived	Fee waived
Nonmember Invited Speakers (Includes one-year ASA membership upon completion of an application)	$110	$110
ASA Early Career Associate or Full Members (For ASA members who transferred from ASA student member status in 2005, 2006, or 2007)	$175	$215
ASA Student Members (with current ID cards)	Fee waived	$25
Nonmember Students (with current ID cards)	$45	$55
Emeritus members of ASA (Emeritus status pre-approved by ASA)	$50	$75
Accompanying Persons (Spouses and other registrants who will not participate in the technical sessions)	$50	$75

*One-day registration is for participants who will attend the meeting for only one day. If you will be at the meeting for more than one day either presenting a paper and/or attending sessions, you must register and pay the full registration fee.

Nonmembers who simultaneously apply for Associate Membership in the Acoustical Society of America will be given a $50 discount off their dues payment for the first year (2008) of membership. Invited speakers who are members of the Acoustical Society of America are expected to pay the registration fee, but nonmember invited speakers may register for one-day only without charge. A nonmember invited speaker who pays the full-week registration fee will be given one free year of membership upon completion of an ASA application form.

Note: A $25 processing fee will be charged to those who wish to cancel their registration after 29 October.

ONLINE REGISTRATION

Online registration is available at ⟨asa.aip.org⟩.

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.

2007

5–8 Oct	123rd Audio Engineering Society Convention, New York, NY [Audio Engineering Society, 60 E. 42 St., Rm. 2520, New York, NY 10165-2520, Tel: 212-661-8528; Fax: 212-682-0477; Web: www.aes.org]
22–24 Oct	NOISE-CON 2007, Reno, NV [Institute of Noise Control Engineering, INCE Business Office, 210 Marston Hall, Ames, IA 50011-2153, Tel.: (515) 294-6142; Fax: (515) 294-3528; E-mail: ibo@inceusa.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

29 June–4 July	Acoustics08, Joint Meeting of the Acoustical Society of America (ASA), European Acoustical Association (EAA), and the Acoustical Society of France (SFA), Paris, France [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; Web: http://asa.aip.org/meetings.html].
27–30 Jul	NOISE-CON 2008, Dearborn, MI [Institute of Noise Control Engineering, INCE Business Office, 210 Marston Hall, Ames, IA 50011-2153, Tel.: (515) 294-6142; Fax: (515) 294-3528; E-mail: ibo@inceusa.org]
28 Jul–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. Email icben2008@att.net.]

Cumulative Indexes to the Journal of the Acoustical Society of America

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

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

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

ACOUSTICAL NEWS—INTERNATIONAL

Walter G. Mayer
Physics Department, Georgetown University, Washington, DC 20057

International Meetings Calendar

Below are announcements of meetings and conferences to be held abroad. Entries preceded by an * are new or updated listings.

August 2007

6–10 **16th International Congress of Phonetic Sciences (ICPhS2007)**, Saarbrücken, Germany (Web: www.icphs2007.de).

27–31 **Interspeech 2007**, Antwerp, Belgium (Web: www.interspeech2007.org).

28–31 **Inter-noise 2007**, Istanbul, Turkey (Web: www.internoise2007.org.tr).

September 2007

2–7 **19th International Congress on Acoustics (ICA2007)**, Madrid, Spain (SEA, Serrano 144, 28006 Madrid, Spain; Web: www.ica2007madrid.org).

9–12 **ICA Satellite Symposium on Musical Acoustics (ISMA2007)**, Barcelona, Spain (SEA, Serano 144, 28006 Madrid, Spain; Web: www.ica2007madrid.org).

9–12 **ICA Satellite Symposium on Room Acoustics (ISRA2007)**, Sevilla, Spain (Web: www.ica2007madrid.org).

10–13 **54th Open Seminar on Acoustics (OSA2007)**, Przemysl, Poland (Web: www.univ.rzeszow.pl/osa2007/).

17–19 **3rd International Symposium on Fan Noise**, Lyon, France (Web: www.fannoise.org).

18–19 **International Conference on Detection and Classification of Underwater Targets**, Edinburgh, UK (Web: ioa.org.uk).

19–21 **Autumn Meeting of the Acoustical Society of Japan**, Kofu, Japan (Acoustical Society of Japan, Nakaura 5th-Bldg., 2-18-20 Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan; Fax: +81 3 5256 1022; Web: www:asj.gr.jp/index-en.html).

20–22 **Wind Turbine Noise 2007**, Lyon, France (Web: www.windturbinenoise2007.org).

24–28 **XIX Session of the Russian Acoustical Society**, Nizhny Novgorod, Russia (Web: www.akin.ru).

27–29 *__3rd Congress of the Alps Adria Acoustical Association__, Graz, Austria (Web: www.alpsadriaacoustics.org).

October 2007

3–5 **Pacific Rim Underwater Acoustics Conference 2007**, Vancouver, BC, Canada (Web: PRUAC.apl.washington.edu).

9–12 **2007 Canadian Acoustic Conference**, Montréal, Québec, Canada (Web: caa-aca.ca).

17–18 *__Institute of Acoustic Autumn Conference 2007__, Oxford, UK (Web: www.ioa.org.uk/viewuocoming.asp).

25–26 **Autumn Meeting of the Swiss Acoustical Society**, Bern, Switzerland (Web: www.sga-ssa.ch).

November 2007

14–16 *__14th Mexsican International Congress on Acoustics__, Leon, Guanajuato, Mexico (Fax:+52 55 5523 4742; e-mail: sberusta@hotmail.com).

29–30 *__Reproduced Sound 23__, The Sage, Gateshead, UK (Web: www.ioa.orguk/viewupcoming.asp).

December 2007

6–9 *__International Symposium on Sonochemistry and Sonprocessing (ISSS2007)__, Kyoto, Japan (Web: www.j-sonochem.org/OSS2007).

June 2008

30–4 **Acoustics'08 Paris: 155th ASA Meeting + 5th Forum Acusticum (EAA) + 9th Congrès Français d'Acoustique (SFA)**, Paris, France (Web: www.acoustics08 -paris.org).—See note below—

July 2008

7–10 **18th International Symposium on Nonlinear Acoustics (ISNA18)**, Stockholm, Sweden (Web: www.conngrex.com/18th_isna).

28–1 **9th International Congress on Noise as a Public Health Problem**, Mashantucket, Pequot Tribal Nation (ICBEN9, P.O.Box 1609, Groton CT 06340-1609, USA: Web: www.icben.org).

August 2008

25–29 **10th International Conference on Music Perception and Cognition (ICMPC 10)**, Sapporo, Japan (Web: icmpc10.typepad.jp).

September 2008

22–26 **INTERSPEECH 2008 - 10th ICSLP**, Brisbane, Australia (Web: www.interspeech2008.org).

October 2008

21–24 *__Acústica 2008__, Coimbra, Portugal (Web: www.spacustica.pt).

26–29 **Inter-noise 2008** Shanghai, China (Web: www.internoise2008.org).

November 2008

2–5 **IEEEInternational Ultrasonics Symposium**, Beijing, China (Web: www.ieee-uffc.org/ulmain.asp?page=symposia).

September 2009

6–10 **Interspeech 2009**, Brighton, UK (Web: www.interspeech2009.org).

August 2010

23–27 **20th International Congress on Acoustics (ICA2010)**, Sydney, Australia (Web: www.ica2010sydney.org).

September 2010

26–30 *__Interspeech 2010__, Makuhari, Japan (Web: www.interspeech2010.org).

News from Paris

The latest news concerning the **Acoustics'08 Paris Conference** (29 June–4 July 2008) shows that the event combines the **155th ASA Meeting**, the **60th Anniversary Celebration** of the Société Française d'Acoustique, the **5th Forum Acusticum (EAA)**, the **9th Congrès Français d'Acoustique**, the **7th European Conference on Noise Control (Euronoise)**, and the **9th European Conference on Underwater Acoustics**.

The Mega-Event will take place at Palais des Congrès de Paris.

 0001-4966/2007/122(2)/687/1/$23.00

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.

Sound and Structural Vibration, 2nd Edition

Frank Fahy and Paolo Gardonio

Academic, New York, 2007. 633 pages, $95 (paperback) ISBN 10: 0-12-373633-1.

I have used the first edition of Frank Fahy's *Sound and Structural Vibration* for my course in Penn State's Graduate Program in Acoustics for several years. I chose the book partly because my predecessor, Dr. Courtney Burroughs, recommended it, but mostly because it strives to teach, using well written passages explaining basic structural acoustics. There are equations, but not too many of them, and Fahy doesn't get bogged down in lengthy mathematical derivations, referring to other books and papers. In the first edition, he often prodded the reader to fill in some of the missing details as exercises.

The first edition was useful for demonstrating fundamentally how structural vibrations interact with neighboring acoustic fields. Fahy explained how various waves propagate through flat and curved structures, and radiate sound. He also showed how sound is transmitted through structures, with a great chapter on transmission loss through various single and double leafed barriers. Finally, he discussed briefly how enclosed acoustic volumes couple with structural walls, and introduced numerical methods for solving structural-acoustic problems.

I found, however, that I needed to supplement the first edition with more detailed material of my own when teaching my students. I developed notes on how the forced vibration of structures is a series summation of a structure's modal responses; more detailed discussions of cylindrical shell vibration and sound radiation; and lengthy explanations of numerical methods in structural acoustics—finite element (FE) modeling of structures, boundary element (BE) modeling of acoustic spaces, and statistical energy analysis (SEA) of coupled fluid-structure systems.

I was very pleased, therefore, to find that all of these areas have been added to the second edition, along with a chapter on active control of sound radiation and transmission. The book has doubled in size—from 309 to 633 pages—but remains affordably priced (a key factor in choosing a book for college students). Little, if any, of the material in the first edition has been removed, but some of it has been rearranged. The reorganization is minor, and I find nearly all of it to be appropriate.

Since the original material has been reviewed already (see, for example: Rich and Peppin's review in *Shock and Vibration Digest*, 1986, Vol. 18, page 18, svd.sagepub.com; and Barry Gibbs' review in the *Journal of Sound and Vibration*, 1987, Vol. 117, No. 3, pages 604–605), I will comment mostly on the new material in the second edition. Some of the additions actually supplement the original material—new and useful plots of simulated and measured data are used to augment the first edition's text. Many of the new illustrations are now reproduced in color. Images of mode shapes and sound fields are clear and help illustrate the concepts embodied in the equations and text.

Chapter 1 - Waves in Fluids and Solid Structures has been expanded. The treatment of cylindrical shell vibrations and modes is more detailed, and new sections on structural modes and their influence on forced response have been added. Modal density and modal overlap, key parameters in Statistical Energy Analysis, are now explained. *Chapter 2* expands on forced response, beginning with simple lumped parameter modeling and continuing with modal summation approaches for beams, plates, and cylindrical shells. Mobility formulas (including those for moment mobility, which is rarely discussed in textbooks) are provided for finite and infinite structures. Finally, more quantities important to SEA—power and energy density—are introduced.

Continuing with the modal summation mobility formulations in Chapter 2, *Chapter 3 – Sound Radiation by Vibrating Structures* now includes modal summation approaches for computing sound radiation. Fahy's original wave number-based approach to explaining sound radiation remains from the first edition, complementing the new approaches. A general discussion of how to use integral (boundary element) techniques to calculate sound radiated by generally shaped structures has been added.

The authors expanded *Chapter 4 – Fluid Loading of Vibrating Structures* slightly to include fluid loading effects on cylindrical shell modes. The strong collection of formulas for transmission loss through various barriers in *Chapter 5* is now augmented with new information on how transmission loss is affected by damping treatments and multilayer composite plates, along with new material on the transmission loss of sound through pipes. *Chapters 6 and 7* have been updated somewhat. *Chapter 7* includes more on boundary element and SEA techniques–this time applied to interior sound fields.

The numerical analysis material in *Chapter 8* has been revised significantly. FE modeling approaches for beams and plates, and for acoustic cavities, are now included. The FE formulations are not meant to be exhaustive, but provide just enough context to the examples in the figures. The classic problem of the coupling between a flexible flat plate and a rigid walled enclosed cavity is solved using an FE/FE approach (panel and cavity modeled with finite elements). The related problem of the sound radiated by a flexible panel on a box is solved with FE/BE methods (panel modeled with finite elements and exterior acoustic space modeled with boundary elements). Throughout the chapter, well conceived color images of mode shapes and sound fields are included.

Chapter 9 is new—and introduces active control of sound radiation and transmission. Basic feedback loops and control algorithms are explained, and examples of the effects of control on actual systems are shown. This is also the only chapter that discusses practical measurements of sound and vibration, along with various actuators used to induce vibration—a welcome addition to the book.

While the book is quite long now (once again, it has doubled in size), and the authors have chosen to address a wide range of topics at a general level, there are still some topics which could have been expanded further. In particular, there is no discussion of thick beam or thick plate bending wave theory. At high frequencies and for thick structures, bending waves in beams and plates depend strongly on in-plane shear stiffness and rotary inertia. This fact has important ramifications on sound radiation, and thick beam/plate wave equations and wave speeds could have been included in this edition without their lengthy and painful derivations. Also, there is little discussion of the effects that heavy fluids (like water) have on structural vibrations. For information on heavy fluid loading effects, I will continue to refer my students to Junger and Feit's *Sound, Structures, and Their Interaction*.

The second edition of *Sound and Structural Vibration* is a well written, well annotated treatment of nearly everything structural-acoustic at a fundamental level. I recommend it strongly for teaching graduate students in vibration and acoustics, even those without extensive mathematical background. Short lists of problems are included at the end of each chapter, with answers provided in the back of the book. The reference list is broad, but not exhaustive, providing guidance to those who need more rigorous treatments of the various topics; and the book is well indexed. Also, with its new material, the second edition of this book makes a fine reference for the practicing structural-acoustician.

STEPHEN HAMBRIC
Applied Research Lab
Penn State University
State College, Pennsylvania
E-mail: sah19@psu.edu

 0001-4966/2007/122(2)/689/1/$23.00

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*
GEOFFREY EDELMANN, *Naval Research Laboratory, Code 7145, 4555 Overlook Ave. SW, Washington, DC 20375*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,178,408

43.28.Tc SODAR SOUNDING OF THE LOWER ATMOSPHERE

Andrew Louis Martin, assignor to Tele-IP Limited
20 February 2007 (Class 73/861.25); filed in Australia 26 February 2003

A sound direction and ranging system is proposed using a received (bistatic or monostatic) chirp to extract arrival amplitude and phase in order to estimate wind speed and bearing in the lower atmosphere. Such techniques have been used for a long time in underwater and medical applications. The patent claims techniques to overcome the practical issues of an atmospheric implementation.—GFE

7,180,827

43.30.Wi SURFACE ACOUSTIC ANTENNA FOR SUBMARINES

François Luc and Eric Sernit, assignors to Thales
20 February 2007 (Class 367/141); filed in France 15 February 2002

This patent proposes methods to overcome self-generated noise contaminating a submarine's flank arrays. Velocity sensors are combined with conventional pressure sensors in order to remove noise in a direction that is normal to the hull.—GFE

7,187,619

43.30.Wi METHOD AND APPARATUS FOR HIGH-FREQUENCY PASSIVE SONAR PERFORMANCE PREDICTION

Juan I. Arvelo, Jr. *et al.*, assignors to The Johns Hopkins University
6 March 2007 (Class 367/13); filed 11 March 2004

A dubious patent claim is made for a software package that predicts the performance of passive high-frequency sonar. The system uses location, time of year, noise, and weather as input to the well known CASS ray tracing model. This reviewer's performance prediction is that the merit of this patent will be undetectable.—GFE

7,177,233

43.30.Yj SYNTHETIC SONAR ANTENNA

Didier Billon, assignor to Thales
13 February 2007 (Class 367/88); filed in France 6 August 2002

In order to reduce the number of sensors in a synthetic aperture system, this patent proposes a self-calibration technique based on a spatial displacement along the array, a time delay, and a change in bearing between two correlated signals transmitted in sufficiently short time intervals.—GFE

7,180,828

43.30.Yj NON-KINKING OIL-FILLED ACOUSTIC SENSOR STAVE

Keith E. Sommer and Henry P. Stottmeister, assignors to The United States of America as represented by the Secretary of the Navy
20 February 2007 (Class 367/154); filed 22 April 2004

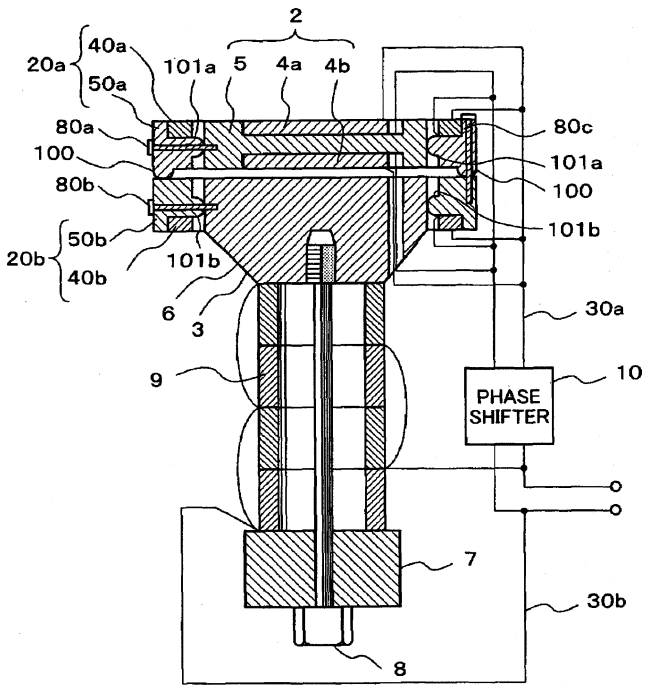
Manufacturing techniques are proposed that prevent commonly used oil-filled acoustic line arrays from kinking.—GFE

7,187,105

43.30.Yj TRANSDUCER WITH COUPLED VIBRATORS

Hiroshi Shiba, assignor to NEC Corporation
6 March 2007 (Class 310/325); filed in Japan 15 June 2004

A combination Langevin (axial mode) and bending (faceplate) mode transducer of wide bandwidth is claimed. Sound radiation is from faceplate **2**. Dimensions of the front mass **3**, rear mass **7**, faceplate **2** diameter, and



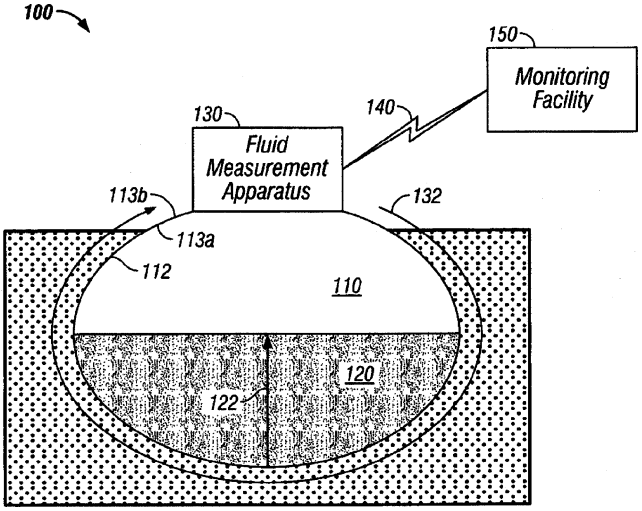
thicknesses **4-5** are selected, all bound by bolt **8**. The Langevin mode provides low-frequency sound radiation while faceplate bending modes provide high-frequency sound radiation.—AJC

7,039,530

43.35.Zc FLUID MEASUREMENT

John H. Bailey *et al.*, assignors to Ashcroft Incorporated
2 May 2006 (Class 702/50); filed 29 December 2003

A method is presented for the measurement of the fluid level in a container such as an underground tank or a railroad car used for the storage or transportation of gas, oil, water, etc. An ultrasonic vibration, typically in the range 30–150 KHz, is introduced by apparatus **130** into the container



wall at point **132**. The vibration travels around the container wall and is affected by the level of fluid in the tank. The resulting vibration is measured at the external tank surface **113b** and the signal travel time is measured to determine the level of fluid in the tank.—DLR

7,174,788

43.35.Zc METHODS AND APPARATUS FOR ROTARY MACHINERY INSPECTION

Gerald John Czerw and Laurie Diane Donovan, assignors to General Electric Company
13 February 2007 (Class 73/620); filed 15 December 2003

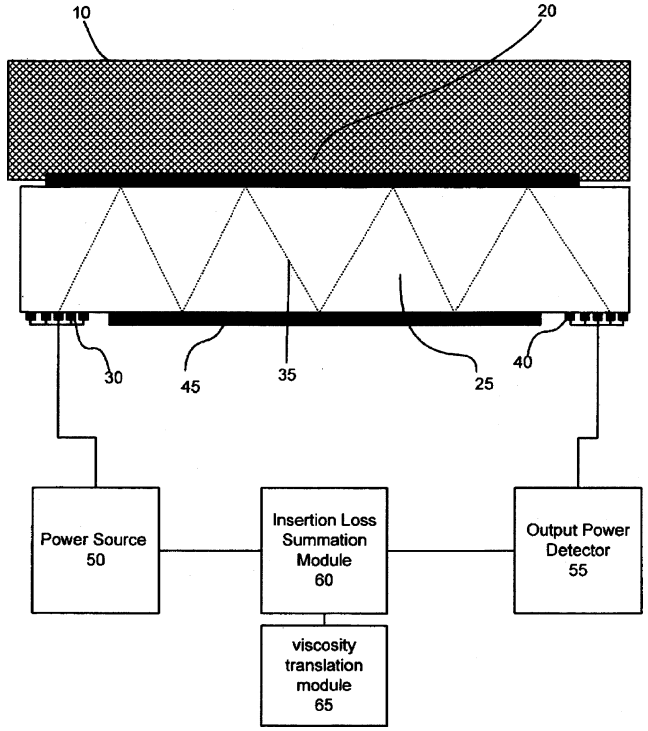
Removing turbine or compressor blades from the rotors for inspection is costly, but the roots of such blades are difficult to inspect when these blades are assembled in a rotor. This patent describes inspection systems in which a steered array of ultrasonic transducers placed near a blade root sends beams into the root at various angles and senses the reflections. A traversing assembly is used to move the transducer system along the blade chord.—EEU

7,181,957

43.35.Zc MEASUREMENT, COMPENSATION AND CONTROL OF EQUIVALENT SHEAR RATE IN ACOUSTIC WAVE SENSORS

Jeffrey C. Andle, assignor to Biode Incorporated
27 February 2007 (Class 73/54.41); filed 19 December 2005

A liquid- and viscoelastic-material viscosity meter is claimed where the shear rate (gradient) in that medium is measured. Shear waves **35** from transmitter **30** travel down a channel after multiple reflections from channel



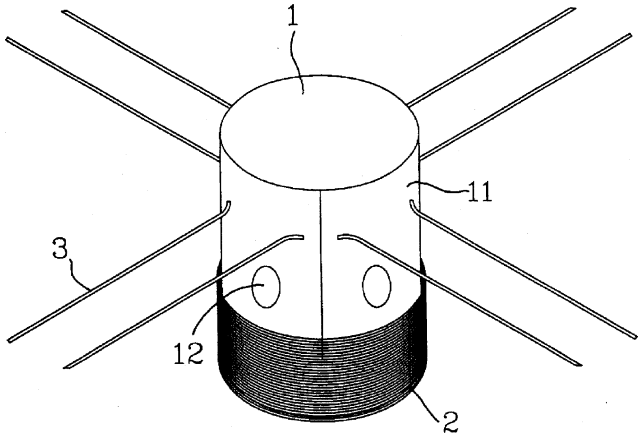
surfaces **45** to detector transducer **40**. The frequency and amplitude of the transmitted wave **35** are adjusted such that the amplitude of the received wave is suitably small. Frequency and amplitude values indicate viscosity.—AJC

7,146,020

43.38.Dv STRUCTURE FOR THE SOUND COIL OF LOUDSPEAKER

Jack Peng, assignor to Meiloon Industrial Company, Limited
5 December 2006 (Class 381/409); filed 20 November 2002

By using an oilpaper former **11** and either two of four independent layers of coils **2**, the ends of which are connected to ''thick wires'' **3**, as well as providing vents **12**, a voice coil that is smaller than a single coil can be produced that, according to the patent, is ''superior to conventional voice



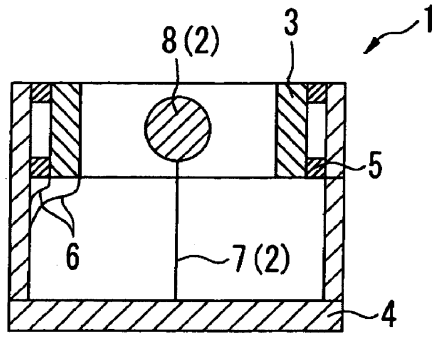
coils'' and is ''industrial valuable.'' This may be, but the added complexity of the invention may offset this superiority, and the very terse (but confusing) patent consisting of less than two and one-half columns of text does not discuss this.—NAS

7,176,601

43.38.Fx PIEZOELECTRIC POWER GENERATION SYSTEM AND SENSOR SYSTEM

Hidetoshi Tanaka and Norio Ohkubo, assignors to Hitachi, Limited
13 February 2007 (Class 310/339); filed in Japan 5 September 2003

This patent should be of interest to those concerned with energy harvesting and structural vibration. The authors describe a simple device (and



1:PIEZOELECTRIC POWER GENERATION SYSTEM
2:VIBRATOR
3:PIEZOELECTRIC ELEMENT
4:BASE
7:BEAM
8:IMPACT ELEMENT

its variations) that can be used to convert vibration energy of large amplitude (such as the motion of the human body, or rolling machinery) to electricity. In the figure, there is a kinetic energy object (a ball is pictured) that rattles around in a cage of piezoelectric material, whose electrodes are wired to a bridge rectifier. The concept is simple enough that it would work, although the efficiency is not mentioned. The patent is not detailed enough to warrant reading, so if you understand the figure, you understand the concept.—JAH

7,146,011

43.38.Hz STEERING OF DIRECTIONAL SOUND BEAMS

Jun Yang *et al.*, assignors to Nanyang Technological University
5 December 2006 (Class 381/77); filed 28 February 2004

A procedure is disclosed for steering a directional audio beam that is self-demodulated from an ultrasound carrier. An array of transducers made from lead zirconate titanate, or possibly another material, are fed signals that are processed using techniques proposed by Blackstock, Kamakura *et al.*, Kite *et al.*, and Pompei, as well as using zeroth-order Bessell functions to minimize spreading (as proposed by the Mayo Foundation) per the Durnin

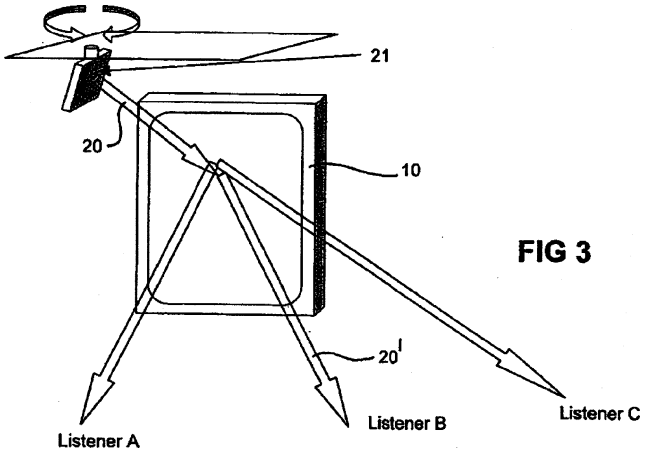


FIG 3

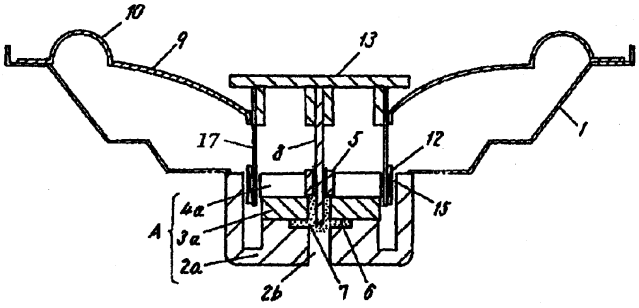
theory for designing non-diffracting beams. The device is said to have use in advertising in public spaces, providing multiple audio beams for same, as well as a means for physically rotating the transducer array to make use of nearby surfaces for reflecting the beam. The patent could also be classified under PACS category 43.60.Dh—NAS

7,149,323

43.38.Ja SPEAKER

Kiyoshi Yamagishi, assignor to Matsushita Electric Industrial Company, Limited
12 December 2006 (Class 381/415); filed in Japan 13 February 2001

The lowest resonant frequency of an electrodynamic cone-type transducer is controlled, in part, by the flexibility of the centering spider or damper. So, why not replace the damper with a centering rod? This has been



done, but the friction due to the center post **8** in bearing **5** causes resonances, among other deleterious effects, that are transmitted to the diaphragm **9**. This friction is reduced by using a magnetic fluid **7** which is confined to

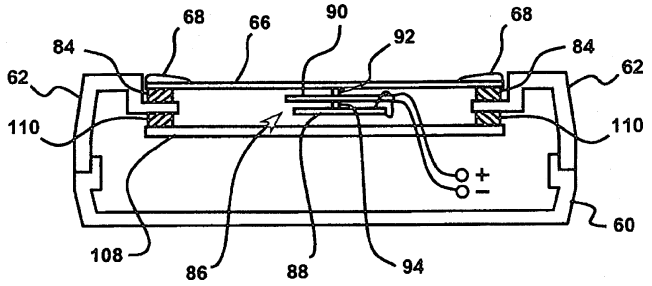
cavity **6** which can then fill the gap between the bearing and the rod. The invention describes the shape of the cavity into which the fluid is introduced.—NAS

7,151,837

43.38.Ja LOUDSPEAKER

Graham Bank *et al.*, assignors to New Transducers Limited
19 December 2006 (Class 381/190); filed in United Kingdom 1 August 2001

A cell phone display **66** that can support bending waves is excited via



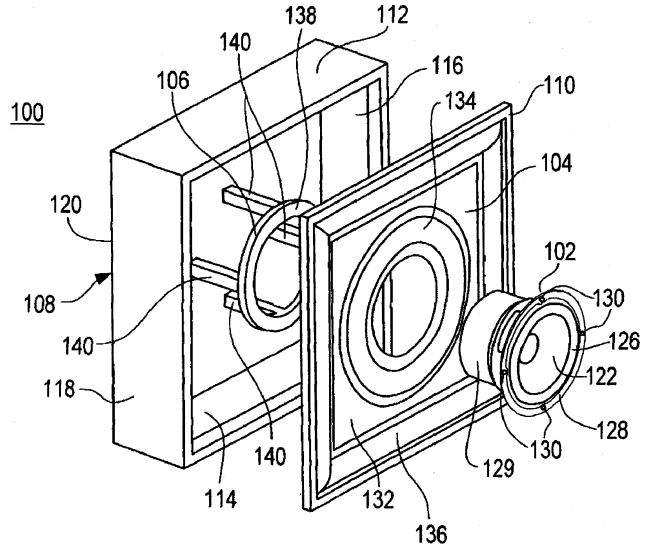
stub **94** by upper **90** and lower **88** transparent piezoelectric bimorph beams.—NAS

7,158,648

43.38.Ja LOUDSPEAKER SYSTEM WITH EXTENDED BASS RESPONSE

Aaron L. Butters and Sargam Patel, assignors to Harman International Industries, Incorporated
2 January 2007 (Class 381/160); filed 15 July 2003

In the eternal search for more from less, one area that has seen a lot of action since the dawn of modern sound reproduction is how to get more bass from a small enclosure. Adding a passive radiator was suggested by Olson in



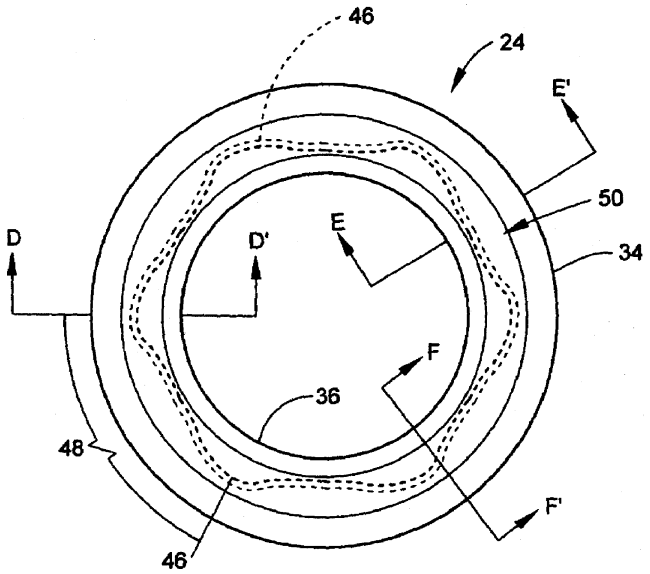
the 1930s and inventors have been incorporating this idea since then. In this instance, passive radiator **104** surrounds the active transducer **102**, which is mechanically supported by mechanism **106**. Note that passive driver surrounds **134** and **136** are not similar in shape.—NAS

7,174,990

43.38.Ja TANGENTIAL STRESS REDUCTION SYSTEM IN A LOUDSPEAKER SUSPENSION

Brendon Stead *et al.*, assignors to Harman International Industries, Incorporated
13 February 2007 (Class 181/172); filed 7 February 2005

This is a revision of United States Patent 6,851,513 [reviewed in J. Acoust. Soc. Am. **118**(2), 586 (2005)]. A major goal of the invention is to increase the linearity and excursion range of a loudspeaker suspension without increasing its size. A conventional half-roll suspension has substantial radial and tangential stresses at large cone excursions, as well as excessive damping losses. Loudspeaker designers have wrestled with the problem for more than 50 years. Some early high-fidelity loudspeakers used soft leather



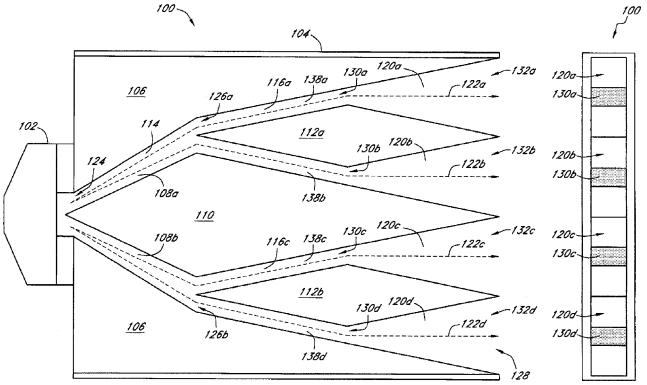
for the outer suspension. A Western Electric aluminum cone loudspeaker had discrete hinged sections separated by large gaps filled with a rubbery damping material. This patent describes an approach in which the cross-sectional shape of the suspension varies around its circumference. In the embodiment shown, the dotted lines indicate the high point of the roll. The undulating geometry allows the suspension elements to stretch more easily.—GLA

7,177,437

43.38.Ja MULTIPLE APERTURE DIFFRACTION DEVICE

Michael Adams, assignor to Duckworth Holding, LLC c/o OSC Audio Products, Incorporated
13 February 2007 (Class 381/340); filed 18 October 2002

Loudspeaker line arrays are currently in vogue for sound reinforcement and concert sound applications. However, it is difficult to create a true line source at frequencies above 2 kHz or so. A number of patents have been granted for various kinds of audio plumbing-convoluted waveguides that conduct sound from a conventional high-frequency driver to a vertical slot



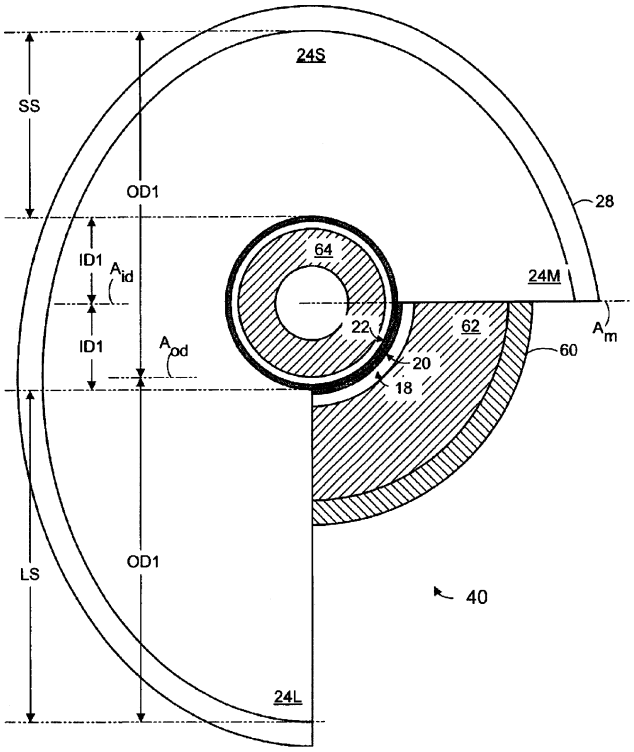
with equal path lengths. The ''tree'' geometry shown here appears to be both simple and effective.—GLA

7,177,440

43.38.Ja ELECTROMAGNETIC TRANSDUCER WITH ASYMMETRIC DIAPHRAGM

Patrick M. Turnmire *et al.*, assignors to Step Technologies Incorporated
13 February 2007 (Class 381/412); filed 31 December 2002

Audiophiles of a certain age will remember a planar loudspeaker whose diaphragm was shaped like an ear to suppress symmetrical breakup modes. It didn't work very well. One might apply the same idea to a more



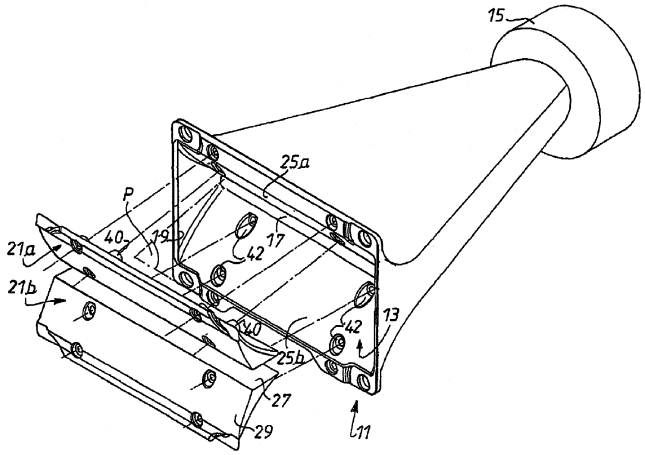
conventional loudspeaker by using an oval cone driven off-center, as shown. Earlier experimenters found that this doesn't work very well either. The patent contains a lot of theory but no test results.—GLA

7,178,629

43.38.Ja ELECTROACOUSTIC PUBLIC ADDRESS UNIT WITH ACOUSTIC HORN OR WAVEGUIDE

Eric Vincenot, assignor to NEXO
20 February 2007 (Class 181/191); filed in France 23 July 2001

A so-called diffraction horn expands in one plane while retaining constant width in the other. The horn mouth normally terminates in a slotted opening with a gentle arc or straight section. In many cases there is a terminating flare at the mouth that further modifies the radiation angle. This



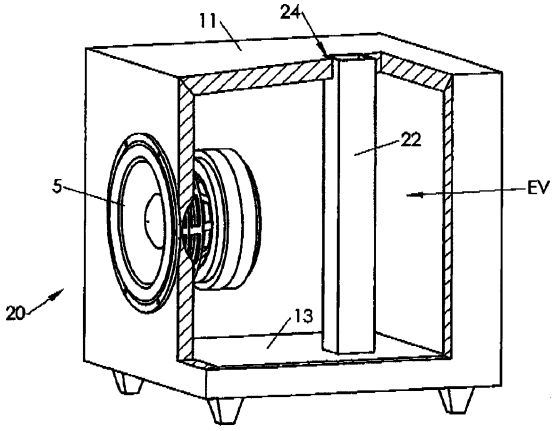
patent describes a horn design that enables the user to modify the terminating flare by attaching another section of different dimensions. It would be interesting to see typical polar plots showing the directional performance, with and without the added section.—JME

7,181,039

43.38.Ja THERMAL CHIMNEY EQUIPPED AUDIO SPEAKER CABINET

Enrique M. Stiles and Richard C. Calderwood, assignors to Step Technologies Incorporated
20 February 2007 (Class 381/397); filed 30 January 2004

There was a time when a ''high-power'' loudspeaker could handle 25 or 30 w without self-destructing. Today, a good woofer is expected to operate reliably with an input of several hundred watts. Since a low-frequency



loudspeaker is only a few percent efficient at best, the resulting heat buildup can be a serious problem, especially if the loudspeaker is mounted in a closed box. Various methods of heat dissipation have been proposed, includ-

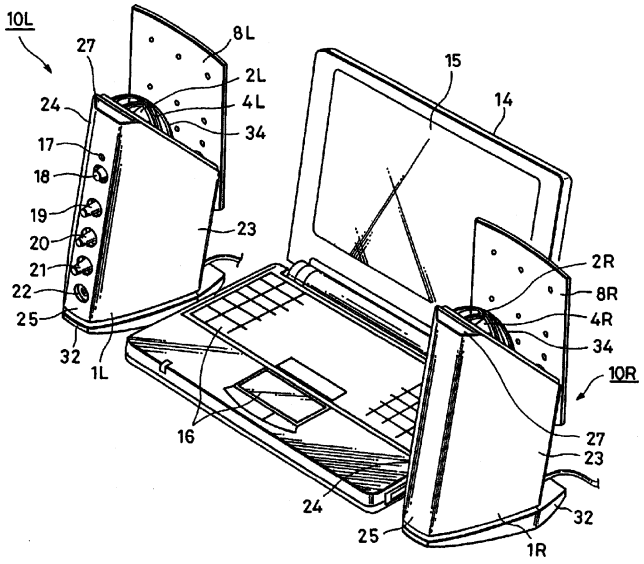
ing fans, finned plates, and water-cooled radiators. This patent suggests that heated air rising through metal chimney **22** can effectively transfer heat from the cabinet to external air.—GLA

7,184,562

43.38.Ja LOUDSPEAKER APPARATUS

Hideki Seki and Makoto Yamagishi, assignors to Sony Corporation
27 February 2007 (Class 381/160); filed in Japan 9 July 2001

Loudspeaker experimenters love to play with sound-reflecting panels. In some designs, sound is reflected from room walls. In others, reflectors are integrated into the overall cabinet design. One single-cabinet stereo system had hinged side panels that served as 45-degree reflectors when opened. The



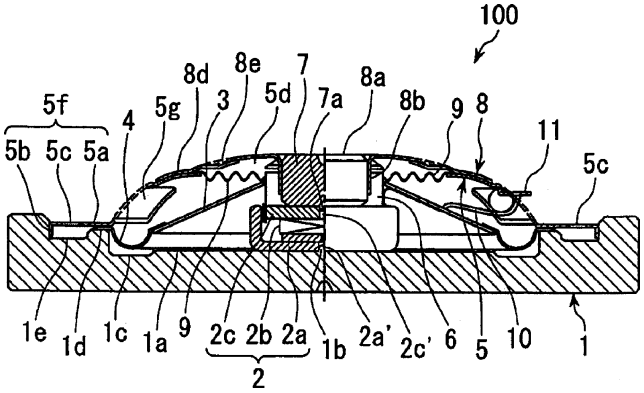
novel feature of this current rehash is that panels **8L**, **8R** ''can swing on a rear plate of the loudspeaker cabinet.'' Oh, the panels can also be illuminated by LEDs mounted on the back of the cabinet.—GLA

7,184,567

43.38.Ja THIN SPEAKER AND METHOD OF MANUFACTURING THE SPEAKER

Hiroshi Sugata and Hirohumi Onohara, assignors to Foster Electric Company, Limited
27 February 2007 (Class 381/433); filed in Japan 25 February 2002

To minimize depth, moving-coil loudspeakers are sometimes made



inside-out with the magnetic assembly nested inside the cone. The variant shown here is both inside-out and upside-down. The novelty of the design appears to lie in the assembly procedure, which is described at some length.—GLA

7,043,434

43.38.Md INTEGRATED SPEECH SYNTHESIZER WITH AN AUTOMATIC IDENTIFICATION OF SPEAKER CONNECTIONS AND IDENTIFICATION METHOD USED THEREOF

Wuu-Trong Shieh, assignor to Elan Microelectronics Corporation
9 May 2006 (Class 704/270); filed in Taiwan 20 December 2000

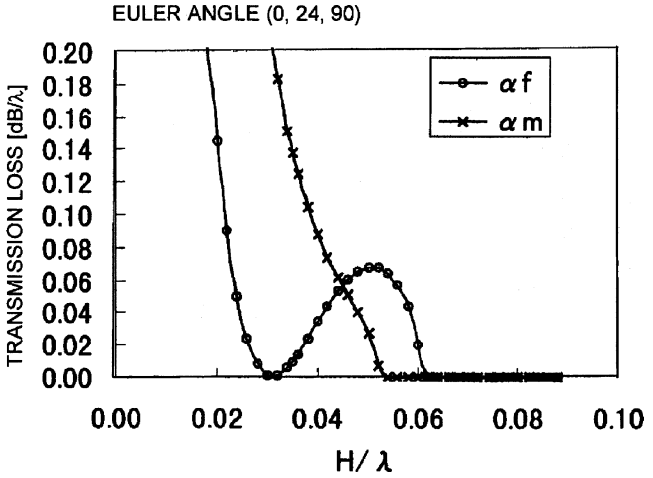
The term ''speech synthesis,'' as used here, in fact refers only to the process of digital-to-analog (D/A) conversion. A well-known problem in the field of D/A conversion of audio signals is the issue of the digital code used to represent the analog zero-voltage level. Between playbacks, the converter is typically driven to some specific ''no code'' value. If the digital ''no code'' value gets converted to a nonzero analog voltage, then some sort of click suppression system, typically an analog ramp generator, must be included in the circuitry. But a new problem arises when different D/A circuits, which may differ in the type of zero-level coding used, are to be interconnected to a common analog output. The patent describes a flag coding standard by which the click suppression circuit could be automatically switched into the D/A playback circuit when required.—DLR

7,180,222

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE

Hajime Kando *et al.*, assignors to Murata Manufacturing Company, Limited
20 February 2007 (Class 310/313 A); filed in Japan 12 August 2002

A surface acoustic wave (SAW) cell-phone, radio-frequency, bandpass filter operating in the second leaky wave mode with low propagation loss is claimed. Gold film electrodes confine the SAWs to the surface. A lithium niobate substrate cut on specified Euler angles and a silicon dioxide protec-



tive coating are used. The figure shows the propagation loss vs gold electrode thickness **H** for propagation for open circuit ***αf*** and closed circuit ***αm*** conditions. Second leaky wave propagation speeds are from 4663 m/s (Vm) to 5025 m/s (Vf).—AJC

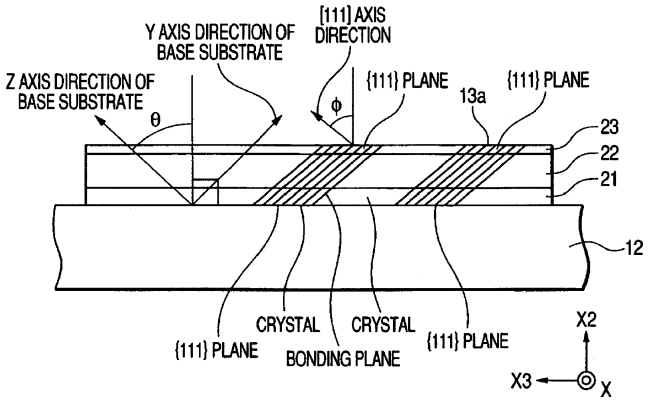
7,180,223

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE

Kyosuke Ozaki *et al.*, assignors to Alps Electric Company, Limited

20 February 2007 (Class 310/313 A); filed in Japan 6 February 2004

To achieve robust electrodes, low losses, and enhanced resonance Q,



laminated structure **21**-**23** has specified electrode crystal axis **13a** orientation with respect to the substrate crystal axes **12**.—AJC

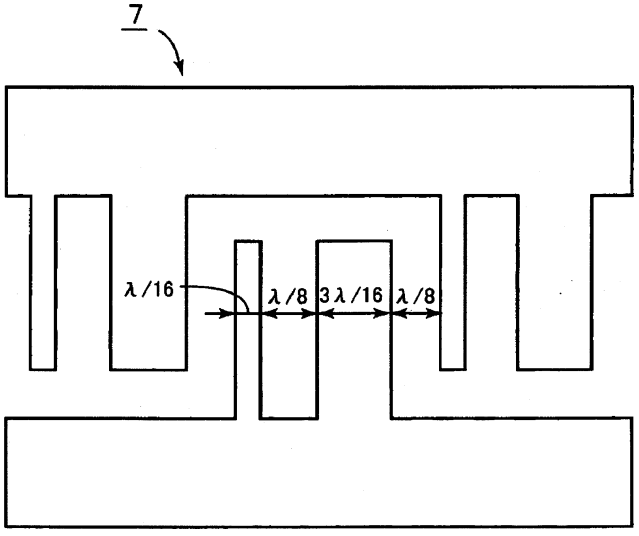
7,187,101

43.38.Rh SURFACE ACOUSTIC WAVE FILTER

Hideo Kidoh, assignor to Murata Manufacturing Company, Limited

6 March 2007 (Class 310/313 A); filed in Japan 4 February 2003

A compact SAW filter is claimed embodying a single phase unidirectional transducer (SPUDT) **7** with finger dimensions as shown in the figure. Electrode film thickness-to-wavelength ratio is in the range of 0.035–0.06. Finger material is an aluminum alloy and a layer that is not aluminum. The



piezoelectric substrate is cut at specified Euler angles. The velocity dispersion of this filter is negative, the temperature stability is good, and the bandwidth is less than 2 %.—AJC

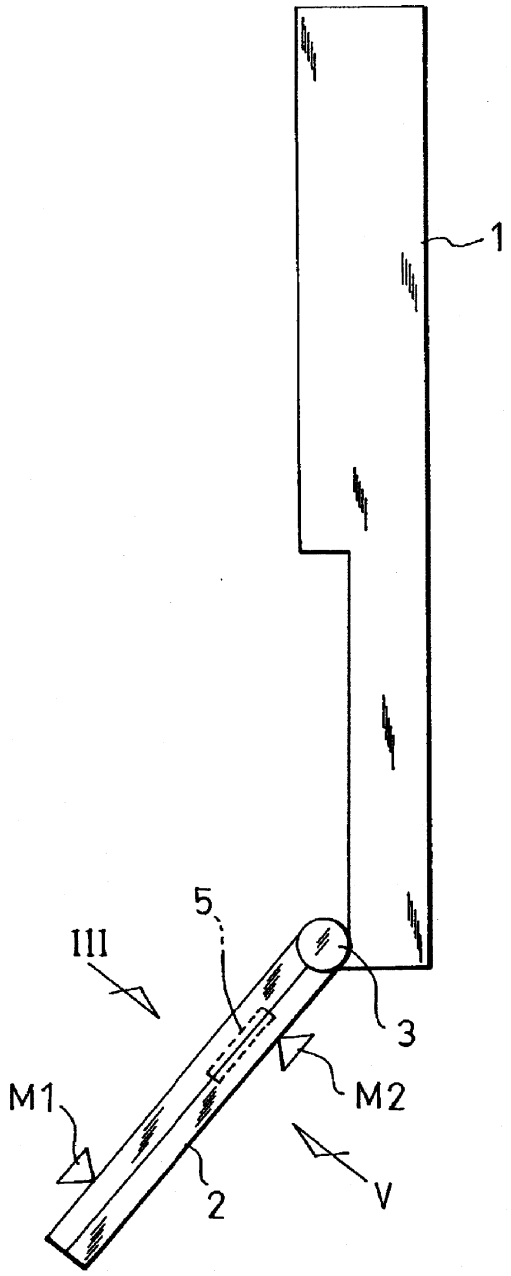
7,187,956

43.38.Si PORTABLE COMMUNICATION APPARATUS AND MICROPHONE DEVICE FOR THE APPARATUS

Shigeru Sugino *et al.*, assignors to Hosiden Corporation

6 March 2007 (Class 455/575.3); filed in Japan 28 September 2001

Ideally, the sound pickup properties of a cell phone should be the same whether the case is open or closed. The problem becomes even trickier if a close-talking, noise-canceling microphone is used. In this design an input aperture **M1** is located on the inner surface of the case and conducts sound to one side of microphone **5**. An ambient sound aperture **M2** is located on



the rear surface and conducts sound to the opposite side of the microphone diaphragm. When the case is closed, **M2** becomes the input aperture and **M1** becomes the ambient sound aperture. Exactly how ambient sound finds its way to **M1** when the case is closed is not explained.—GLA

7,042,990

43.38.Si METHOD FOR PARAMETRIZING THE GREETING MESSAGE OF A VOICE MAILBOX

Rodolphe Marsot, assignor to Cegetel Groupe
9 May 2006 (Class 379/88.23); filed 2 October 2003

In a typical cell phone system, a voice mailbox of some type is made available for the user to record a message to be presented to a caller when the user cannot answer the phone. Such a message might include a voice message plus one or more of a video clip, text, or a Touch Tone™ sequence. With a prior art cell phone system, such a message is typically produced by the process of calling in to the server and going through various trial runs until a satisfactory message is completed. During the time the message is being prepared, both the user's phone and the user's voice mailbox are unavailable to take incoming calls. This patent would provide all the necessary functions in the user's phone to prepare and evaluate the message locally, including trial playbacks and rerecording. Once a satisfactory message has been created, the user would then call into the server and upload the message, thus tying up the message system for only a short period of time. The message would be stored and transmitted using a format described here as multimedia messaging service (MMS). It is unknown to this reviewer whether MMS is related to any widely accepted standard.—DLR

7,184,786

43.38.Si TECHNIQUES FOR COMBINING VOICE WITH WIRELESS TEXT SHORT MESSAGE SERVICES

Inderpal Singh Mumick *et al.*, assignors to Kirusa, Incorporated
27 February 2007 (Class 455/466); filed 21 December 2004

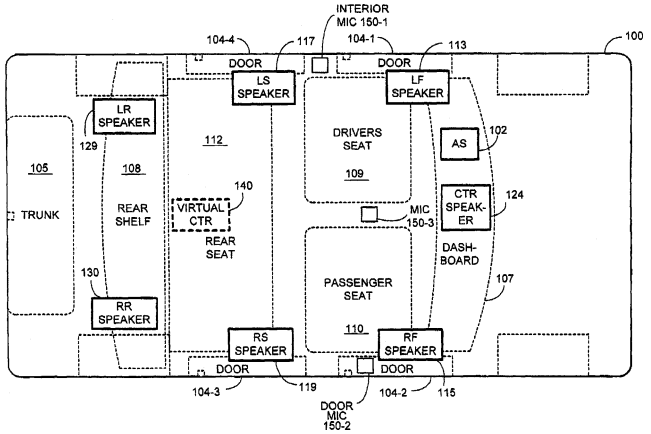
It's not enough that a caller can leave a voice recording or send a text message to your cellular phone. This patent describes a method to do both at the same time by including a short voice message along with a text message.—GFE

7,177,432

43.38.Vk SOUND PROCESSING SYSTEM WITH DEGRADED SIGNAL OPTIMIZATION

Bradley F. Eid and William Neal House, assignors to Harman International Industries, Incorporated
13 February 2007 (Class 381/22); filed 31 July 2002

Degraded optimization is an idea whose time has come. The seeming oxymoron is actually intended to be read as "optimization of degraded signals" and is concerned with automotive sound reproduction. The acoustics of the listening space, the listener/loudspeaker geometry, varying



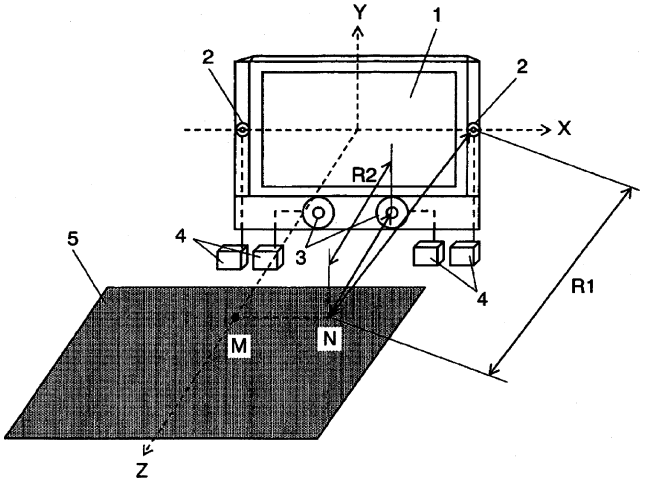
background noise, and less-than-optimum FM reception define a listening environment quite different from a home theater. The patent describes suitable techniques for decoding stereo signals, adding ambience to mono signals, and adjusting volume in relation to background noise.—GLA

7,181,029

43.38.Vk SPEAKER SYSTEM FOR PICTURE RECEIVER AND SPEAKER INSTALLING METHOD

Kazuhiko Ikeuchi *et al.*, assignors to Matsushita Electric Industrial Company, Limited
20 February 2007 (Class 381/306); filed in Japan 7 August 2003

The front panel of a TV receiver is largely taken up by the video display, leaving little room for left and right loudspeakers. Earlier patents have proposed complicated waveguides or arrays of small speakers to deal with the problem. Another common approach is to augment limited-range left and right speakers with larger woofers located under the screen. To maintain good stereo localization, such two-way systems use a low crossover frequency, perhaps 200 Hz, and try to position the left and right speakers close to their respective woofers. This patent asserts that a much higher



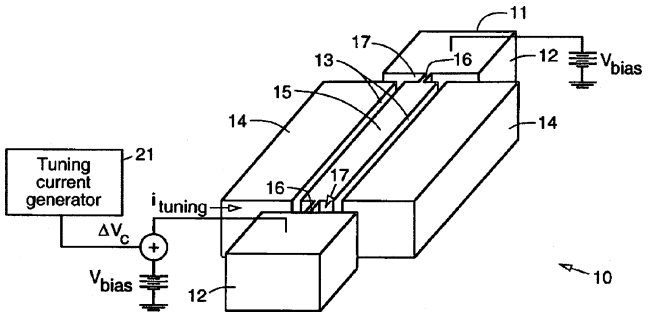
crossover frequency can be used, the left and right speakers **2** can be located higher, and woofers **3** can (and should) be located closer together, without compromising stereo imaging over a fairly large listening area. By tweaking crossover phase characteristics in relation to relative distances, listeners **M** and **N** presumably receive the same balance of low-frequency and high-frequency sound. Five magic formulas are presented to prove that it must be so, but this reviewer is skeptical. Moreover, other requirements for good stereo reproduction seem to have been ignored in the process.—GLA

7,176,770

43.40.Cw CAPACITIVE VERTICAL SILICON BULK ACOUSTIC RESONATOR

Farrokh Ayazi *et al.*, assignors to Georgia Tech Research Corporation
13 February 2007 (Class 333/186); filed 22 August 2005

This patent describes a MEMS-based mechanical resonator for use in tunable resonator applications operating in the VHF and UHF bands. These resonators are incorporated as movable plate pieces in variable capacitors, suggesting tuning capability and modulation capability as well. The primary function is to operate as a tunable bandpass filter, with tuning accomplished by a dc voltage in series with the input signal as shown in the figure. In



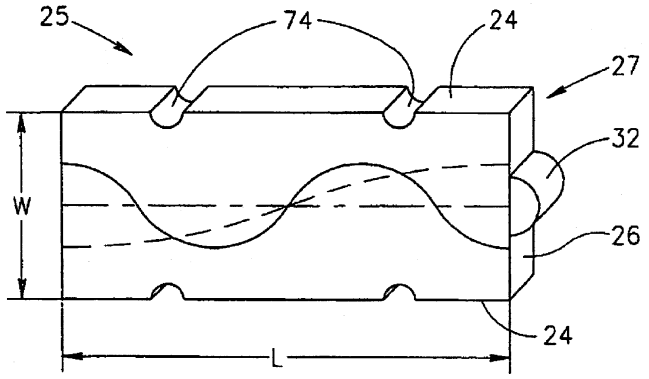
operation, the signal is fed into **14** (left) and taken out at **14** (right). A bias voltage is applied to bar **15**, allowing the gap capacitance to be used and driven by the rf signal, while a tuning voltage applied in series with the rf input allows the filter passband to be tuned. The concept is not novel, but the realization in this case is well thought out. Q and impedance specifications are given, and the authors say it has been operated up to 500 MHz.—JAH

7,183,690

43.40.Cw RESONANCE SHIFTING

Lior Shiv *et al.*, assignors to Nanomotion Limited
27 February 2007 (Class 310/312); filed 11 April 2005

This patent describes the use of holes and/or notches drilled into the sides of a resonant beam to affect the resonant frequencies of one or more modes of the beam. In this way, the phase shift between tangential and perpendicular motions can be controlled when the beam is used as the driving pawl of a linear motor. The softening of each mode of vibration is controlled by locating the holes at the appropriate points with respect to the resonance nodes and antinodes, allowing it to be tuned after production.—JAH

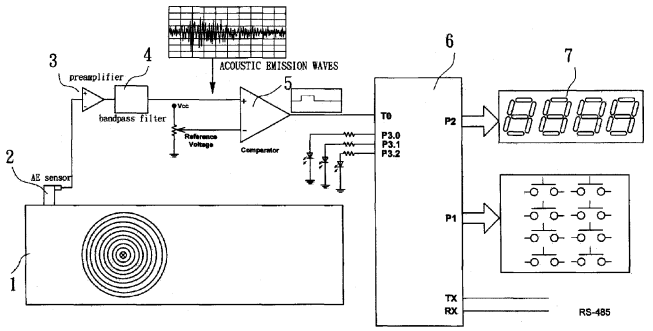


7,180,303

43.40.Le SIMPLE PARTIAL DISCHARGE DETECTOR FOR POWER EQUIPMENT USING ACOUSTIC EMISSION TECHNIQUE

Jiann-Fuh Chen *et al.*, assignors to Unelectra International Corporation
20 February 2007 (Class 324/536); filed 1 December 2005

The author claims an acoustic-emission, high-voltage, component voltage-breakdown sensor for transformers, switchgear, etc. Acoustic emission sensor **2** (essentially a piezoelectric accelerometer) attached to component **1** sends breakdown pulses to preamplifier **3**, amplifier **4**, bias circuit **5**



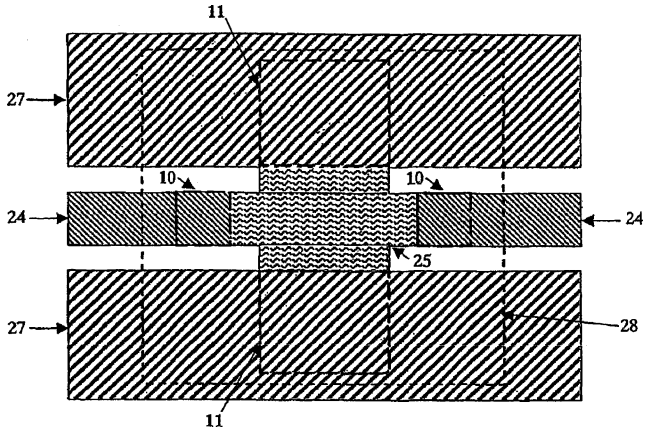
and signal processor **6** where pulses above a reference value are shaped into longer rectangular pulses for counter output **P2**, **7**. Multiple pulses per second indicate incipient voltage breakdown within **1**. Signals **P1** can also be sent for subsequent alerts and indicator lights to signal that maintenance is required.—AJC

7,187,254

43.40.Sk FILM BULK ACOUSTIC RESONATOR FILTERS WITH A COPLANAR WAVEGUIDE

Qingxin Su *et al.*, assignors to TDK Corporation
6 March 2007 (Class 333/189); filed in United Kingdom 29 November 2000

This patent discloses the use of piezoelectric film bulk acoustic resonators (FBARs) embedded in a coplanar waveguide as a means of implementing electrical bandpass filters. FBARs **25** are incorporated into coplanar waveguide formed by **24** and **27** by bridging the ground electrodes **27** and fitting in series with the signal electrode **24**. By this means, the authors assert that the filter can be made more compact than the usual wirebonded FBAR, and the ground plane characteristics can be more easily predicted. An example is given of how this replaces a discrete ladder filter network.—JAH



7,178,794

43.40.Tm FLUID ISOLATOR ASSEMBLY AND FLOATING ELASTOMERIC DAMPING ELEMENT

John Frederick Runyon, assignor to Seagate Technology LLC
20 February 2007 (Class 267/64.27); filed 10 September 2003

It is intended that several isolators of the type disclosed in this patent be used to support a platform on which sensitive equipment is located. An isolator according to this patent consists in essence of two isolators in series. The first isolator consists of a diaphragm atop a pressurized chamber and is intended to attenuate vibrations at relatively low frequencies; the chamber pressure may be adjusted to accommodate different payloads. The second isolator consists of two flexible disks that are joined at their edges by a

relatively thick annulus of viscoelastic material; it is intended to provide additional attenuation at higher frequencies.—EEU

7,178,818

43.40.Tm VIBRATION DAMPING DEVICE FOR USE IN AUTOMOTIVE SUSPENSION SYSTEM AND SUSPENSION SYSTEM USING THE SAME

Akira Katagiri *et al.*, assignors to Tokai Rubber Industries, Limited
20 February 2007 (Class 280/124.144); filed in Japan 10 March 2003

A device as described in this patent in essence consists of a rubber bushing that is located where a suspension member is connected to the vehicle body. A sensor that is included in the bushing's housing is used to sense relative motion between the outer and inner portions of the bushings, so as to provide a signal that may be used to activate an antilock brake system, for example.—EEU

7,182,187

43.40.Tm DAMPER AND VIBRATION DAMPING STRUCTURE USING THE SAME

Masami Mochimaru *et al.*, assignors to Oiles Corporation
27 February 2007 (Class 188/297); filed in Japan 21 February 2002

Dampers according to this patent are intended to be used in diagonal bracing for buildings to assist in the reduction of earthquake-induced vibrations. A typical damper consists of two relatively close-fitting concentric tubes with viscous or viscoelastic material between them. The outer tube may be connected near the top of a column, for example, while the inner tube is connected near the bottom of an adjacent column. Distortion of the building frame then results in relative motion between the tubes, thus inducing shear and attendant energy dissipation in the viscous or viscoelastic material.—EEU

7,174,879

43.40.Vn VIBRATION-BASED NVH CONTROL DURING IDLE OPERATION OF AN AUTOMOBILE POWERTRAIN

Michael Chol *et al.*, assignors to Ford Global Technologies, LLC
13 February 2007 (Class 123/406.21); filed 10 February 2006

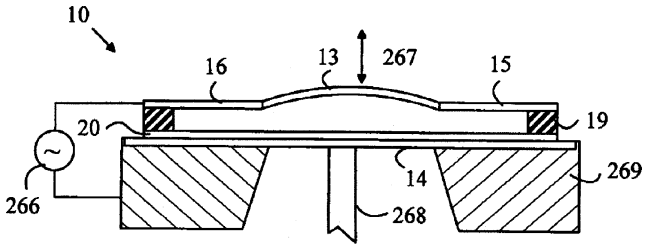
In order to reduce an automotive vehicle's vibrations when the engine is idling, the engine's spark timing is adjusted so as to reduce the idling speed in the event that a vibration signal from a sensor exceeds a predetermined threshold. Although the idling speed can be adjusted a priori, this system is said to be useful in cases where the idling conditions change with wear and varying dynamic conditions.—EEU

7,180,605

43.40.Yq VIBRATION SENSOR UTILIZING A FEEDBACK STABILIZED FABRY-PEROT FILTER

Dan Huber and Paul Corredoura, assignors to Agilent Technologies, Incorporated
20 February 2007 (Class 356/519); filed 28 September 2004

An acceleration transducer is claimed comprising Fabry-Perot mirrors **14** (fixed) and **13** mounted on cantilever springs **16**. Mirror **13** is subject to displacement by acceleration force **267**. A countering electrostatic charge **266** is also applied via the control circuit. This electrostatic force maintains the mirror position. Detection of mirror error displacement is via cavity light



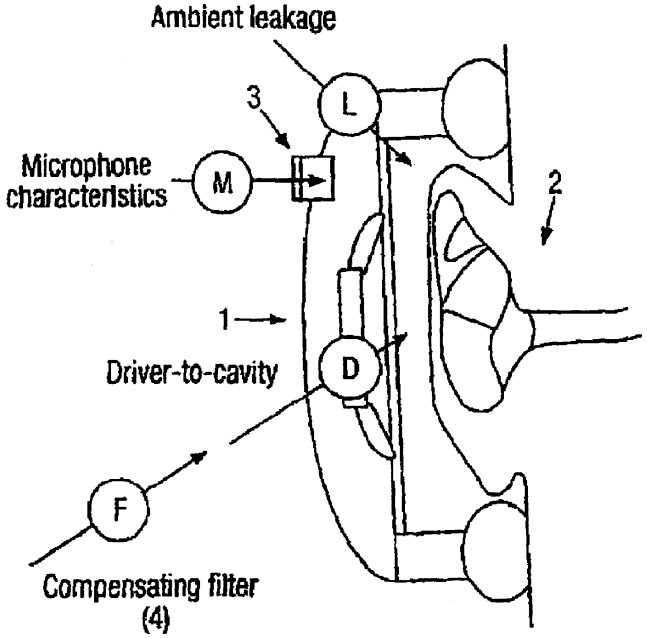
at two optical frequencies from electrically pumped laser-medium transmitter **20** by optical fiber **268** to a light sensor. The amplitude of Fabry-Perot light transmitted to the fiber is a function of the spacing between mirrors **13** and **14**. A control circuit changes the electrostatic force voltage to maintain mirror **13** at its static location. That control voltage then becomes the acceleration output signal of this vibration sensor.—AJC

7,177,433

43.50.Ki METHOD OF IMPROVING THE AUDIBILITY OF SOUND FROM A LOUDSPEAKER LOCATED CLOSE TO AN EAR

Alastair Sibbald, assignor to Creative Technology Limited
13 February 2007 (Class 381/71.6); filed in United Kingdom 7 March 2000

Lightweight noise-cancelling headsets are surprisingly effective in removing annoying background noise. These typically include a sensing microphone in the outer surface of each ear cup. The electrical signal from the microphone is amplified, lowpass filtered, and then subtracted from the pro



gram signal. High frequencies are passively attenuated by the ear cup itself. This patent asserts that more effective noise cancellation can be achieved by including higher frequencies and electrically compensating for head-related effects, including individual ear geometries.—GLA

7,182,994

43.55.Dt ACOUSTIC FLOOR MAT

Cooksey Timothy Scott, assignor to Pretty Products, Incorporated
27 February 2007 (Class 428/131); filed 8 January 2003

This proposed floor mat is designed to keep water out, but let sound through. It is envisioned for automobiles or other acoustic spaces where

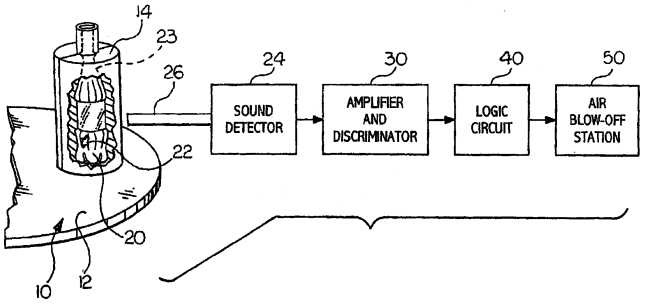
sound quality might be diminished by undermining the original acoustic design. This is not a psychological acoustics patent.—GFE

7,040,167

43.58.Fm METHOD AND APPARATUS FOR DETECTING HOLES IN PLASTIC CONTAINERS

Donald W. Hayward *et al.*, assignors to Plastic Technologies, Incorporated
9 May 2006 (Class 73/592); filed 31 October 2002

A defect monitoring system is described for use in the production of plastic containers **14**, which are to be tested as airtight. In order to detect any air leaks in a container, compressed air is injected into the container and an acoustic pickup **26** detects any air leaks. Described as an ultrasonic system, the operating frequency is said to be in the range of 5–40 KHz. In setting up



and testing the leak detection system, the best results were obtained by using a polyvinyl chloride tube of 1 in. internal diameter and a length of 12.9 in., placed on the pickup of a Radio Shack sound level meter. The claims include specification of the tube, but not the meter.—DLR

7,183,699

43.58.Wc STACKED TYPE ELECTRO-MECHANICAL ENERGY CONVERSION ELEMENT AND VIBRATION WAVE DRIVING APPARATUS

Yutaka Maruyama and Kiyoshi Nitto, assignors to Canon Kabushi Kaisha
27 February 2007 (Class 310/365); filed in Japan 15 June 2004

This patent describes a rotating-wave motor that is based on stacking thin-film piezo elements that are segmented into quadrant electrodes, allowing a phase-shifted drive signal to be easily applied. The size is on the order of 1 cm, and it seems to be designed for camera focusing applications where low power consumption and low vibration are essential. There is no apparent novelty to this design, but the discussion and drawings are clear and detailed.—JAH

7,184,521

43.60.Bf METHOD AND SYSTEM FOR IDENTIFYING A PARTY ANSWERING A TELEPHONE CALL BASED ON SIMULTANEOUS ACTIVITY

Scott Edward Sikora *et al.*, assignors to Par3 Communications, Incorporated
27 February 2007 (Class 379/69); filed 10 June 2004

An automatic device is proposed that would detect whether a person or a computer is speaking on the phone. This patent asserts that automated devices put sound simultaneously on input and output audio channels. In other words, us polite humans don't speak while spoken too—the same can't be said about your answering machine.—GFE

7,177,416

43.60.Dh CHANNEL CONTROL AND POST FILTER FOR ACOUSTIC ECHO CANCELLATION

Ming Zhang and Kuo Yu Lin, assignors to ForteMedia, Incorporated
13 February 2007 (Class 379/387.01); filed 10 July 2002

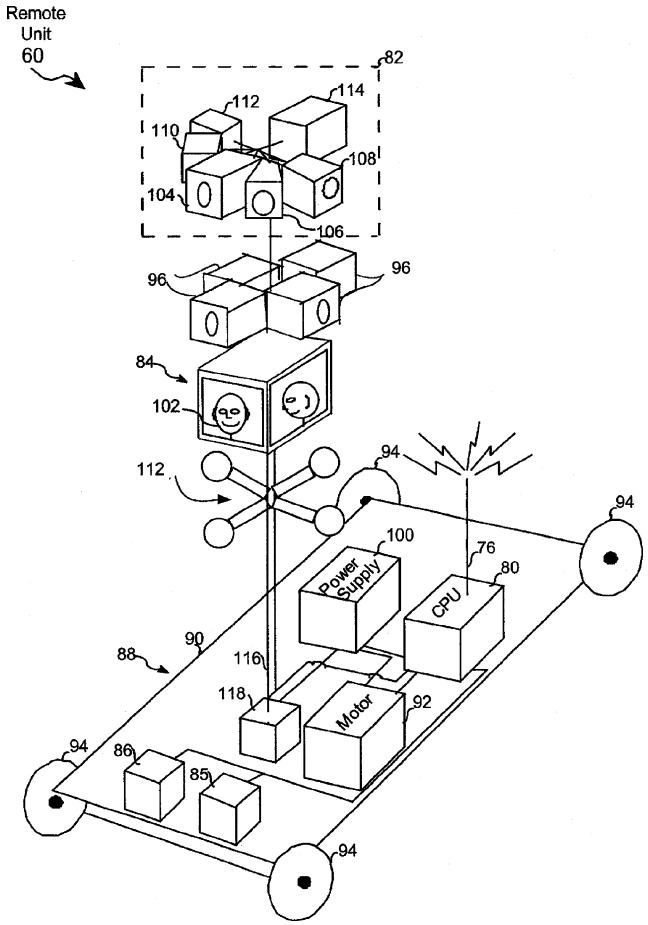
This patent describes a method to remove echoes from various hands-free or speakerphone devices that cannot be removed via coherent techniques such as adaptive filtering. The post filter presented calculates the cross-correlation between the digitized input signal and the residual adaptive filter error which then provides a set of coefficients to an FIR filter.—GFE

7,184,559

43.60.Dh SYSTEM AND METHOD FOR AUDIO TELEPRESENCE

Norman P. Jouppi, assignor to Hewlett-Packard Development Company, L.P.
27 February 2007 (Class 381/92); filed 23 February 2001

The patent describes an elaborate way of capturing the sonic signature of one space for eventual mapping into another space, hence the term ''tele-



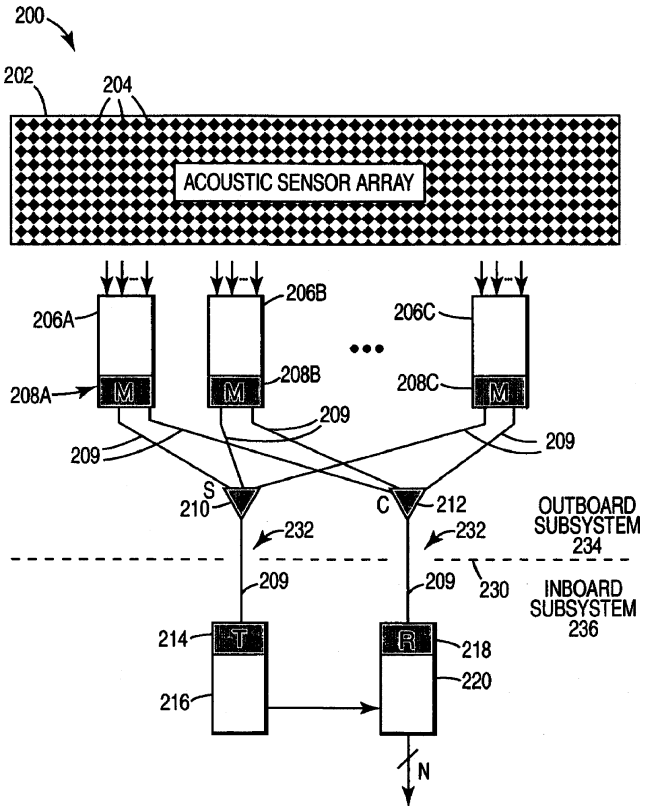
presence.'' The technique is intended as a supplement to standard video teleconferencing as a means of enhancing overall performance.—JME

7,184,670

43.60.Dh TELEMETRY SYSTEM AND METHOD FOR ACOUSTIC ARRAYS

VanWinkle (Van) T. Townsend, assignor to Lockheed Martin Corporation
27 February 2007 (Class 398/169); filed 2 May 2001

This patent deals with acoustic telemetry in the form of large arrays of



transducers such as may be used in underwater surveillance. Matters of sequential sampling, modulation, and multiplexing are discussed.—JME

7,187,623

43.60.Dh UNDERWATER DATA COMMUNICATION AND INSTRUMENT RELEASE MANAGEMENT SYSTEM

Maurice D. Green and Kenneth F. Scussel, assignors to Teledyne Benthos, Incorporated
6 March 2007 (Class 367/133); filed 10 March 2005

This patent describes a novel combination acoustic release and modem. This conveniently packaged device can transmit acoustic telemetry while the battery is strong before switching to a purely passive mode to await the release command.—GFE

7,187,718

43.60.Dh SYSTEM AND METHOD FOR ENCODING AND DECODING DIGITAL DATA USING ACOUSTICAL TONES

James Robert Jensen, assignor to Disney Enterprises, Incorporated
6 March 2007 (Class 375/260); filed 27 October 2003

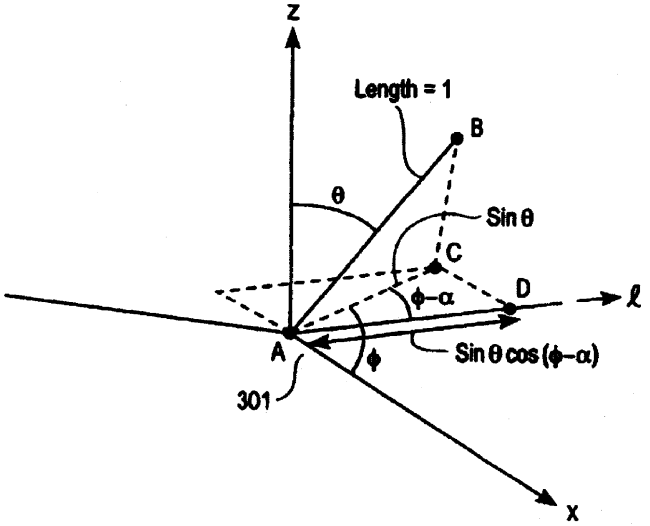
In this egregiously suspect patent, Disney claims to have invented the very concept of a digital acoustic modem and receiver. This claim includes any information transfer via acoustic tones. Doesn't anyone remember the ultrasonic television "clickers" from the 50s? Disney apparently wants a cheap way for already speaking toys to exchange digital information with a computer.—GFE

7,039,198

43.60.Fg ACOUSTIC SOURCE LOCALIZATION SYSTEM AND METHOD

Stanley T. Birchfield and Daniel K. Gillmor, assignors to Quindi
2 May 2006 (Class 381/92); filed 2 August 2001

This patent presents the argument that prior beamforming methods for source localization involved the processing of signals from one or more microphone pairs, identification of a possible source location for each mic pair, then some manner of combination of the possible solutions into a



single best fit. The method presented here essentially seems to be a matrix solution for simulataneously solving for the source location using all mic inputs. However, the math for this is not presented in a matrix formulation. Perhaps that is because the matrix solutions have been patented elsewhere.—DLR

7,039,199

43.60.Fg SYSTEM AND PROCESS FOR LOCATING A SPEAKER USING 360 DEGREE SOUND SOURCE LOCALIZATION

Yong Rui, assignor to Microsoft Corporation
2 May 2006 (Class 381/92); filed 26 August 2002

The acoustic source localization method presented here seems to be a traditional one, based on the calculation of a possible source location for each microphone pair. A direction of arrival is computed for each pair of microphones. These estimates are then combined for all mic pairs. What

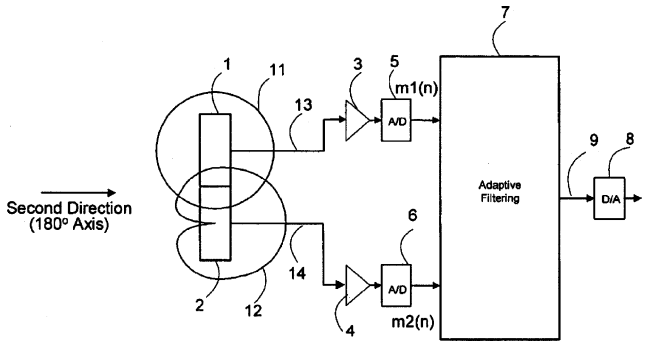
differs in the patented version is that temporal signal alignments are done only during the presence of a speech signal. Speech presence detection is based on energy levels observed within each processing frame.—DLR

7,181,026

43.60.Mn POST-PROCESSING SCHEME FOR ADAPTIVE DIRECTIONAL MICROPHONE SYSTEM WITH NOISE/INTERFERENCE SUPPRESSION

Ming Zhang, Singapore, Singapore *et al.*
20 February 2007 (Class 381/92); filed 13 August 2001

Omnidirectional and unidirectional microphones are located fairly close together and coupled via an adaptive filter arrangement, as shown in



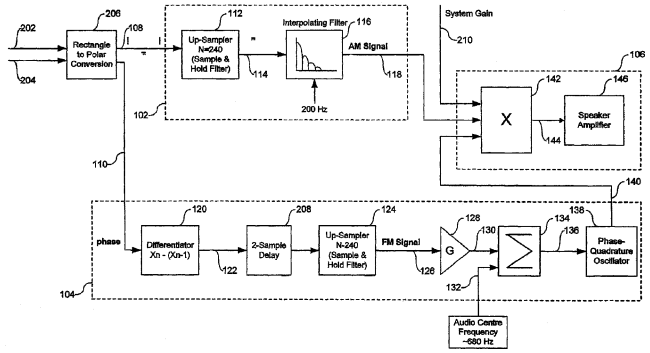
the figure. The effect of the adaptive filtering function, and further post-processing, is to null out a distant noise source through ''zeroing in'' on it.—JME

7,184,951

43.60.Qv METHODS AND SYSTEMS FOR GENERATING PHASE-DERIVATIVE SOUND

John Mark Royle *et al.*, assignors to Radiodetection Limited
27 February 2007 (Class 704/205); filed 15 February 2002

Underground object locator systems operate at alternating current frequencies **202** of several kilohertz. The resulting object signals **204** are narrow-band wave perturbations which, in themselves, are often indistinguishable and unrecognizable to the operator. The author claims an up-



converting signal process **112**, **124** for these narrow-band perturbations **204**, thus making them audible **146** if not distinguishable to the object locator operator.—AJC

7,181,955

43.60.Rw APPARATUS AND METHOD FOR MEASURING MULTI-PHASE FLOWS IN PULP AND PAPER INDUSTRY APPLICATIONS

Daniel L. Gysling, assignor to Weatherford/Lamb, Incorporated
27 February 2007 (Class 73/53.03); filed 7 August 2003

Presented is a paper-production patent that pertains to pulp parameters; specifically, the percentage of pulp to water, flow rate, and consistency. These parameters are measured via effective sound speed estimations.—GFE

7,187,777

43.60.Uv SOUND REPRODUCING SYSTEM SIMULATING

Richard E. Saffran, assignor to Bose Corporation
6 March 2007 (Class 381/306); filed 12 May 1995

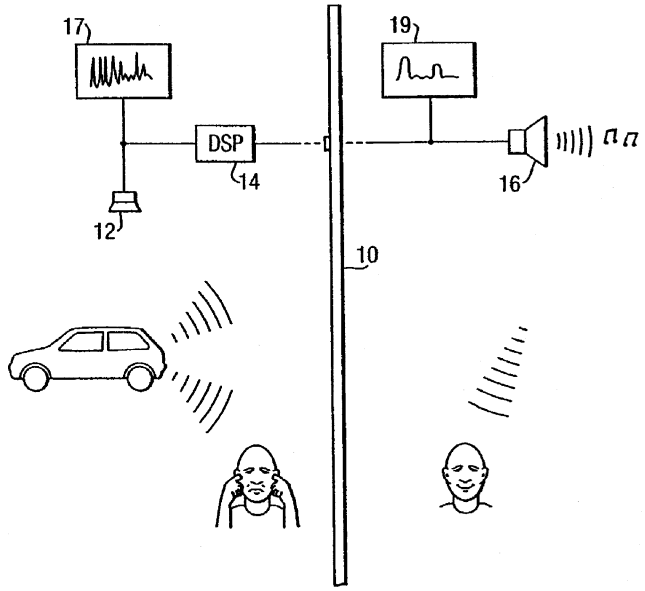
Perhaps the most difficult task faced by a potential buyer of a sound system is the objective evaluation of many different systems located in the same show room. This patent discusses one way in which the user can ''tare out'' these confusing differences. The solution is to place the listener in a small, acoustically inert booth and reproduce program material binaurally encoded over a pair of left-right loudspeakers with requisite crosstalk cancellation. Presumably, the various loudspeaker types to be tested will have been carefully modeled via impulse measurements so that only the differences among them will be apparent to the subject. The effects of various intended playback environments can similarly be auditioned.—JME

7,181,021

43.66.Ba APPARATUS FOR ACOUSTICALLY IMPROVING AN ENVIRONMENT

Andreas Raptopoulos, London, United Kingdom *et al.*
20 February 2007 (Class 381/71.14); filed in United Kingdom 21 September 2000

A device is described that listens for unwanted sounds (man-made and natural) that could be described as annoying or distracting. Pleasant masking



sounds, tailored to the user, are then transmitted in order to block out or change the perception of the noise.—GFE

7,181,019

43.66.Lj AUDIO CODING

Dirk Jeroen Breebaart and Arnoldus Werner Johannes Oomen, assignors to Koninklijke Philips Electronics N. V.
20 February 2007 (Class 381/23); filed in the European Patent Office 11 February 2003

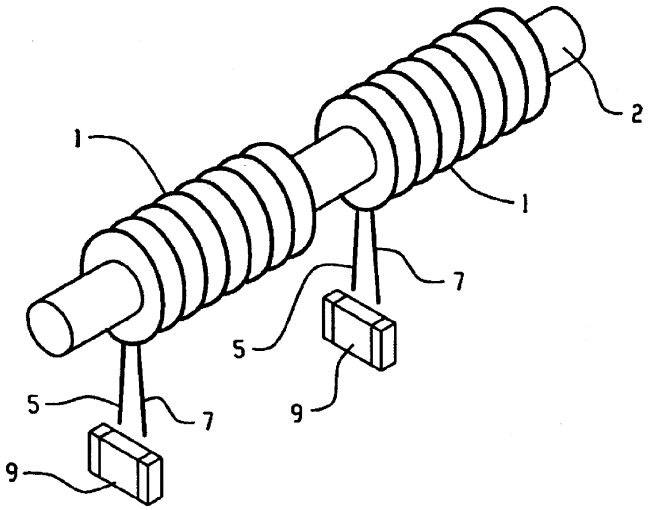
In many modern consumer applications, stereo programs may be transmitted via perceptual coding in order to save valuable signal space. Two-channel stereo lends itself to special treatment in terms of its sum (L+R) component and its difference (L−R) component. In the system described here, the sum, or monophonic, signal is transmitted directly. The difference signal is derived entirely from relative timing, phase, and amplitude cues between L and R program components on a near-instantaneous basis. These then act as ''steering components,'' enabling the sum signal to be redirected as stereo. With sufficient attention to fine tuning, such systems can work surprisingly well. If all of this sounds too good to be true, read up on color television, possibly the earliest example of perceptual coding.—JME

7,177,436

43.66.Ts COMPONENT ARRANGED DIRECTLY ON A T-COIL

Jan Frieding and Jürg Sudan, assignors to Phonak AG
13 February 2007 (Class 381/331); filed 26 March 2004

Additional capacitors required to bring coils to the desired resonant frequency in small, remote-controlled hearing aids are incorporated by attaching them directly to connection taps brought out on the coils.—DAP

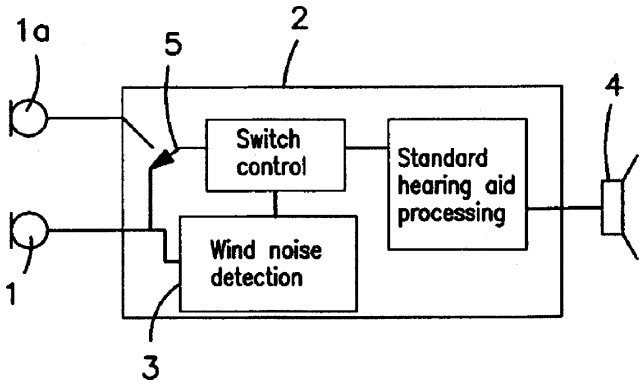


7,181,030

43.66.Ts WIND NOISE INSENSITIVE HEARING AID

Karsten Bo Rasmussen *et al.*, assignors to Oticon A/S
20 February 2007 (Class 381/312); filed in Denmark 12 January 2002

Two microphone types are utilized in a hearing aid: a primary microphone **1**, comparatively sensitive to wind noise, is utilized when there is little or no wind, and a secondary microphone **1a**, less sensitive to wind noise, is employed when there is significant wind present. Preferred embodiments include manufacturing the secondary microphone on a chip with



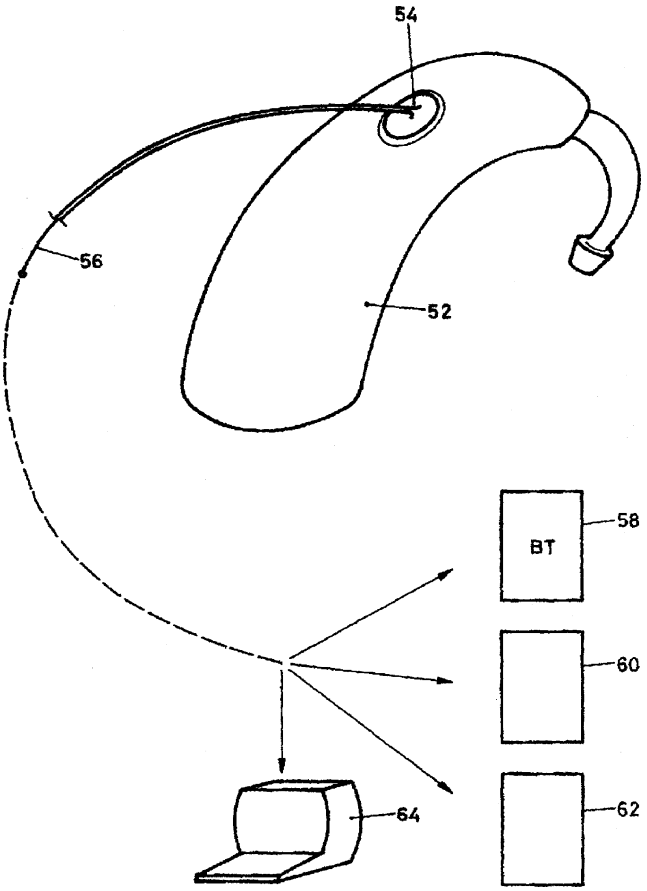
MEMS technology and utilizing a wind filter or wind-protected sound inlet location. The hearing aid signal processor may utilize cross correlation between the signals from the microphones to make decisions on when to switch between microphones.—DAP

7,181,032

43.66.Ts METHOD FOR ESTABLISHING A DETACHABLE MECHANICAL AND/OR ELECTRICAL CONNECTION

Andreas Jakob and Herbert Baechler, assignors to Phonak AG
20 February 2007 (Class 381/314); filed 13 March 2001

The hearing aid casing acts as the dielectric for a capacitor formed using conductive plates on either side of part of the case. On the outside of the case, the attractive force of a magnetic link is used to establish and maintain the mechanical linkage to a cable. The cable terminates in a detachable electrical connection to the case and connects to components inside the hearing aid, such as the signal processing module, via the hearing-aid-



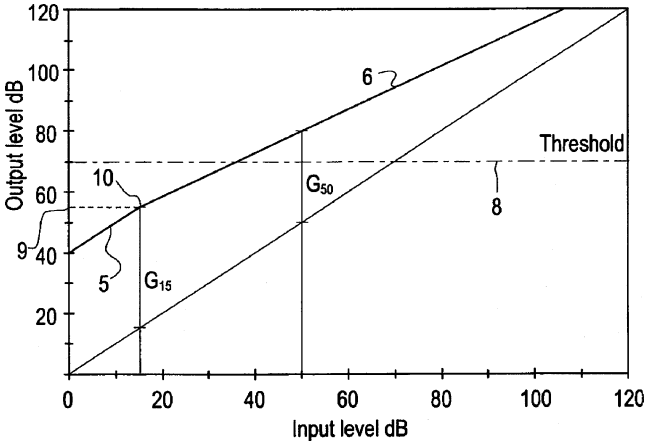
case capacitor. The other end of the detachable cable may connect to devices such as a radio for communication to external devices.—DAP

7,181,031

43.66.Ts METHOD OF PROCESSING A SOUND SIGNAL IN A HEARING AID

Carl Ludvigsen, assignor to Widex A/S
20 February 2007 (Class 381/312); filed 8 January 2004

A method of implementing compression in a hearing aid includes setting, in at least one channel, a kneepoint level that produces an output level below the hearing threshold. Below the kneepoint amplification is linear and above the kneepoint the compression ratio is typically 1.4:1-2:1. The attack



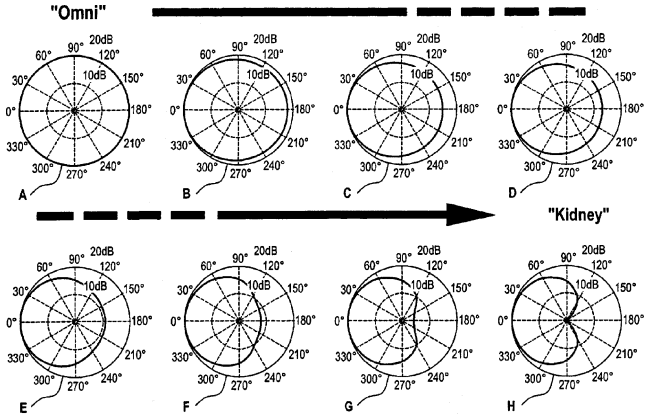
and release times of the compressor are typically 0.5–2 s and 5–20 s, respectively. The kneepoint is at a very low input level such as 15–25 dB SPL.—DAP

7,181,033

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

Eghart Fischer and Volkmar Hamacher, assignors to Siemens Audiologische Technik GmbH
20 February 2007 (Class 381/317); filed in Germany 17 October 2001

Switching within the hearing aid from one program to another or one signal processing algorithm to another is done gradually to avoid audio artifacts such as popping or level discontinuities. This soft transition is



achieved by having both operating conditions present in parallel during the switching event, with the relative weighting of the first to second signals gradually changing during switching.—DAP

7,181,034

43.66.Ts INTER-CHANNEL COMMUNICATION IN A MULTI-CHANNEL DIGITAL HEARING INSTRUMENT

Stephen W. Armstrong, assignor to Gennum Corporation
20 February 2007 (Class 381/321); filed 18 April 2002

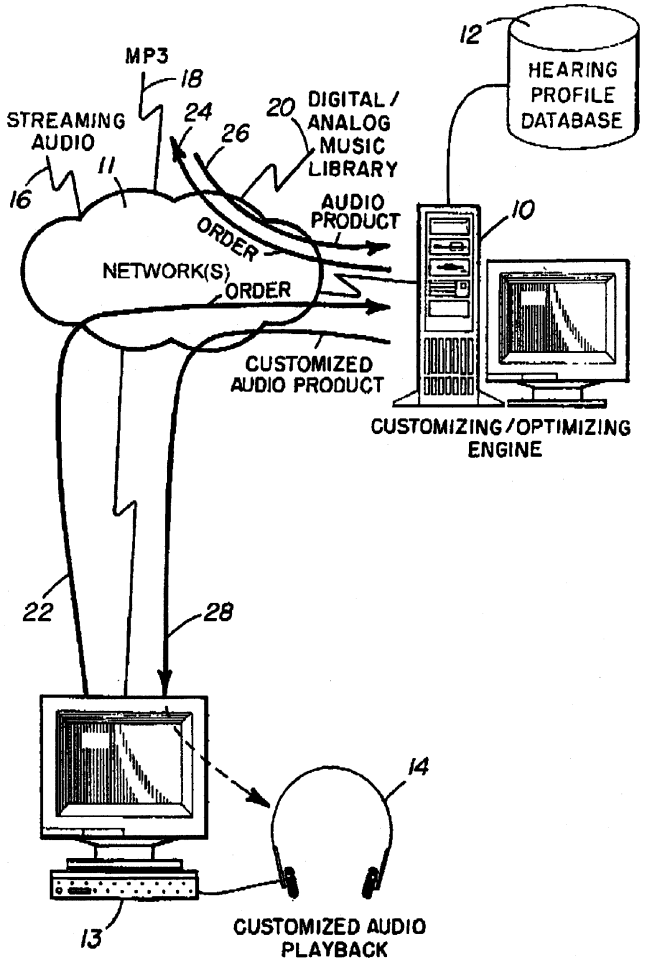
The amount of compression for each channel of a multichannel hearing aid is determined by the energy levels in that channel and at least one other channel as well as the total energy in the wideband audio signal. The one other channel signal may be at a higher frequency than the channel signal to help prevent masking weaker high-frequency signals.—DAP

7,181,297

43.66.Ts SYSTEM AND METHOD FOR DELIVERING CUSTOMIZED AUDIO DATA

Vincent Pluvinage and Rodney Perkins, assignors to Sound ID
20 February 2007 (Class 700/94); filed 28 September 1999

Steps cited in the method described include storing a machine-readable hearing profile for the user, accepting a machine readable order from the user for a particular audio data product via an input device or network,



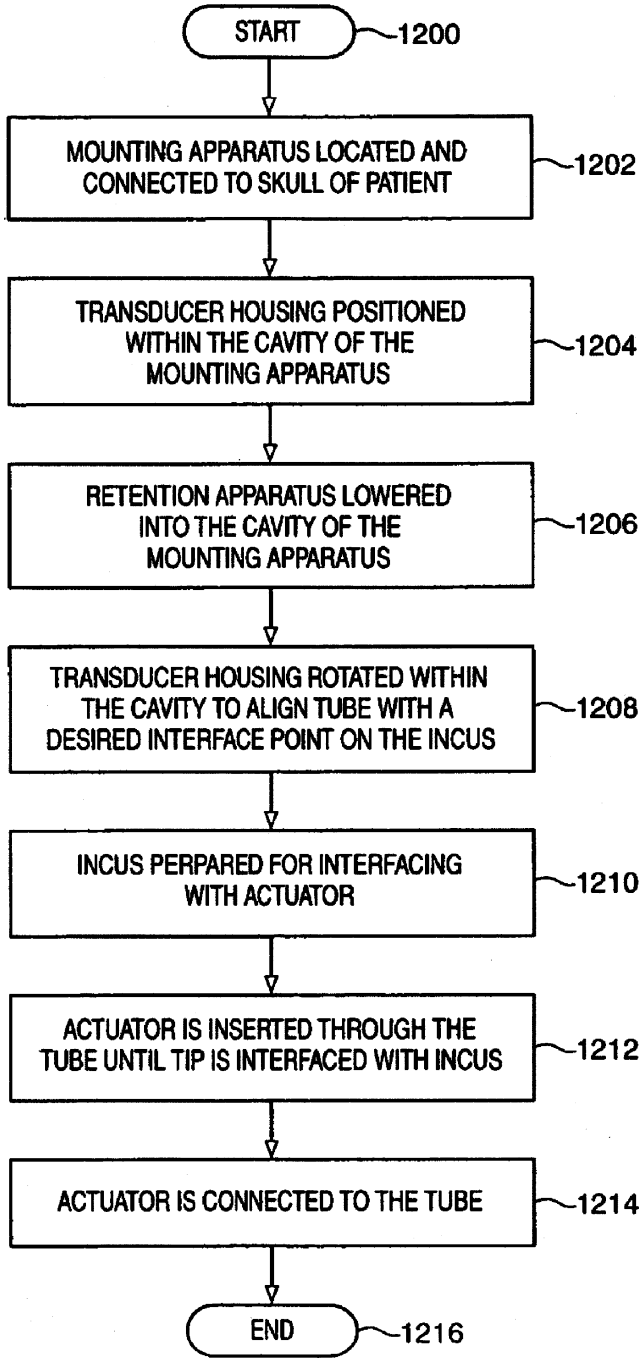
customizing the audio data in accordance with the hearing profile of the user, delivering the audio data product via machine readable storage medium or a network, and using feedback about the performance of the product from the user to modify the hearing profile.—DAP

7,186,211

43.66.Ts TRANSDUCER TO ACTUATOR INTERFACE

Robert Edwin Schneider and Scott Allan Miller III, assignors to Otologics, LLC
6 March 2007 (Class 600/25); filed 9 April 2004

Transducers in implanted hearing devices must be oriented and located at a desired position relative to the ossicles. Methodology is recommended for achieving a sealed interconnection between a movable actuator that stimulates an ossicle and the transducer housing. To facilitate easier positioning, at least one portion of the transducer housing that encloses the



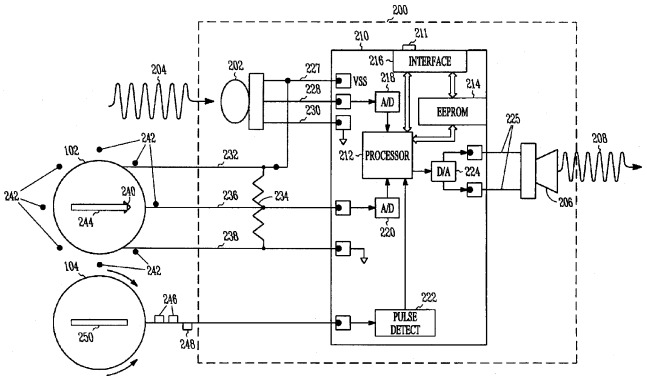
magnet and coil of the driver is rotatable relative to the transducer mounting so that the center of rotation does not move during rotation. A seal is formed around the magnet and coil connected to the actuator to protect against body fluids.—DAP

7,184,564

43.66.Ts MULTI-PARAMETER HEARING AID

Joyce Rosenthal, assignor to Starkey Laboratories, Incorporated
27 February 2007 (Class 381/322); filed 9 September 2003

Some hearing aid dispensers do not have computer access to program digital hearing aids. However, use of digital hearing aids is still possible by controlling them with potentiometers instead of a computer. Small hearing aids have space for very few potentiometers, thus limiting flexibility. Externally manipulated parameter-select and parameter-adjust methods are pro-



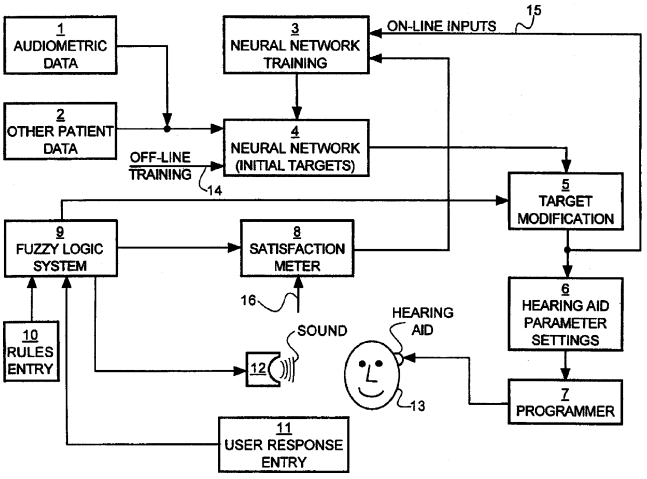
posed to increase the number of parameters controlled and to adjust the parameter values, respectively. A parameter-select potentiometer interfaces to a digital processor to select a particular parameter in one portion of memory. Then, rotation of a parameter-adjust potentiometer sends pulses to the digital processor to adjust the value of the selected parameter.—DAP

7,187,778

43.66.Ts NEUROFUZZY BASED DEVICE FOR PROGRAMMABLE HEARING AIDS

Stavros Photios Basseas, assignor to Beltone Electronics Corporation
6 March 2007 (Class 381/314); filed 29 December 2003

Programmable hearing aid performance may be changed and optimized for individuals with hearing loss by taking into account wearer preferences and complaints. Target electroacoustic parameters representing ''optimum'' performance are first generated by a multilayer neural network that



represents a priori knowledge entered initially during an offline training session. The neural network is modified by wearer comments while listening to sound stimuli during and after the initial hearing aid fitting.—DAP

7,039,578

43.70.Kv METHOD AND APPARATUS FOR TRAINING FOREIGN LANGUAGES

Yoon-Yong Ko and Sang-Hyun Bae, both of Kwangju, Republic of Korea
2 May 2006 (Class 704/8); filed in Republic of Korea 25 April 2000

A device for use in learning a foreign language plays audio or audio/visual items consisting of the presentation of phrases in the target language. The patent describes such a system in which the time interval between playback of items is determined as a combination of a delay that can be set by the user and the length of the previously spoken item. Thus, the listener would have a suitable length of time available in which to repeat the phrase before the next item is played. A playback controller allows items to be selected from a particular set of the available material, such as items from a particular lesson. Items from the chosen set may be played in a particular sequence or may be selected randomly.—DLR

7,046,300

43.71.Ma ASSESSING CONSISTENCY BETWEEN FACIAL MOTION AND SPEECH SIGNALS IN VIDEO

Giridharan Iyengar *et al.*, assignors to International Business Machines Corporation
16 May 2006 (Class 348/515); filed 29 November 2002

The patent lists a number of cases for which it is desirable that a computer be able to analyze and track lip movements in a video signal, correlating these with an accompanying sound track. Such cases might include quality evaluation during video production, speaker detection during a conference, speaker verification during biometric (voice) evaluation, and the ability of the computer to recognize when it is being spoken to. Details of both audio and video signal analyses are described elsewhere in referenced publications. Even the details of audio/video feature consistency are glossed over here, with only a listing (included in the claims) of the type of analysis features expected from both audio and video signals. In other words, what is patented here is little more than the idea that this kind of tracking could be done.—DLR

7,047,189

43.72.Dv SOUND SOURCE SEPARATION USING CONVOLUTIONAL MIXING AND A PRIORI SOUND SOURCE KNOWLEDGE

Alejandro Acero *et al.*, assignors to Microsoft Corporation
16 May 2006 (Class 704/222); filed 18 November 2004

Perhaps to be considered as the ultimate in background noise reduction for speech signals, methods of sound source separation go a long way toward providing a clean speech signal in various environments when speech recognition is to be performed. Success, however, typically requires more sophisticated signal processing than the usual reference mic subtraction or beamforming methods used in multiple-mic systems. For example, delay-and-sum beamforming as well as blind-source separation techniques will often fail to locate a source when room reverberation is present in the signal. The system described here would use either a vector codebook or a hidden Markov model speech recognizer to build a reconstruction filter for each voice signal to be separated from a mix. As the analysis proceeds, additional a priori information about each speaker is accumulated. Maximum a posteriori estimates are used to build up FIR reconstruction filters for each speaker. Detailed equations are presented for the methods described to process multiple microphone signals to achieve the patented results. The final concept is covered by just eight claims.—DLR

7,177,430

43.72.Gy DIGITAL ENTROPING FOR DIGITAL AUDIO REPRODUCTIONS

Jason Seung-Min Kim, assignor to Portalplayer, Incorporated
13 February 2007 (Class 380/252); filed 31 October 2001

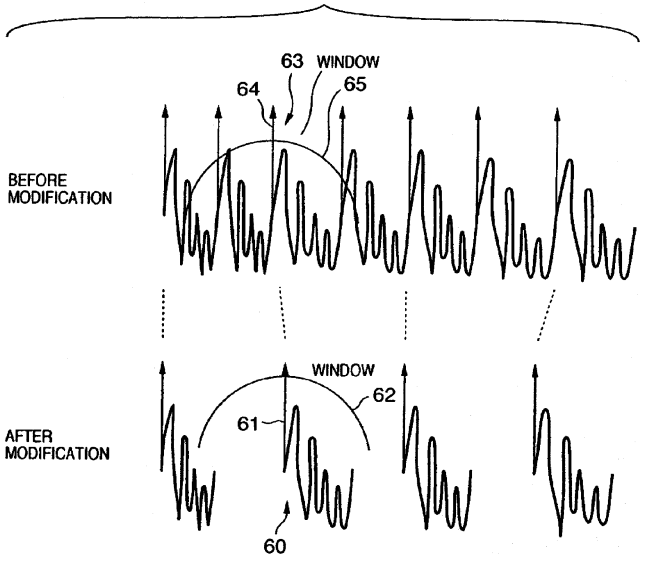
Unauthorized copying is discouraged by adding white noise to the digital audio signal to degrade it. The noise level increment is increased after each successive reproduction. The magnitude of the white noise signal is produced by a random number generator and is adjusted typically to initially produce an approximately 3 dB degradation of the digital audio signal.—DAP

7,039,588

43.72.Ja SYNTHESIS UNIT SELECTION APPARATUS AND METHOD, AND STORAGE MEDIUM

Yasuo Okutani and Yasuhiro Komori, assignors to Canon Kabushiki Kaisha
2 May 2006 (Class 704/258); filed in Japan 31 March 2000

The speech synthesizer presented here is based on the concatenation of short units of speech from a large database. The patent covers several choices in the size of the stored speech units, primarily either diphones or half-diphones. Several methods are presented for searching the database and for modifying the concatenated speech units so as to reduce the distortion



introduced during the generation of the output prosodic profile. The speech units are combined using pitch-synchronous overlap-and-add methods. The database search methods and the techniques used for adjusting the amplitude and duration of the concatenated units are presented in considerable detail.—DLR

7,043,432

43.72.Ja METHOD AND SYSTEM FOR TEXT-TO-SPEECH CACHING

Raimo Bakis *et al.*, assignors to International Business Machines Corporation
9 May 2006 (Class 704/260); filed 29 August 2001

This speech synthesis system would maintain a cache of previously synthesized text strings along with the corresponding synthetic speech output. If a new text to be synthesized is found in the cache, then the previously synthesized speech signal can be replayed immediately, saving the time and

computational resources needed to resynthesize that item. There is a brief mention of the possibility of storing additional information in the cache along with the synthesized speech, which could include, for example, a text string to be displayed, different from that used as the synthesizer input. A score is also maintained with each item stored in the cache, allowing selected cache items to be deleted as the cache memory space is filled.—DLR

7,039,590

43.72.Ne GENERAL REMOTE USING SPOKEN COMMANDS

Daniel Luchaup, assignor to Sun Microsystems, Incorporated
2 May 2006 (Class 704/275); filed 30 March 2001

This patent initially defines a small-vocabulary command recognizer as a ''trivial'' speech recognition system and then takes pains to clarify that the voice-operated remote control presented is not a trivial system, but, rather, would be capable of recognizing an extended command, such as ''VCR, tape the program from 8 to 9 pm and from 10 to 11 pm tonight.'' Curiously, however, after all this introduction of how fancy the system would be, there is absolutely no discussion of the recognition mechanism; no mention of spectral features, vocabulary, grammars, or any of the other typical accouterments of a recognition system, let alone processor time or memory requirements. The claims at least provide that this non-trivial recognition system may include a microphone in addition to the usual infrared or rf link and possible LCD display.—DLR

7,043,427

43.72.Ne APPARATUS AND METHOD FOR SPEECH RECOGNITION

Ralf Kern and Karl-Heinz Pflaum, assignors to Siemens Aktiengesellschaft
9 May 2006 (Class 704/233); filed in Germany 18 March 1998

This patent deals with the issue of speech recognition difficulties resulting from differences in the speech signal due to the acoustic characteristics of the room, which may be present or absent in the signal depending on the distance of the speaker's mouth from the microphone. The solution presented here would ''correct'' a close-talking signal by convolving it with a room reverberation impulse response such that the resulting signal includes a reverberation matching that which was used during the initial recognizer training. There is no discussion of how such an approach would deal with a speech signal collected during reverberant (speaker-phone) conditions differing from the original room used during the recognizer training.—DLR

7,043,436

43.72.Ne APPARATUS FOR SYNTHESIZING SPEECH SOUNDS OF A SHORT MESSAGE IN A HANDS FREE KIT FOR A MOBILE PHONE

Byung-Seok Ryu, assignor to Samsung Electronics Company, Limited
9 May 2006 (Class 704/270.1); filed in Republic of Korea 5 March 1998

This patent describes a relatively simple addition to the signal-processing (DSP) software in a mobile phone to complete the scenario of full hands-free operation of the phone in an automobile. The missing piece, according to the patent, is that when a call comes in, certain information, such as caller ID, is transmitted to the phone using a service called short message service (SMS). These SMS messages typically cause a beep or similar notification and are then displayed as text on the phone. This patent would add the ability to synthesize a speech output signal based on the received SMS message. Brief voice commands, uttered by the user, would guide the phone through the sequence of operations, including speaking the SMS message and handling the call as the user wishes.—DLR

7,046,777

43.72.Ne IVR CUSTOMER ADDRESS ACQUISITION METHOD

Vicki L. Colson *et al.*, assignors to International Business Machines Corporation
16 May 2006 (Class 379/142.06); filed 2 June 2003

When a customer calls a company for the purpose of obtaining information or, more particularly, to place an order, such calls are frequently handled by a computerized interactive voice response (IVR) system. However, one of the more difficult tasks for such a system is to correctly understand and log in the caller's information, such as e-mail and postal mailing addresses. But, this is exactly the information that must be logged correctly in order to properly serve the caller. The approach patented here would supplement the IVR system with an online directory assistance database. The caller's telephone number can typically be obtained using a caller ID system. This number may optionally be verified verbally by the caller. That number may then be used to obtain further caller address details from the directory assistance database. Is this all obvious, or what?—DLR

7,177,800

43.72.Ne METHOD AND DEVICE FOR THE PROCESSING OF SPEECH INFORMATION

Joseph Wallers, assignor to digital design GmbH
13 February 2007 (Class 704/201); filed in Germany 3 November 2000

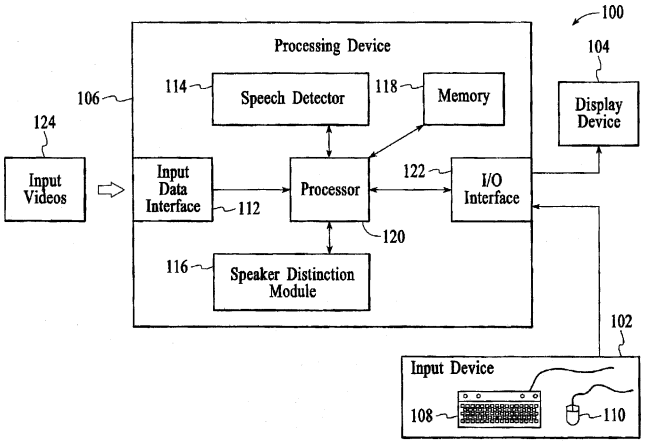
Methodology for a speech-based memory is proposed by recording and searching/reproducing speech information without training, assignment of certain words by the user, or establishing a vocabulary. At least one spoken voice memorandum and one spoken search word are recorded, transformed into a representation suitable for pattern matching, and are compared.—DAP

7,184,955

43.72.Ne SYSTEM AND METHOD FOR INDEXING VIDEOS BASED ON SPEAKER DISTINCTION

Pere Obrador and Tong Zhang, assignors to Hewlett-Packard Development Company, L.P.
27 February 2007 (Class 704/231); filed 25 March 2002

Audio content is used to index, distinguish, and retrieve desired video scenes according to the speech characteristics of speakers. An energy detector, zero-crossing-rate (ZCR) variance analyzer, and ZCR amplitude span



analysis are used to determine if the audio portions contain speech. Spectral features are generated for audio portions of the video segments containing speech.—DAP

7,184,957

43.72.Ne MULTIPLE PASS SPEECH RECOGNITION METHOD AND SYSTEM

John R. Brookes and Norikazu Endo, assignors to Toyota Info Technology Center Company, Limited
27 February 2007 (Class 704/246); filed 10 October 2002

More accurate speech recognition performance is said to result for car navigation systems by recognizing speech multiple times in parts rather than attempting to recognize the entire utterance in one pass. Context of the first pass is determined by matching a first part of the input speech signal and the first pass result is used to generate a second pass grammar for the input speech to the second pass.—DAP

7,175,598

43.80.Vj ULTRASOUND DIAGNOSIS APPARATUS THAT ADJUSTS A TIME PHASE BETWEEN A PLURALITY OF IMAGE SERIES

Naoki Yoneyama, assignor to Kabushiki Kaisha Toshiba
13 February 2007 (Class 600/443); filed in Japan 18 June 2002

Two image sequences are obtained, each under a different condition. A processor in the system utilizes a variable such as an ECG measured with each set of images. Based on the variable, the time phase of the second set of images is adjusted by the processor relative to the time phase of the first set of images.—RCW

7,179,449

43.80.Vj ENHANCED ULTRASOUND DETECTION WITH TEMPERATURE-DEPENDENT CONTRAST AGENTS

Gregory M. Lanza *et al.*, assignors to Barnes-Jewish Hospital
20 February 2007 (Class 424/9.321); filed 30 January 2001

Ultrasound reflectivity changes that depend on the temperature of a contrast agent are used to enhance ultrasound images obtained either alone or in conjunction with drug delivery, hypothermia, cryotherapy, or with other imaging modalities.—RCW

7,186,219

43.80.Vj CALIBRATION OF A DOPPLER VELOCIMETER FOR STROKE VOLUME DETERMINATION

Markus J. Osypka and Donald P. Bernstein, assignors to Osypka Medical GmbH
6 March 2007 (Class 600/504); filed 10 October 2002

The cross-sectional area of the aortic valve is obtained using a calibration method. In this method, a reference stroke volume is determined by a method other than Doppler velocimetery and a reference systolic velocity integral assumed to be proportional to stroke volume is obtained by Doppler velocimetery. A constant of proportionality determined by the calibration process is then used with further Doppler velocimetery measurements to calculate cardiac stroke volume.—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.

A_0 mode interaction with a plate free edge: Theory and experiments at very low frequency by thickness product (L)

Guillemette Ribay,[a] Stefan Catheline, Dominique Clorennec, Ros Kiri Ing, and Mathias Fink

Laboratoire Ondes et Acoustique, Université Denis Diderot, UMR CNRS 7587 ESPCI, 10 rue Vauquelin, 75231 Paris Cedex 05, France

(Received 24 November 2006; revised 4 May 2007; accepted 18 May 2007)

When a plane acoustic wave reaches a medium with an impedance infinite or null, it experiences a phase shift of zero or π and its amplitude on the edge is maximum or vanishes. The case of a flexion wave (A_0 Lamb wave) at a free end is also simple; its amplitude is multiplied by a factor $2\sqrt{2}$ and the phase shift is $\pi/2$. The evanescent wave at the origin of these phenomena, perfectly described by the classical flexural plate theory, is identified as the imaginary A_1 mode of the exact Rayleigh-Lamb theory. The experiences confirm the theoretical predictions. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2749444]

PACS number(s): 43.40.Dx, 43.20.Mv, 43.20.El, 43.20.Ye [DF] Pages: 711–714

I. INTRODUCTION

When reflected on a plate's free edge, the first antisymmetric Lamb wave A_0 gives birth to evanescent waves: Indeed, according to the three-dimensional elastic theory, no combination of the incident and reflected A_0 mode can verify the free edge conditions.[1] The Lamb wave is phase shifted by a value depending on the frequency by thickness product (fh). Recent studies were done for such a product $fh \geq 1$ MHz mm.[2] Diligent *et al.*[3] could compute with a semi-analytical model the phase shift of the reflected A_0 mode. In this letter, the attention is focused at $fh \leq 0.1$ MHz mm. At this low frequency by thickness product, the A_0 wave becomes a flexion wave.

The ability of the simple and purely analytical flexural plate theory to account for the propagating A_0 mode is well known.[4] It also describes an evanescent wave and a phase shift created by the reflection on a free edge. However, to our knowledge, a quantitative experimental study of the flexural plate model accuracy is yet to be done. Therefore, the objective of this letter is on one hand to show that this evanescent wave can formally be identified as the imaginary A_1 mode of the exact Rayleigh-Lamb theory and, on the other hand, to demonstrate through experiments the high degree of accuracy of the simple flexural plate theory.

All these results are of interest in the field of tactile interactive experiments since it was shown that the energy of a touch contact propagates as low frequency flexion waves within a plate,[5,6] and that the wave is reflected many times before vanishing.

II. PREDICTIONS OF THE FLEXURAL PLATE THEORY AND THE LAMB WAVE THEORY

According to the classical plate theory,[4] the transverse displacement w of a free vibrating plate obeys the following equation:

$$\Delta^2 w + \frac{\rho h}{D}\frac{\partial^2 w}{\partial t^2} = 0, \tag{1}$$

where $D = Eh^3/12(1-\nu^2)$, E being the Young's modulus, h the plate thickness, ν the Poisson ratio, ρ the density ($\mathrm{kg/m^3}$).

The assumptions made to obtain this equation are the following: w does not depend on the thickness coordinate, and the shear deformations are neglected.

However, when are these assumptions valid? To answer this question, one needs the Rayleigh-Lamb theory. Actually, as demonstrated, for example, by Royer,[7] if k is the wave number, the first antisymmetric Lamb mode A_0 is almost transverse when $kh \ll 1$, and w does not depend on the thickness anymore; moreover, the Rayleigh Lamb dispersion equation becomes

[a]Author to whom correspondence should be addressed. Electronic mail: guillemette.ribay@loa.espci.fr

$$\omega^2 = \frac{V_P^2 . h^2}{12} k^4 \tag{2}$$

with Vp the plate velocity defined as $2V_S\sqrt{1-V_S^2/V_L^2}$, V_S being the shear velocity, V_L the longitudinal velocity, ω the pulsation.

Actually, in the Fourier domain in space and in time, Eq. (2) can be considered as an equation between operators: A multiplication by $i\omega$ corresponds to a derivation versus time, and a multiplication by ik is a derivation versus the spatial coordinates. Now in an isotropic plate, it is easy to show that the square plate velocity V_P^2 is equal to $E/\rho(1-\nu^2)$; then Eq. (2) is the same as Eq. (1). As a consequence, the assumptions made in the flexural plate theory are valid for frequency by thickness product small compared to one.

Now consider a plate infinite in the y direction, defined for $x>0$, and with a free edge at $x=0$. If a plane harmonic A_0 wave is incident orthogonally to the free edge, w is a function of x and t. The well-known solutions w of Eq. (1) are then linear combinations of

$$e^{i(\omega t+kx)}, \quad e^{i(\omega t-kx)}, \quad e^{i(\omega t+ikx)}, \quad e^{i(\omega t-ikx)}. \tag{3}$$

The first term is the incident wave on the free edge, propagating in the decreasing direction of x; the second term is the reflected wave, propagating in the opposite direction (increasing x), the third one is the evanescent wave created at $x=0$, whose amplitude decreases with x; the last term would be the evanescent wave created at the other edge of the plate, but we suppose that it is situated at x tending towards infinity, and therefore this last term is negligible.

Thus, $w(x,t)=e^{i\omega t}(Ae^{ikx}+Be^{-ikx}+Ce^{-kx})$, where A, B, C are arbitrary constants (complex). Now, according to the classical plate theory,[1] the free edge condition is

$$\frac{\partial^2 w}{\partial x^2} = \frac{\partial^3 w}{\partial x^3} = 0. \tag{4}$$

Therefore, A, B, and C are solutions of the following equations:

$$-A-B+C=0,$$

$$-iA+iB-C=0. \tag{5}$$

As a consequence, if the amplitude of the incident wave is A, we deduce that the amplitude of the reflected wave is $B=-iA$; that is, $B=A.e^{-i\pi/2}$: the wave is phase shifted by $\pi/2$.

Moreover, the amplitude of the evanescent part is $C=A(1-i)=A\sqrt{2}e^{-i\pi/4}$. The expression of the evanescent part of the wave is: $w(x,t)=A\sqrt{2}e^{i(\omega t-\pi/4)}e^{-kx}$. Its amplitude is thus divided by 10 when $e^{-kx}=0.1$; that is $2\pi x/\lambda=2.3$. As a conclusion, the evanescent wave can be considered as negligible for x greater than half the wavelength.

Actually, according to the exact Rayleigh-Lamb theory, many nonpropagating evanescent modes exist. Which one corresponds to the evanescent wave of the flexural plate theory? Figure 1 shows the exact dispersion curves obtained from the Rayleigh-Lamb theory for an aluminium plate. When k is real, the result is the A_0 dispersion curve (continuous line). When k is imaginary, the result is the evanescent A_1 mode (dashed line). When fh tends towards zero, the Rayleigh-Lamb curves merge. Thus, wave numbers k of same modulus are found for a propagative and a nonpropagative wave as described by the general solution of flexural waves, Eq. (3). The dispersion curve of the classical plate theory (dotted line), Eq. (2) is shown indeed to coincide with the two former exact dispersion curves. We can therefore conclude that the evanescent wave predicted by the classical plate theory corresponds to the imaginary nonpropagating branch of the A_1 mode.

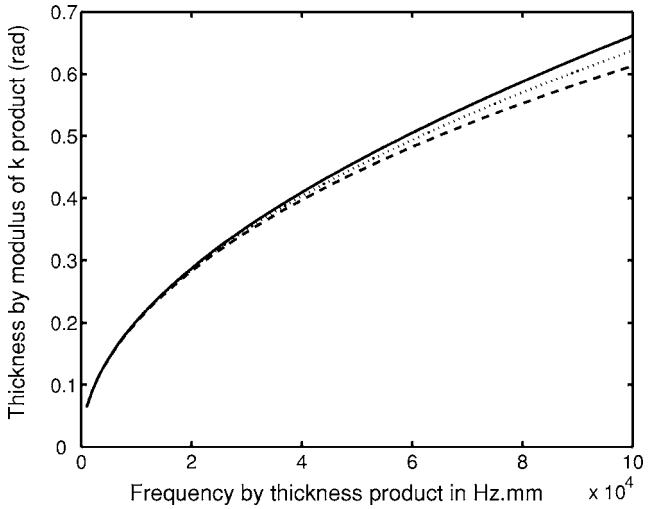


FIG. 1. Modulus of wave number by thickness product in rad versus fh product in Hz mm: Continuous line: Propagating mode A_0 predicted by Lamb theory, dashed line: evanescent mode A_1, purely imaginary, predicted by Lamb theory; dotted line: Same modes predicted by the classical flexural plate theory.

The other nonpropagating modes predicted by Rayleigh-Lamb theory are complex modes, i.e. k, is not purely imaginary anymore. As shown by Lowe and Diligent,[2] their attenuation at fh less than 0.1 MHz mm is extremely high: In a 1-mm-thick plate, their minimum attenuation (i.e., imaginary part of wave number) is greater than 60 dB/mm, whereas according to Fig. 1, at fh equal to 0.1 MHz mm, it is around 5 dB/mm for the imaginary evanescent mode A_1: Therefore, the complex modes can be neglected here.

III. EXPERIMENTS AT VERY LOW FREQUENCY BY THICKNESS PRODUCT

In order to study the phase shift of a plane wave reflected on a plate's free edge, we have chosen to work in a continuous monochromatic regime, instead of using a transient technique,[3] unsuitable here because of the strong dispersion occurring at such fh product. We could thus observe the interference between incident and reflected waves. Indeed, according to the above mentioned theory, vibration nods will appear. The total transverse displacement will be equal to

$$w(x,t) = Ae^{i(\omega t+kx)} + Ae^{-i\frac{\pi}{2}}e^{i(\omega t-kx)} + A\sqrt{2}e^{-i\frac{\pi}{4}}e^{i\omega t}e^{-kx}.$$

The real part w_r of w is $A\left(\cos(\omega t+kx)+\cos\left(\omega t-kx-\pi/2\right)+\sqrt{2}e^{-kx}\cos\left(\omega t-\pi/4\right)\right)$. It can also be written as

$$w_r(x,t) = A\cos\left(\omega t - \frac{\pi}{4}\right)\left(2\cos\left(kx + \frac{\pi}{4}\right) + \sqrt{2}e^{-kx}\right). \quad (6)$$

At the edge ($x=0$), the amplitude of the vibration is equal to $2\sqrt{2}A$, whereas, far from the edge, the amplitude of the ventral vibration is equal to $2A$. This "resonance" phenomenon is very different from the well-known "edge resonance"[8,9] that occurs when a symmetric mode reflects on a plate free edge, provided that its frequency range contains the proper frequency component. Indeed, the resonance observed in the case of A_0 incident mode is not frequency dependent. At fh product <0.1 MHz mm, i.e., much less than the cutoff frequencies of higher modes (propagating or nonpropagating), no frequency resonant behavior is expected. At higher fh product, when other modes are non-negligible, one can wonder whether such a frequency dependent resonance could occur in the antisymmetric case or not. However, as the attention of this letter is being focused on very low fh product, we will not answer this question here.

A nod appears at position x if $w_r(x,t)$ is equal to zero at any time t. Therefore, the positions of the nods are the solutions of the following equation:

$$2\cos\left(kx + \frac{\pi}{4}\right) = -\sqrt{2}e^{-kx}. \quad (7)$$

The right-hand side is simply zero when the evanescent wave is negligible, that is for x greater than half the wavelength. In such a case, the nods positions are:

$$x_{\text{nod}} = \frac{\lambda}{8} + \frac{n\lambda}{2}, \quad (8)$$

n being an integer. Only the first nod position is influenced by the evanescent part: It takes place for $kx=1.038$; that is $x=0.1652.\lambda$. (In the absence of evanescent waves, it would have been equal to $0.125.\lambda$.)

Therefore, the nods position takes into account (a) the phase shift when A_0 is reflected, (b) the evanescent wave (for the first nod).

More generally, suppose that one does not know the phase shift value (noted $\Delta\varphi$). For x greater than half a wavelength, the nods positions would be: $x_{\text{nod}}=\lambda/2(\pi-\Delta\varphi/2\pi)+(n-1)\lambda/2$. Then the measured phase shift would be

$$\Delta\varphi = \pi - 2\pi.\left(x_{\text{nod number } n}.\frac{2}{\lambda} - (n-1)\right). \quad (9)$$

As illustrated in Fig. 2, a first experiment was performed in a bar, because it is equivalent to the case where a plane wave is incident orthogonally on the free edge of a plate infinite in the other direction, except that, because of Poisson effect, the square plate velocity has to be replaced by $V_P^2 \times(1-\nu^2)$. Absorbing material (acoustic foam) is put at the other edge of the bar, so that the wave reflected at this edge can be considered as negligible; the medium is semi-infinite.

A piezoelectric transducer, glued on a 1-mm-thick and 1-m-long stainless steel bar, emits a continuous harmonic wave at 5 kHz, and the transverse component of the wave is measured thanks to a heterodyne interferometer[10] coupled to a low frequency demodulator. Figure 3 shows the normalized

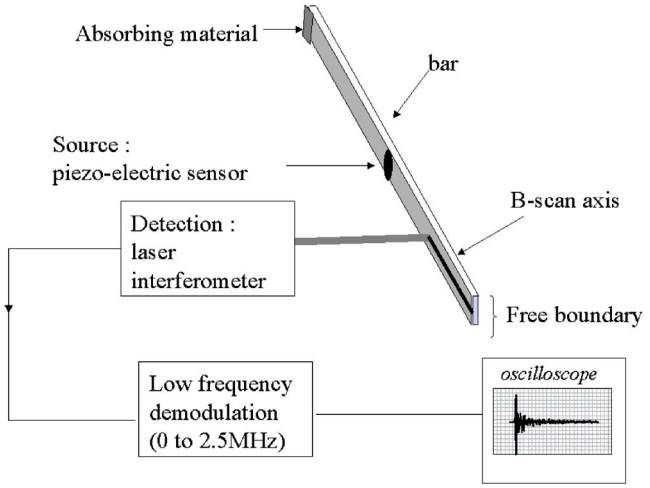


FIG. 2. (Color online) Experimental setup for the study of flexion waves reflecting at the free end of a bar.

amplitude of the transverse velocity of the wave measured at distance x from the free edge, and the equivalent theoretical quantity (based on Eq. (6)). Both are in very good agreement. Moreover, the amplitude enhancement at the free edge can be clearly seen on both curves. The experimental ratio of the amplitude at the edge to the ventral vibration amplitude far from the edge is equal to $0.98\times\sqrt{2}$, which is very close to the theoretical value ($\sqrt{2}$).

The first nod is measured at 6.8 mm+/−0.2 mm. The half wavelength is the distance between two consecutive nods away from the edge, that is 20.2 mm; therefore, the theoretical position of the first nod is 6.68 mm, which is in very good agreement with the experiment.

Experiments on a plate were also achieved; the difficulty consists in creating a line-shaped source so that the wave can be considered as a plane one. To this end, we used a 30 cm by 30 cm by 0.5 mm Duraluminum plate, and a Q-switched Nd: yttrium–aluminum–garnet laser is used as a source. A beam expander and a cylindrical lens were used to focus the beam onto a 13-mm-long line.

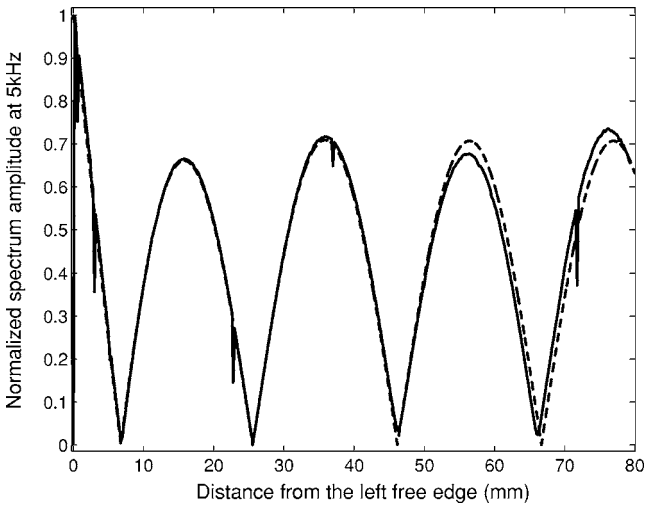


FIG. 3. Amplitude of the transverse velocity measured at position x, versus x (continuous line) and theoretical amplitude (dashed line) for a 5 kHz mm flexural wave.

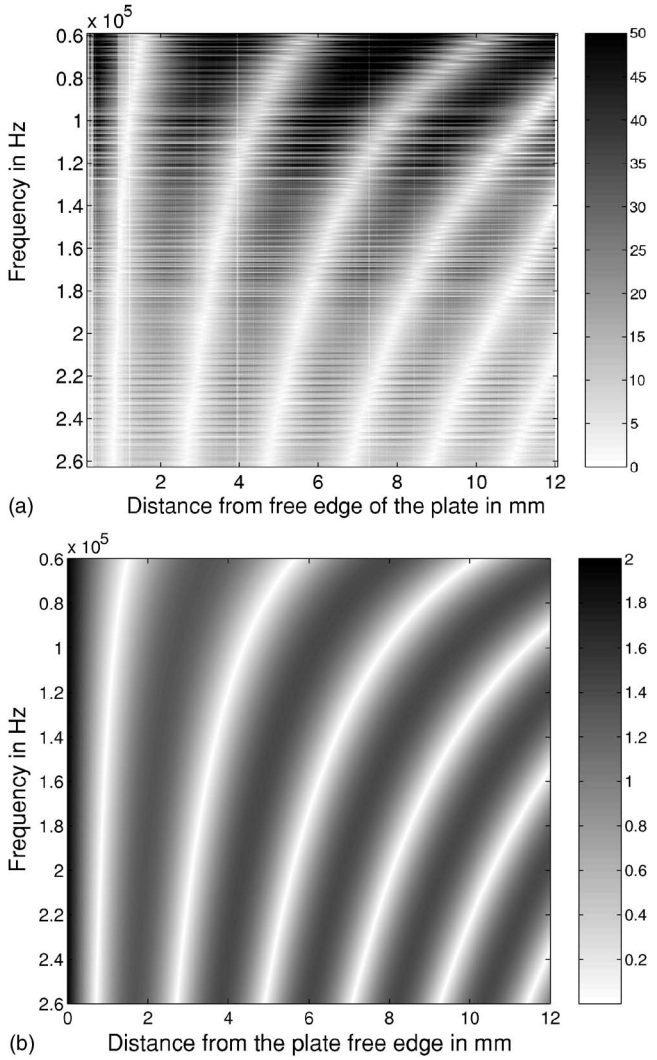


FIG. 4. (a) Experiment on a 0.5-mm-thick Duraluminum plate: Amplitude (coded in gray scale, arbitrary unit) of FFT of displacements measured at position x, versus distance x and versus frequency. (b) Theoretical amplitude of FFT of displacements at position x, versus distance x and versus frequency, for a 0.5-mm-thick Duraluminum plate (shear velocity: 3130 m/s and longitudinal velocity: 6320 m/s). The theory is in good agreement with the experiment (Fig. 4(a)).

The source signal is a pulse; the nods will be observable by taking the fast Fourier transform (FFT) of all signals at the same frequency. Because of the width of the line source, to ensure that the wave is a quasi-plane one, the source is placed 13 mm from the edge, and we have to focus on a frequency at which the wavelength is smaller than 13 mm (so that several nods can be created between the source and the edge). The transverse displacement component is measured thanks to the same heterodyne interferometer coupled to the low frequency demodulator. Figure 4 shows the amplitude of the FFT of each signal versus the abscissa x and the frequency f; the nods can be clearly seen.

The free edge position is at $x=0+/-30$ μm. The nods positions are then: 0.9, 2.99, 5.44, 7.87, and 10.3 mm, measured with $+/-5$ μm precision. The wavelength is then 4.86 mm$+/-10$ μm; therefore, the theoretical first nod position is 0.8 mm. The deviation from the measured one is probably due to the fact that, contrary to the previous experiment where $fh=5$ kHz mm, at this new fh product, as shown in Fig. 1, the modulus of the wave number of the imaginary evanescent mode A_1 cannot be considered the same as the modulus of A_0 wave number.

However, it does not prevent us from measuring the phase shift of A_0, because we do so far away enough from the edge so that the evanescent waves are negligible.

According to Eq. (9), the measured phase shift is $\Delta\varphi = \pi - 2\pi.(10,3.2/4,86-(5-1)) = 1.64$ rad$+/-0.165$ rad $(+/-0.11$ rad due to the uncertainty of the wavelength measurement, and $+/-0.055$ rad due to the free edge position measurement uncertainty.). As a conclusion, the measured phase shift is 94° (+/−9.2°).

IV. CONCLUSION

In this letter, experiments in bars and plate at low frequency by thickness product (<0.1 MHz mm) confirm quantitative predictions of the flexural plate theory. This perfect agreement in phase and amplitude is illustrated by Figs. 3 and 4. It perfectly accounts for the evanescent wave contribution known as the imaginary A_1 mode in the complete Rayleigh Lamb theory. These results are of importance in the frame of tactile interactive experiments. Indeed the localization of finger impacts on plates relies on the whole vibration pattern, involving many reflections on the free edges.

ACKNOWLEDGMENTS

This work was partially financed by the European FP6 IST Project "Tangible Acoustic Interfaces for Computer-Human Interaction (TAI-CHI)." The support of the European Commission is gratefully acknowledged.

[1] P. J. Torvik, "Reflection of wave trains in semi-infinite plates," J. Acoust. Soc. Am. **41**, 346–353 (1967).
[2] M. J. S. Lowe and O. Diligent, "Reflection of the fundamental Lamb modes from the ends of plates," Rev. Prog. Quant. Nondestr. Eval. **20**, 89–96 (2001).
[3] O. Diligent, M. J. S. Lowe, E. Le Clézio, M. Castaings, and B. Hosten, "Prediction and measurement of nonpropagating Lamb modes at the free end of a plate when the fundamental antisymmetric mode A_0 is incident," J. Acoust. Soc. Am. **113**(6), 3032–3042 (2003).
[4] K. F. Graff, *Wave Motion in Elastic Solids* (Dover, New York, 1991).
[5] R. K. Ing, N. Quieffin, S. Catheline, and M. Fink, "In solid localization of finger impacts using acoustic time-reversal process," Appl. Phys. Lett. **87**, 204104 (2005).
[6] G. Ribay, S. Catheline, D. Clorennec, R. K. Ing, N. Quieffin, and M. Fink, "Acoustic impact localization in plates: Properties and stability to temperature variation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 378–385 (2007).
[7] D. Royer and E. Dieulesaint, *Elastic Waves in Solids* (Springer, Berlin, 1999), Vol. **1**.
[8] E. Le Clezio, M. V. Predoi, M. Castaings, B. Hosten, and M. Rousseau, "Numerical predictions and experiments on the free-plate edge mode," Ultrasonics **41**, 25–40 (2003).
[9] Pagneux "Revisiting the edge resonance for Lamb waves in a semi-infinite plate," J. Acoust. Soc. Am. **120**, 649–656 (2006).
[10] D. Royer and E. Dieulesaint, "Optical detection of sub-angstrom transient mechanical displacement," *Proceedings of the 1986 IEEE Ultrasonics Symposium* (IEEE, New York, 1986), p. 527.

The ontogeny of echolocation in a Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientalis*) (L)

Songhai Li
Institute of Hydrobiology, The Chinese Academy of Sciences, Wuhan, 430072, People's Republic of China and Graduate School of the Chinese Academy of Sciences, Beijing, 100039, People's Republic of China

Ding Wang,[a] Kexiong Wang, and Jianqiang Xiao[b]
Institute of Hydrobiology, The Chinese Academy of Sciences, Wuhan, 430072, People's Republic of China

Tomonari Akamatsu
National Research Institute of Fisheries Engineering, Fisheries Research Agency, Hasaki, Kamisu, Ibaraki 314-0408, Japan

(Received 21 January 2007; revised 7 May 2007; accepted 14 May 2007)

Acoustic and concurrent behavioral data from one neonatal male Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientalis*) in captivity were presented. The calf click train was first recorded at 22 days postnatal, and the frequency of hydrophone-exploration behavior with head scanning motions in conjunction with emissions of click trains by the calf increased gradually with age. The echolocation clicks in the first recorded click train were indistinguishable from those of adults. Calf echolocation trains were found to decrease in maximum click-repetition rate, duration, and number of clicks per train with age while the minimum click-repetition rate remained more consistent. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2747203]

PACS number(s): 43.80.Ka [WWA] Pages: 715–718

I. INTRODUCTION

Odontocetes use echolocation to orient in the environment and capture prey (Au, 1993). Before neonatal animals begin to hunt fish successfully, they must be capable of producing high frequency sonar signals and of processing the environment information contained in the sonar echoes. The ontogeny of echolocation system in the neonatal odontocetes, however, is poorly understood. Data on the neonatal appearance of echolocation in odontocetes only exist for a few species, including bottlenose dolphin (*Tursiops truncates*; Carder and Ridgway, 1983; Lindhard, 1988; Reiss, 1988; Kamminga and Terry, 1994; Hendry, 2004), killer whale (*Orcinus orca*; Bowles *et al.*, 1988), sperm whales (*Physeter macrocephalus*; Watkins *et al.*, 1988; Madsen *et al.*, 2003), Dall's porpoise (*Phocoenoides dalli*; Bain, 1988), harbor porpoise (*Phocoena phocoena*; Kamminga and Terry, 1994), and beluga whale (*Delphinapterus leucas*; Vergara and Barrett-Lennard, 2003). Which components of echolocation are learned through social contact and which ones are innate, unconditioned acoustic behaviors is an important area still open to investigation.

With the birth of one male Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientalis*) in our facility (the Baiji Aquarium), we got an opportunity to investigate the neonatal appearance and ontogeny of echolocation in this species, which has never been studied before. The current study aimed to identify the components of echolocation that are fully developed shortly after birth and those that changed as the newborn calf matured. The investigated features of echolocation include time and frequency characteristics of single click, train duration, number of clicks per click train, and click-repetition rate (CRR).

II. MATERIALS AND METHODS

The investigated neonatal male finless porpoise was born in a 3-m-deep, 25×7 m kidney-shaped pool in the Baiji Aquarium on July 5, 2005 (Wang *et al.*, 2005). During the study period, the calf resided together with the mother porpoise, approximately 8 years old, and an adult female tank mate, approximately 11 years old.

Sound recording and ad lib notation using event based sampling methods were utilized to document sound production and behavioral development from day 1-181 after birth, under favorable conditions. The data presented in this paper are based on at least daily 1 h observations and recordings from the birth through 30 days postnatal, roughly weekly 1 h observations and recordings from day 30-60, and monthly 1 h from day 60-181.

A hydrophone (ST1020, Oki Electric Co. Ltd., Japan) with sensitivity of −180 dB *re*: 1 V μPa^{-1} +3/−12 dB, up to 150 kHz, was utilized to input the porpoise vocalizations to a Sony PCHB244 digital data recorder, with a flat frequency response from dc to 147 kHz within 3 dB. The hydrophone was located at 0.5 m depth and 0.5 m to tank wall. A high-pass filter of 100 Hz was incorporated to the vocalization recordings.

[a]Author to whom correspondence should be addressed. Electronic mail: wangd@ihb.ac.cn
[b]Present address: Psychology Department, Hunter College, CUNY, New York, 10021 NY.

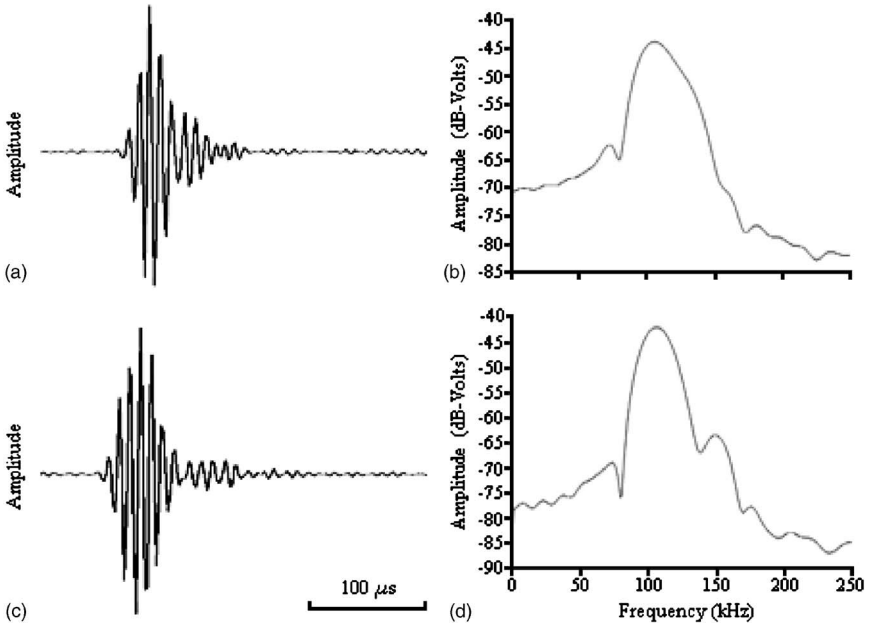


FIG. 1. (a) Wave form of a representative click from the earliest recorded calf click train; (b) power spectrum of the click depicted in (a); (c) and (d): Comparison examples of click wave form and spectrum from the mother porpoise in the same tank.

Analysis of vocalizations was done using PC-based DT-Disk™ (Direct-to-disk Recorder, Version 1.11, November 2002, American Engineering Design) and SIGNAL™ software (Version 4.03, December 2005; American Engineering Design). After a preliminary aural inspection, the taped vocalizations were re-digitized at a sampling rate of 500 kHz by 12 bit Data Translation-3010 analog I/O cards. By detailed review and inspection of raw recordings (time and frequency characteristics), no low-frequency ($<$70 kHz) components were found in all the vocalizations. Subsequently, for the sake of convenience to analyze, a high-pass digital filter of 70 kHz was performed to the raw recordings to eliminate environmental noise. In order to identify the vocalizing animal and avoid analysis of off-axis signals, only instances in which an individual would vocalize directly at the hydrophone within 2 m were analyzed.

Clicks were defined as very short (less than 100 μs; Li *et al.*, 2005) high-frequency pulses, and a click train was defined as a group of three or more clicks with regular or gradual change of the interclick interval (Moreno *et al.*, 2003). If the interval was changed abruptly within subsequent intervals but was then maintained to be regular or gradual, it was still considered as part of the same click train (Moreno *et al.*, 2003). Once a click train was fully processed and identified, the train duration, which was defined as the span of time between the visually determined onset of the first click in a train and the termination of the last click in a train (Hendry, 2004), and the number of clicks were documented. Click-repetition rate (CRR, clicks/s) was defined as the inverse of the time interval between two subsequent clicks. Several internal commands in SIGNAL™ software were associatively used to extract the CRR contours of click trains. From these CRR contours, the maximum click-repetition rate (CRRmax) and minimum click-repetition rate (CRRmin) of each click train were determined. The acoustic parameters of click trains were fed into EXCEL software and STATISTICA software to be analyzed.

III. RESULTS

At 4 hours postnatal, the mother began to swim with the calf and then nurse him. By day 2, the neonate was able to maneuver well to avoid the tank wall without maternal guidance. From the second through the 20th postnatal day, the neonate swimming was mainly characterized by almost constant mother-infant contact in echelon formation with brief infant departures and re-approaches. During this period, even the calf occasionally passed the hydrophone in isolation, there was no any behavioral evidence showing he was interested in this object and no vocalization was recorded. By days 22 postnatal, the neonate was first observed to exhibit head scanning motions in conjunction with emission of a long click train to the hydrophone when approaching it alone. The first recorded click train had a duration of over 4 s, and number of clicks over 750. In this case, the CRR changed and fluctuated between 90 and 190 Hz (clicks/s, i.e., inverse of the time interval between two subsequent clicks) through the first 0.6 s, gradually increased to 220 Hz between 0.6 and 3.0 s, and then decreased to $\sim$100 Hz at the end. The clicks have apparently short multi-cycle wave form [Fig. 1(a)] and high-frequency spectrum [Fig. 1(b)], and are indistinguishable from adult clicks both in temporal and frequency domains. Comparison examples of click wave form and spectrum from the mother porpoise in the same tank were shown in Figs. 1(c) and 1(d), respectively. The cross-correlation comparison as a quantitative measure of similarity between the click from the click train and the clicks from the two adults, recorded from the same pool and with the same conditions, showed correlation values of 0.86 ± 0.07 ($N=200$) when comparing with temporal domain, and values of 0.98 ± 0.01 ($N=200$) when comparing with frequency domain (power spectra). No open-mouth posture and air bubble production were noticed while scanning and echolocating the hydrophone. Subsequently, the calf exhibited a gradually increased frequency of hydrophone-exploration behavior with head scanning motions in conjunction with emissions of

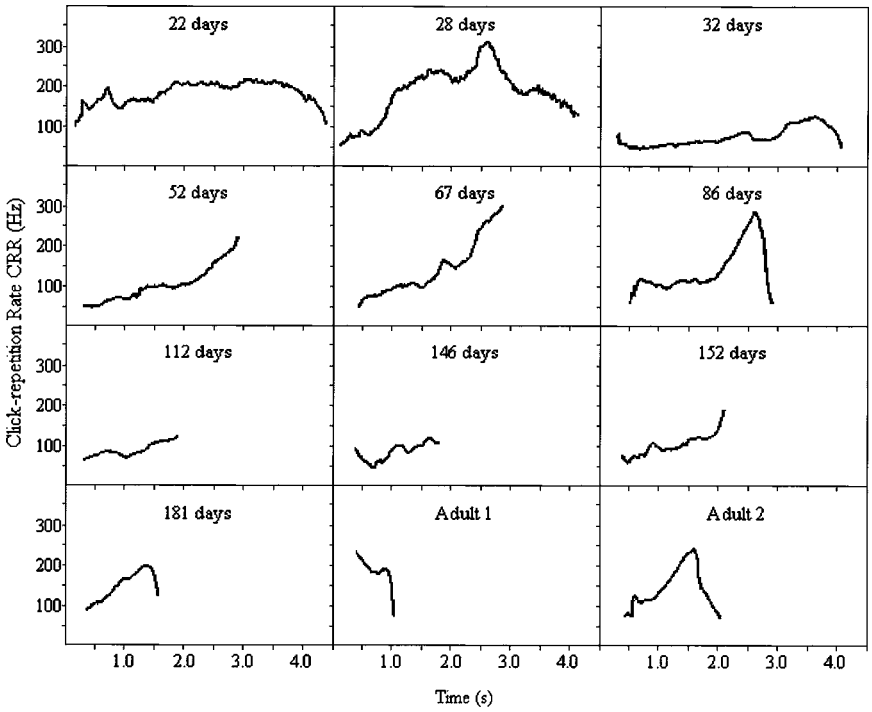


FIG. 2. Click-repetition rate (CRR) contours of click trains produced by neonatal finless porpoise ranging in age from 22 to 181 days. Data for two adults are shown for comparison. Adult 1: the mother, Adult 2: an adult female tank mate.

click trains, from one time in a 1 h recording at days 22 postnatal to at least 11 times in a 15 min recording at days 181 postnatal without the company of other individuals. By about three months, the calf was noticed scanning and toying with fish, with which the adult animals were fed. On day 98 postnatal, the neonatal porpoise was observed swallowing a fish for the first time.

In total, 76 postnatal click trains were registered at days 22 ($N=1$), 28 ($N=2$), 32 ($N=1$), 52 ($N=1$), 67 ($N=8$), 86 ($N=5$), 112 ($N=6$), 146 ($N=1$), 152 ($N=35$), and 181 ($N=16$). As shown in Fig. 2, the click trains generally decreased in duration from over 4 s at days 22 to less than 1.5 s at days from 112 through 181 postnatal. After 112 days of age, the durations of click trains resembled those of adult echolocation trains [Figs. 2 and 3(c)]. Quantitative data on CRRmax, CRRmin, click train duration, and number of clicks per train were plotted as functions of age showed in Fig. 3. The logarithmic regression lines in Fig. 3 indicated that the CRRmax presented a trend to decrease [correlation analysis showed $r=-0.42$, $p<0.001$, $N=76$; Fig. 3(a)] with

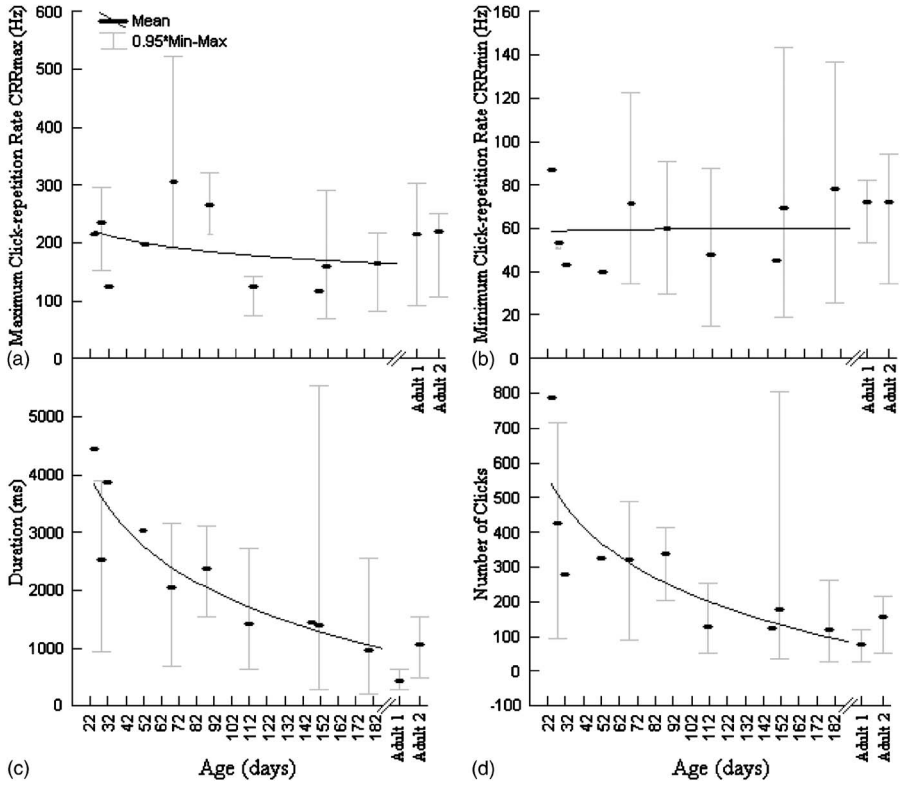


FIG. 3. The following parameters of click trains recorded from neonatal finless porpoise are plotted as a function of age: (a) maximum click-repetition rate (CRRmax); (b) minimum click-repetition rate (CRRmin); (c) duration; and (d) number of clicks. Data points represent the mean values, and bars represent the 95% of minimum-maximum range. A logarithmic regression line was fitted with each plot. Data for two adults are shown for comparison. Adult 1: the mother, Adult 2: an adult female tank mate.

age; the CRRmin trended to be quite stable [$r=0.15$, $p>0.1$; Fig. 3(b)]; the click train duration and the number of clicks gradually decreased from day 22 through day 181 [$r=-0.50$, $p<0.001$ and $r=-0.49$, $p<0.001$, respectively; Figs. 3(c) and 3(d)]. The CRRmax, duration, and number of clicks decreased most dramatically up to 112 days of age, after which time they remained relatively stable, and approximate to those of adult click trains [Figs. 3(a), 3(c), and 3(d)].

IV. DISCUSSION AND CONCLUSIONS

The data presented here are based on only one calf and a limited number of observations. However, this intriguing finding, of the similarity of the wave form and power spectrum of clicks between adults and the calf shortly after birth (i.e., within 25 days), is consistent with those findings described previously in *Phocoenoides dalli* (Bain, 1988) and *Phocoena phocoena* (Kamminga and Terry, 1994), and thus seems to show an innate, unconditioned acoustic behavior nature in neonate of the Phocoenidae; The appearance of echolocation shortly after birth (at least at days 22 postnatal), changes in echolocation components such as CRRmax, train duration, and number of clicks, increased head scanning motions, and increased frequency of hydrophone exploration, also suggest some similar behavioral patterns in a calf's initial development of echolocation abilities with other odontocetes, such as bottlenose dolphin (Lindhard, 1988; Reiss, 1988; Kamminga and Terry, 1994; Hendry, 2004), killer whale (Bowles *et al.*, 1988), sperm whales (Watkins *et al.*, 1988; Madsen *et al.*, 2003), and beluga whale (Vergara and Barrett-Lennard, 2003).

The CRRmax, train duration, and number of clicks gradually decreasing with age (Figs. 2 and 3) in neonatal finless porpoise might indicate maturational changes in sensory systems. As over time, the calf might be more familiar with the object, or be able to acquire a greater degree of skill in click production and interpretation of the returning echoes. Whereas, the change trends of train duration and number of clicks with age were contrary to those found in bottlenose dolphin (Hendry, 2004), where train duration and number of clicks increased with age. These differences might indicate an interspecific variation in the echolocation development between finless porpoise and bottlenose dolphin. Looking back at the plots in Figs. 2 and 3, there emerge to be an abrupt change in the CRR contours and acoustic parameters between 86 and 112 days of age, and after which they remained relatively stable and approximate to those of adult click trains. In addition, this period, during which a significant change happened, overlapped the time while the calf was observed scanning, manipulating, and then capturing the fish. Thus, if it was not mainly following the cue of sight, it seems possible that the calf had been capable of using his sonar skillfully to capture fish successfully. According to these recordings and observations, it can be concluded that the first 100 days after birth are essential for the echolocation and related behaviors to develop to become a crucial sensory device in finless porpoise.

It should be kept in mind that the present data are from a captive environment. It is not yet clear how the nature of the captive environment may have affected the development of echolocation and related behaviors. Complicated and specific environment in the wild, accompanied by the more frequent social contact of a larger social community, may well affect the speed at which echolocation becomes truly functional.

ACKNOWLEDGMENTS

Grateful thanks are given to the staff at Baiji Aquarium, Institute of Hydrobiology, for their support and assistance. This research was supported by grants from the Chinese Academy of Sciences (CAS, The President Fund), the Institute of Hydrobiology, CAS (220103), and Program for Promotion of Basic Research Activities for Innovative Biosciences of Japan.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).

Bain, D. (**1988**). "Sound production by a neonatal Dall's porpoise (*Phocoenoides dalli*)," in *Proceedings of the International Marine Animals Trainers' Association 16th Annual Conference* (San Antonio, Texas), pp. 62–69.

Bowles, A. E., Young, W. G., and Asper, E. D. (**1988**). "Ontogeny of stereotyped calling of a killer whale calf, *Orcinus orca*, during her first year," Rit Fiskideildar **11**, 251–275.

Carder, D. A., and Ridgway, S. H. (**1983**). "Apparent echolocation by a sixty-day-old bottlenosed dolphin, *Tursiops truncates*," J. Acoust. Soc. Am. **74**, S74.

Hendry, J. L. (2004). "The ontogeny of echolocation in the Atlantic bottlenose dolphin (*Tursiops truncatus*)," Ph.D. dissertation, the University of Southern Mississippi, MS.

Kamminga, C., and Terry, R. P. (**1994**). "Preliminary results of research on the ontogeny of the odontocete sonar signal," in *Research on Dolphin Sounds*, edited by C. Kamminga (Ph.D. dissertation, Delft University of Technology, Delft), pp. 171–184.

Li, S., Wang, K., Wang, D., and Acamatsu, T. (**2005**). "Echolocation signals of the free-ranging Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientalis*)," J. Acoust. Soc. Am. **117**, 3288–3296.

Lindhard, M. (**1988**). "Apparent sonar clicks from a captive bottlenosed dolphin, *Tursiops truncates*, when 2, 7 and 38 weeks old," in *Animal Sonar, Process and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 109–114.

Madsen, P. T., Carder, D. A., Møhl, B., and Ridgway, S. H. (**2003**). "Sound production in neonate sperm whales," J. Acoust. Soc. Am. **113**, 2988–2991.

Moreno, P., Kamminga, C., and Cohen Stuart, A. B. (**2003**). "Clicks produced by captive Amazon river dolphins (*Inia geoffrensis*) in sexual context," in *Echolocation in Bats and Dolphins*, edited by J. Thomas, C. F. Moss, and M. Vater (Univ. of Chicago, Chicago), pp. 419–425.

Reiss, D. (**1988**). "Observations on the development of echolocation in young bottlenose dolphins," in *Animal Sonar, Process and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 121–128.

Vergara, V., and Barrett-Lennard, L. G. (**2003**). "Vocal development in a captive beluga (*Delphinapterus leucas*) calf," *Fifteenth Biennial Meeting of the Society for Marine Mammalogy*, Greensboro, N.C., Dec. 14–19, **2003**.

Wang, D., Hao, Y., Wang, K., Zhao, Q., Chen, D., Wei, Z., and Zhang, X. (**2005**). "The first Yangtze finless porpoise successfully born in captivity," Environ. Sci. Pollut. Res. **5**, 247–250.

Watkins, W. A., Moore, K. E., Clark, C. W., and Dahlheim, M. E. (**1988**). "The sounds of sperm whale calves," in *Animal Sonar, Process and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 99–107.

Analytical approximations for the modal acoustic impedances of simply supported, rectangular plates

W. R. Graham[a)]
Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, England

(Received 24 August 2006; revised 4 May 2007; accepted 4 May 2007)

Coupling of the *in vacuo* modes of a fluid-loaded, vibrating structure by the resulting acoustic field, while known to be negligible for sufficiently light fluids, is still only partially understood. A particularly useful structural geometry for the study of this problem is the simply supported, rectangular flat plate, since it exhibits all the relevant physical features while still admitting an analytical description of the modes. Here the influence of the fluid can be expressed in terms of a set of doubly infinite integrals over wave number: the modal acoustic impedances. Closed-form solutions for these impedances do not exist and, while their numerical evaluation is possible, it greatly increases the computational cost of solving the coupled system of modal equations. There is thus a need for accurate analytical approximations. In this work, such approximations are sought in the limit where the modal wavelength is small in comparison with the acoustic wavelength and the plate dimensions. It is shown that contour integration techniques can be used to derive analytical formulas for this regime and that these formulas agree closely with the results of numerical evaluations. Previous approximations [Davies, J. Sound Vib. **15**(1), 107–126 (1971)] are assessed in the light of the new results and are shown to give a satisfactory description of real impedance components, but (in general) erroneous expressions for imaginary parts. [DOI: 10.1121/1.2747094]

PACS number(s): 43.20.Tb, 43.20.Rz, 43.30.Jx, 43.40.Rj [JGM] Pages: 719–730

I. INTRODUCTION

The modal approach to the analysis of vibrating structures is both effective and highly developed. It is thus natural to extend it to vibro-acoustic problems. Here, however, one encounters a complication—the fluid pressure induced by the vibrating structure provides an additional forcing, and one which couples the *in vacuo* modes (Davies, 1971a; Graham, 1995). Even if the fluid is light enough for this effect to be negligible, coupling also occurs in the expression for radiated acoustic power (Davies, 1971b; Keltie and Peng, 1987; Cunefare, 1992). Finally, when the fluid region is bounded, and hence also naturally analyzed in terms of modes, these too are coupled (Sum and Pan, 2000).

The inclusion of coupling terms greatly increases the computational demands of a modal analysis, and there is thus a strong incentive to neglect them. However, the circumstances under which this simplification is valid are still not well understood, in spite of a number of fundamental investigations involving beams (Leppington *et al.*, 1986; Keltie and Peng, 1987; Cunefare, 1992), flat plates (Davies, 1971a; Mkhitarov, 1972; Bano *et al.*, 1992; Graham, 1995), and disks (Lee and Singh, 2005). Further study of generic geometries is therefore still required.

It is tempting to restrict such study to one-dimensional geometries, because of the associated simplifications in analysis and computation. However, Graham (1995) has shown that the step from a 1D to 2D geometry has a crucial impact on the influence of modal coupling. Among 2D geometries, the simply supported flat plate is an ideal test case, because its modes are known in analytical form [see, for example, Chang and Leehey (1979)].

The effect of fluid loading is conveniently expressed in terms of acoustic modal impedances (Chang and Leehey, 1979). In cases where relatively few modes are involved, these can be computed numerically (e.g., Bano *et al.*, 1992) without the vibro-acoustic response calculation becoming excessively arduous. The inclusion of large numbers of modes, however, makes it necessary to find more efficient ways of evaluating the impedances.

The expressions to be evaluated can be written either as quadruple integrals over the plate surface or as double integrals over the (infinite) wave number domain. The first case is of a form that has recently been shown by Pierce *et al.* (2002) to be reducible to a finite summation of single integrals, each of which can straightforwardly be found via numerical quadrature. Alternatively (Snyder and Tanaka, 1995; Li, 2001), a suitable series expansion of the integrand leads to an infinite summation that is convergent over all frequencies. Nonetheless, some problems remain; Pierce *et al.* state that "numerical difficulties could arise" in their method for large plates and/or high mode numbers, while the number of terms needed for Li's summation to converge can become excessive in the same circumstances. There is thus a requirement for analytical approximations to the impedances.

Early attempts to derive such approximations were made by Wallace (1972), Davies (1971b), and Pope and Leibovitz (1974). More recently, Graham (1995) has presented a systematic asymptotic analysis for the case of an acoustically large plate. This, however, is not the only instance where useful results are available; one can also consider the regime

[a)]Electronic mail: wrg11@cam.ac.uk

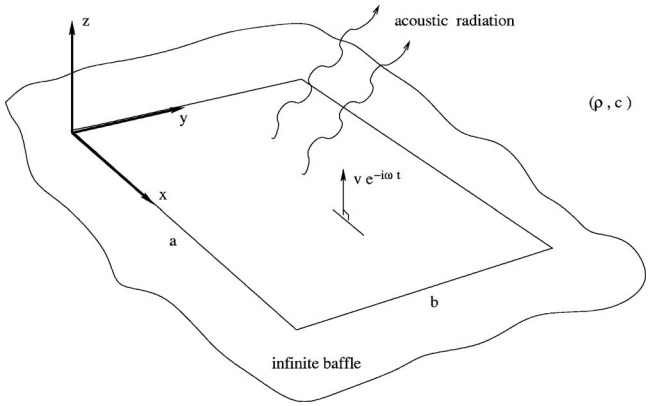


FIG. 1. The simply supported flat plate. The plate is set in an infinite, rigid baffle lying in the $x-y$ plane. A compressible fluid, with mean density ρ and sound speed c, occupies the region $z>0$.

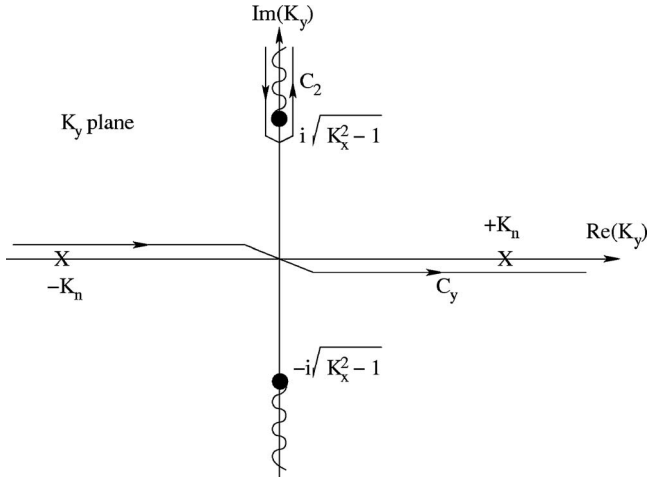


FIG. 2. The complex K_y plane. The integration contour C_y lies just above the real axis for $\Re(K_y)<0$, and just below for $\Re(K_y)>0$. The integration contour C_2 follows the upper half-plane branch cut. The poles shown at $K_y=\pm K_n$ exist only for the case $n=q$.

where the plate is large in terms of the modal wavelength, but not the acoustic wavelength. Furthermore, this case applies to lower frequencies, where fluid loading effects are more likely to be significant. The aim of the current work is therefore to derive and validate approximate expressions for the acoustic modal impedances in this parameter regime.

The paper is organized as follows. In Sec. II the double integral expression for the modal impedances is presented, and then manipulated into a form amenable to analysis by contour integration techniques. The evaluation of the first integral is discussed in Sec. III, and that of the second in Sec. IV. Finally, in Sec. V, the resulting asymptotic expressions are validated by comparison with numerical results. The paper also includes an appendix, in which recommendations for practical implementation of the approximations are given.

II. PRELIMINARY ANALYSIS

Figure 1 shows the geometry of the problem. A simply supported, rectangular, flat plate, of length a and breadth b, is set in an infinite rigid baffle and vibrates with (normal) velocity $v(x,y)e^{-i\omega t}$. As a result, it radiates acoustic waves into the compressible fluid in $z>0$, which has mean density ρ and sound speed c.

When the plate is thin enough to obey the bending wave equation, its structural modes are given by (Davies, 1971a)

$$\psi_{mn}(x,y)=\frac{2}{\sqrt{ab}}\sin\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right). \tag{1}$$

Vibration in mode (p,q), i.e., $v(x,y)=v_{pq}\psi_{pq}(x,y)$, generates a plate surface pressure field $p(x,y,0)e^{-i\omega t}$, with $p(x,y,0)=\Sigma p_{mn}\psi_{mn}(x,y)$. The modal pressures p_{mn} are linked to v_{pq} through the acoustic wave equation, and it can be shown (Chang and Leehey, 1979) that $p_{mn}=\rho c Z_{mnpq}v_{pq}$, where

$$Z_{mnpq}=0, \quad (m+p) \text{ or } (n+q) \text{ odd}; \tag{2}$$

$$Z_{mnpq}=\frac{4K_mK_nK_pK_q}{\pi^2\mu\eta}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1-(-1)^m\cos(\mu K_x)}{(K_x^2-K_m^2)(K_x^2-K_p^2)} \times\frac{1-(-1)^n\cos(\eta K_y)}{(K_y^2-K_n^2)(K_y^2-K_q^2)}\frac{dK_ydK_x}{\sqrt{1-K_x^2-K_y^2}} \quad \text{otherwise.} \tag{3}$$

Here the square root is to be taken as either positive real or positive imaginary, according to whether $K_x^2+K_y^2$ is less or greater than one. The term Z_{mnpq} is a (dimensionless) modal impedance; when $m=p$ and $n=q$ its real part is also often referred to as the "radiation efficiency" of a mode.

The "self-impedance," Z_{mnmn}, will be evaluated directly from (3). However, it must first be expressed in a form suitable for contour integration in the complex K_x and K_y planes. This is achieved by decomposing the cosine terms in the integrand into complex exponential components, and exploiting the even nature of the other terms, leading to

$$Z_{mnmn}=\frac{4K_m^2K_n^2}{\pi^2\mu\eta}\int_{C_x}\int_{C_y}\frac{1-(-1)^me^{i\mu K_x}}{(K_x^2-K_m^2)^2}\frac{1-(-1)^ne^{i\eta K_y}}{(K_y^2-K_n^2)^2} \times\frac{dK_ydK_x}{\sqrt{1-K_x^2-K_y^2}}. \tag{4}$$

Note that this process introduces singularities on the real K_x and K_y axes, at $\pm K_m$ and $\pm K_n$, respectively, and the integration paths have first been deformed onto the contour C_y shown in Fig. 2 and its K_x plane counterpart.

When $m\neq p$ and/or $n\neq q$, the evaluation of the modal impedance is simplified by expressing the components of (3) as partial fractions, yielding

$$Z_{mnpn}=\frac{K_mK_p}{K_p^2-K_m^2}[J_{pn}^x-J_{mn}^x], \tag{5}$$

$$Z_{mnpq}=\frac{K_mK_nK_pK_q}{(K_p^2-K_m^2)(K_q^2-K_n^2)}[J_{pq}^{xx}-J_{mq}^{xx}-J_{pn}^{xx}+J_{mn}^{xx}]. \tag{6}$$

Here the "singly cross partial impedance," J_{mn}^x, is given by

$$J^{x}_{mn} = \frac{4K_n^2}{\pi^2 \mu \eta} \int_{-\infty}^{\infty} \int_{C_y} \frac{1-(-1)^m e^{i\mu K_x}}{K_x^2 - K_m^2} \frac{1-(-1)^n e^{i\eta K_y}}{(K_y^2 - K_n^2)^2} \times \frac{dK_y dK_x}{\sqrt{1-K_x^2-K_y^2}}, \tag{7}$$

and its "doubly cross" counterpart, J^{xx}_{mn}, by

$$J^{xx}_{mn} = \frac{4}{\pi^2 \mu \eta} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1-(-1)^m e^{i\mu K_x}}{K_x^2 - K_m^2} \frac{1-(1)^n e^{i\eta K_y}}{K_y^2 - K_n^2} \times \frac{dK_y dK_x}{\sqrt{1-K_x^2-K_y^2}}. \tag{8}$$

Once again, the cosine terms have been decomposed into complex exponentials, and the integration contours moved off the real axis where necessary. Note also that the impedance Z_{mnmq} is straightforwardly obtained from (5), with appropriate variable exchanges.

The previous, high frequency asymptotic analysis for the modal impedance (Graham, 1995) requires $\mu,\ \eta \gg 1$ with K_m, $K_n \sim O(1)$. Here, we evaluate the integrals (4), (7), and (8) in the lower frequency region $\mu,\ \eta \sim O(1)$, subject to the condition $K_m,\ K_n \gg 1$ (i.e., high mode numbers). Note that this regime differs from that of Wallace (1972), whose approximations for the real part of (3) apply for $\mu,\ \eta \ll 1$.

III. THE INNER INTEGRAL

A. The case $n \neq q$

This case applies for the doubly cross impedance, J^{xx}_{mn}. The integral over K_y is written as

$$I_n = \int_{-\infty}^{\infty} \frac{1-(-1)^n e^{i\eta K_y}}{K_y^2 - K_n^2} \frac{dK_y}{\sqrt{1-K_x^2-K_y^2}}. \tag{9}$$

In principle, one must conduct separate analyses for $K_x > 1$ and $K_x < 1$, but, in practice, the latter case is of lesser importance. When $K_x > 1$, the branch cuts needed to ensure that the square root remains single-valued and satisfies the conditions required of it lie on the imaginary axis, starting at $K_y = \pm i\sqrt{K_x^2 - 1}$, as shown in Fig. 2. Then (Graham, 1995) the integration contour can be deformed upwards to give $I_n = I_{n1} - I_{n2}$, where

$$I_{n1}(K_x) = \frac{2i}{K_n \sqrt{K_x^2 + K_n^2 - 1}} \log\left(\frac{K_n + \sqrt{K_x^2 + K_n^2 - 1}}{\sqrt{K_x^2 - 1}} \right), \tag{10}$$

$$I_{n2}(K_x) = \int_{C_2} \frac{(-1)^n e^{i\eta K_y}}{K_y^2 - K_n^2} \frac{dK_y}{\sqrt{1-K_x^2-K_y^2}}. \tag{11}$$

On normalizing the integration variable in (11) by K_n, i.e., $K_y = i\sqrt{K_x^2 - 1} + iK_n u$, it can be seen that the dominant contribution when $K_n \gg 1$ comes from the branch point region. On the basis of Watson's lemma (Crighton *et al.*, 1992), the integral can then be estimated asymptotically by expanding the nonexponential part of the integrand as a power series in y, giving

$$I_{n2} \sim \frac{2i(-1)^n e^{-\eta\sqrt{K_x^2-1}}}{K_x^2 + K_n^2 - 1} \times \int_0^{\infty} \frac{1+O(u)}{[u(u+2\sqrt{K_x^2-1}/K_n)]^{1/2}} e^{-\eta K_n u}\, du. \tag{12}$$

Note that, although $K_n \gg 1$, the same cannot be assumed of K_x; this is why the square root term has been excluded from the expansion in u. Equation (12) can be evaluated with the help of formula 3.364(3) of Gradshteyn and Ryzhik (1994), giving

$$I_{n2} \sim \frac{2i(-1)^n}{K_x^2 + K_n^2 - 1} K_0\left(\eta\sqrt{K_x^2 - 1}\right) + e^{-\eta\sqrt{K_x^2-1}} O(K_n^{-3}). \tag{13}$$

Although (10) and (13) are the main results that will be required for I_n, the integration over K_x means that they should be applicable over the entire complex K_x plane. This can be shown by repeating the above calculations for the case $K_x < 1$. The expression for I_{n1} is found to be valid as long as $\sqrt{K_x^2 - 1} = -i\sqrt{1 - K_x^2}$ for $K_x < 1$ (Graham, 1995). The analysis for I_{n2} becomes slightly more involved, requiring a further contour deformation in the complex u plane, with eventual result

$$I_{n2} \sim -\frac{\pi(-1)^n}{K_x^2 + K_n^2 - 1} H_0^{(1)}\left(\eta\sqrt{1 - K_x^2}\right) + O(K_n^{-3}). \tag{14}$$

This is in agreement with (13) under the same condition on $\sqrt{K_x^2 - 1}$ [Abramowitz and Stegun (1970) give $2K_0(-is) = i\pi H_0^{(1)}(s)$]. Equations (10) and (13) thus provide an analytic continuation for I_n into the entire complex K_x plane, as long as branch cuts guaranteeing the required square root behavior are provided.

B. The case $n = q$

This case applies for the self- and singly cross impedances [see (4) and (7)]. The integral over K_y is now

$$I_{nn} = \int_{C_y} \frac{1-(-1)^n e^{i\eta K_y}}{(K_y^2 - K_n^2)^2} \frac{dK_y}{\sqrt{1-K_x^2-K_y^2}}. \tag{15}$$

It is evaluated in exactly the same way as I_n, giving $I_{nn} = I_{nn1} - I_{nn2}$, with

$$I_{nn1} = \frac{-i}{K_n^2 \sqrt{K_x^2 + K_n^2 - 1}} \left[\frac{\pi\eta}{2} + \frac{K_x^2 + 2K_n^2 - 1}{K_n(K_x^2 + K_n^2 - 1)} \times \log\left(\frac{K_n + \sqrt{K_x^2 + K_n^2 - 1}}{\sqrt{K_x^2 - 1}} \right) \right] + \frac{i}{K_n^2(K_x^2 + K_n^2 - 1)}, \tag{16}$$

$$I_{nn2} \sim -\frac{2i(-1)^n}{(K_x^2 + K_n^2 - 1)^2} K_0(\eta\sqrt{K_x^2 - 1}) + e^{-\eta\sqrt{K_x^2-1}} O(K_n^{-5}). \tag{17}$$

Like (10) and (13), these expressions hold throughout the K_x plane.

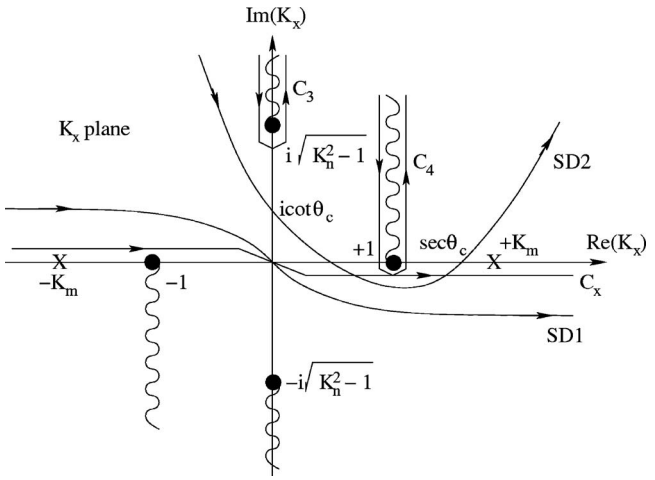


FIG. 3. The complex K_x plane. The integration path C_x is the counterpart of C_y in the K_y plane. C_3 and C_4 follow the upper half-plane branch cuts, while SD1 and SD2 are steepest descent contours. The latter detours around the imaginary axis branch cut if $K_n < \text{cosec}\ \theta_c$. Poles lie at $K_x = \pm K_m$.

IV. THE OUTER INTEGRAL

A. Branch cuts

The need for branch cuts in the expressions for I_n and I_{nn} has already been mentioned. The requirements on $\sqrt{K_x^2-1}$ (positive real or negative imaginary on the real K_x axis) are fulfilled by the cuts from $K_x=1$ to $1+i\infty$ and from $K_x=-1$ to $-1-i\infty$ shown in Fig. 3.

The formulas for I_{n1} and I_{nn1}, (10) and (16), also contain the term $\sqrt{K_x^2+K_n^2-1}$. Given the condition $K_n>1$ implicit in the asymptotic analysis, this root is always positive real on the real K_x axis. Branch cuts starting at $K_x=\pm i\sqrt{K_n^2-1}$ ensure that this requirement is maintained.

B. Contour deformations

The K_x-plane contour deformations are identical for the self-, singly cross partial, and doubly cross partial impedances. They are thus described here for the latter only. Equation (8) is first written as

$$J_{mn}^{xx} = \frac{4}{\pi^2 \mu \eta} \int_{-\infty}^{\infty} \frac{1-(-1)^m e^{i\mu K_x}}{K_x^2 - K_m^2} [I_{n1}(K_x) - I_{n2}(K_x)]\, dK_x. \tag{18}$$

When $K_n > \text{cosec}\ \theta_c$, the integration path for the term involving I_{n1} is deformed upwards onto the branch cut contours C_3 and C_4, while the term involving I_{n2} is split into two components whose paths are deformed onto the steepest descent contours $SD1$ and $SD2$. Further details, and the modified analysis necessary when $K_n < \text{cosec}\ \theta_c$, are given by Graham (1995). The final result is

$$J_{mn}^{xx} = J_{ex}^{xx} + H(K_m - \sec\ \theta_c) J_{mo}^{xx} + H(K_n - \text{cosec}\ \theta_c) J_{no}^{xx} + J_{m1}^{xx} + J_{n1}^{xx} + J_{sd}^{xx}, \tag{19}$$

where

$$J_{ex}^{xx} = \frac{4}{\pi^2 \mu \eta} \int_{C_3+C_4} \frac{I_{n1}(K_x)}{K_x^2 - K_m^2}\, dK_x, \tag{20}$$

$$J_{mo}^{xx} = \frac{4i}{\pi \mu \eta K_m} I_{n2}(K_m), \tag{21}$$

$$J_{no}^{xx} = -\frac{4}{\pi^2 \mu \eta} \int_{C_3} \frac{(-1)^m e^{i\mu K_x}}{K_x^2 - K_m^2} I_{n1}(K_x)\, dK_x, \tag{22}$$

$$J_{m1}^{xx} = -\frac{4}{\pi^2 \mu \eta} \int_{C_4} \frac{(-1)^m e^{i\mu K_x}}{K_x^2 - K_m^2} I_{n1}(K_x)\, dK_x, \tag{23}$$

$$J_{n1}^{xx} = -\frac{4}{\pi^2 \mu \eta} \int_{SD1} \frac{I_{n2}(K_x)}{K_x^2 - K_m^2}\, dK_x, \tag{24}$$

$$J_{sd}^{xx} = \frac{4}{\pi^2 \mu \eta} \int_{SD2} \frac{(-1)^m e^{i\mu K_x}}{K_x^2 - K_m^2} I_{n2}(K_x)\, dK_x. \tag{25}$$

Note that the contour $SD2$ includes a detour around the branch point $i\sqrt{K_n^2-1}$ if $K_n < \text{cosec}\ \theta_c$.

C. Evaluation of the components of J_{mn}^{xx}

1. J_{ex}^{xx}

Equation (20) can be evaluated exactly, with result

$$J_{ex}^{xx} = \frac{4}{\pi \mu \eta K_m K_n P} \left[\log\left(\frac{K_m K_n + P}{K_m K_n - P} \right) - i\pi \right]. \tag{26}$$

2. J_{mo}^{xx}

This term is given explicitly by (21). Noting that $K_o(s) \sim e^{-s}/\sqrt{s}$ for large s, one finds from (13) that J_{mo}^{xx} is exponentially small when $K_m \gg 1$.

3. J_{no}^{xx}

This term is given by the branch cut integral, (22). On writing the integration variable in the form applicable to this branch cut, i.e., $K_x = i(\sqrt{K_n^2-1}+u)$, it becomes immediately clear that J_{no}^{xx} is exponentially small when $K_n \gg 1$.

4. J_{m1}^{xx}

The nonexponential terms involving K_x in (23) vary slowly, as they are dominated by the large parameters K_m and K_n. The exponential term decays rapidly with distance away from the real axis. Watson's lemma thus applies, and the result is

$$J_{m1}^{xx} \sim \frac{8i(-1)^m e^{i\mu}}{\pi \mu^2 \eta (K_m^2 - 1) K_n^2} + O(K_{mn}^{-5}). \tag{27}$$

Note that the term (K_m^2-1) could be replaced by K_m^2 without altering the asymptotic validity of (27), but slight improvements in numerical accuracy are observed when it is retained. The same comment applies to any asymptotically inconsistent terms in other expressions.

5. J_{n1}^{xx}

This integral lies on the contour $SD1$, which is the steepest descent path for the exponential $e^{-\eta\sqrt{K_x^2-1}}$ (Graham, 1995). The saddle point is at $K_x=0$. Although the relevant

exponential does not appear explicitly in (24) and (13), it matches the large argument behavior of the modified Bessel function, and the contributions from large $|K_x|$ are therefore exponentially small. The integral thus depends on the saddle point region, where the nonexponential terms in K_x are slowly varying when K_m, $K_n \gg 1$. It is therefore asserted that they can be expanded about this point, giving

$$J_{n1}^{xx} \sim \frac{8i(-1)^n}{\pi^2 \mu \eta K_m^2 (K_n^2 - 1)} \int_{SD1} K_0(\eta \sqrt{K_x^2 - 1})\, dK_x + O(K_{mn}^{-5}). \tag{28}$$

The integration contour may now be deformed further round, to the imaginary axis, since the Bessel function remains exponentially small at large radius, except for a finite region very close to the imaginary axis where it is algebraically small (like $|K_x|^{-1/2}$). Noting that $\sqrt{K_x^2 - 1} = -i\sqrt{1+s^2}$ for $K_x = is$, one has now

$$J_{n1}^{xx} \sim \frac{8i(-1)^n}{\pi \mu \eta K_m^2 (K_n^2 - 1)} \int_0^\infty H_0^{(1)}(\eta \sqrt{1+s^2})\, ds + O(K_{mn}^{-5}). \tag{29}$$

On making the substitution $1 + s^2 = u$ and employing relation 6.592(14) of Gradshteyn and Ryzhik (1994), it is found that

$$J_{n1}^{xx} \sim \frac{8i(-1)^n e^{i\eta}}{\pi \mu \eta^2 K_m^2 (K_n^2 - 1)} + O(K_{mn}^{-5}). \tag{30}$$

Note that this component could also have been obtained by reversing the order of integration in (8), which would lead to the counterpart of J_{m1}^{xx} with x- and y-associated terms exchanged. The confirmation of that result [compare (27) and (30)] by this alternative approach justifies the claim that (28) is a valid asymptotic approximation of (24).

6. J_{sd}^{xx}

Here the integration contour is $SD2$, which is the steepest descent path for $e^{-\eta\sqrt{K_x^2-1}} e^{i\mu K_x}$ (Graham, 1995). The saddle point is at $K_x = \cos\theta_c$, after which the path enters the fourth quadrant of the K_x plane before crossing the real axis again at $K_x = \sec\theta_c$. Before the saddle point the contour crosses the imaginary axis at $K_x = i\cot\theta_c$.

Once again, it is noted that the large K_x form of the integrand in (25) matches the exponential function appropriate to $SD2$, and thus that the integral is dominated by the saddle point region. (This point holds even if $K_n < \operatorname{cosec}\theta_c$ and $SD2$ includes part of the branch cut, as long as $K_n \gg 1$.) On forming the series expansion of the slowly varying integrand terms about the saddle point, (25) becomes

$$J_{sd}^{xx} \sim \frac{4(-1)^{m+n}}{\pi \mu \eta (K_m^2 - \cos^2\theta_c)(K_n^2 - \sin^2\theta_c)} \int_{SD2} e^{i\mu K_x} H_0^{(1)}(\eta \sqrt{1 - K_x^2})\, dK_x + O(K_{mn}^{-5}), \tag{31}$$

where the alternative form for I_{n2}, (14), has been used instead of (13). The integration contour may now be deformed back onto the real axis, and the symmetry properties of the Bessel function invoked in conjunction with formula 6.677(8) of Gradshteyn and Ryzhik (1994), to give

$$J_{sd}^{xx} \sim -\frac{8i(-1)^{m+n}}{\pi \mu \eta (K_m^2 - \cos^2\theta_c)(K_n^2 - \sin^2\theta_c)} \frac{e^{i\sqrt{\mu^2+\eta^2}}}{\sqrt{\mu^2+\eta^2}} + O(K_{mn}^{-5}). \tag{32}$$

This completes the analysis for the doubly cross partial impedance. When combined as in (19), Eqs. (26), (27), (30), and (32) provide asymptotic expressions up to $O(K_{mn}^{-4})$; the leading order term is the imaginary part of (26), at $O(K_{mn}^{-3})$.

D. The components of J_{mn}^{x}

The analysis for the singly cross partial impedance differs only in minor details, so the resulting expressions are simply stated here without further explanation:

$$J_{mn}^{xx} = J_{ex}^{x} + H(K_m - \sec\theta_c) J_{mo}^{x} + H(K_n - \operatorname{cosec}\theta_c) J_{no}^{x} + J_{m1}^{x} + J_{n1}^{x} + J_{sd}^{x}, \tag{33}$$

where J_{mo}^{x} and J_{no}^{x} are exponentially small, and

$$J_{ex}^{x} = \frac{2i}{\pi \mu K_m P} \log\left(\frac{P + K_m}{P - K_m}\right) - \frac{4}{\pi \mu \eta (K_n^2 - 1) P^2} - \frac{2(K_n^2 + P^2)}{\pi \mu \eta K_m K_n P^3} \left[\log\left(\frac{K_m K_n + P}{K_m K_n - P}\right) - i\pi \right], \tag{34}$$

$$J_{m1}^{x} \sim -\frac{8i(-1)^m e^{i\mu}}{\pi \mu^2 \eta (K_m^2 - 1) K_n^2} + O(K_{mn}^{-5}), \tag{35}$$

$$J_{n1}^{x} \sim -\frac{8i(-1)^n K_n^2 e^{i\eta}}{\pi \mu \eta^2 K_m^2 (K_n^2 - 1)^2} + O(K_{mn}^{-5}), \tag{36}$$

$$J_{sd}^{x} \sim \frac{8i(-1)^{m+n} K_n^2}{\pi \mu \eta (K_m^2 - \cos^2\theta_c)(K_n^2 - \sin^2\theta_c)^2} \frac{e^{i\sqrt{\mu^2+\eta^2}}}{\sqrt{\mu^2+\eta^2}} + O(K_{mn}^{-5}). \tag{37}$$

Again, these expressions give the components of the asymptotic expansion up to $O(K_{mn}^{-4})$. The leading order term is now $O(K_{mn}^{-2})$, in (34), but the real part remains $O(K_{mn}^{-4})$.

E. The components of Z_{mnmn}

Again, the analysis essentially follows that of Sec. IV C. The results are

$$Z_{mnmn} = Z_{ex} + H(K_m - \sec\theta_c) Z_{mo} + H(K_n - \operatorname{cosec}\theta_c) Z_{no} + Z_{m1} + Z_{n1} + Z_{sd}, \tag{38}$$

with Z_{mo}, Z_{no} exponentially small, and

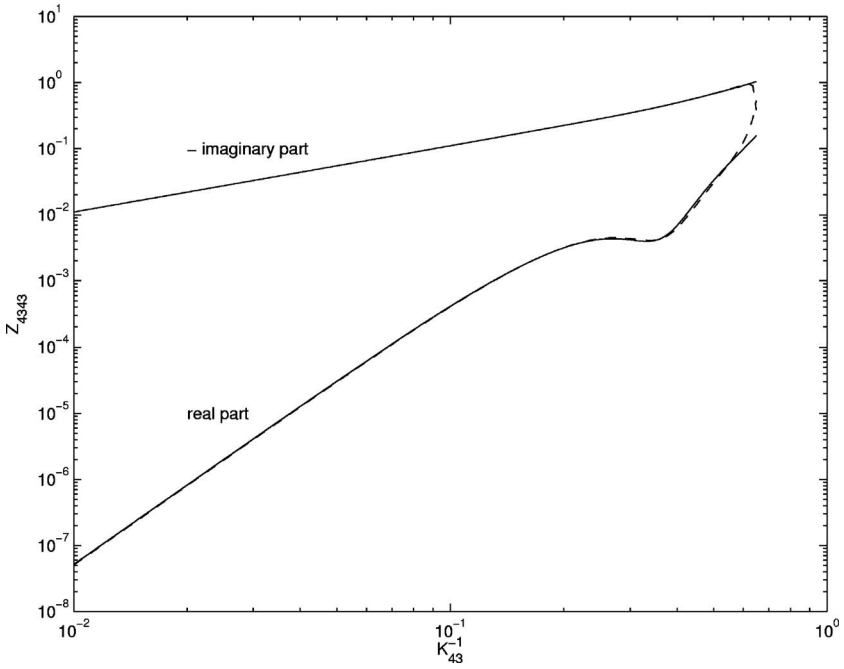


FIG. 4. Numerical (—) and asymptotic (- - -) results for the real and imaginary parts of the self-impedance Z_{4343}. Plate dimensions $a=1.5$ m, $b=1.0$ m.

$$Z_{ex} = -\frac{i}{P} - \frac{i(K_m^2+P^2)}{\pi\mu K_m P^3}\log\left(\frac{P+K_m}{P-K_m}\right) - \frac{i(K_n^2+P^2)}{\pi\eta K_n P^3}\log\left(\frac{P+K_n}{P-K_n}\right) + \frac{2i(\mu+\eta)}{\pi\,\mu\,\eta P^2} + \frac{2(2P^4+P^2-3K_m^2K_n^2)}{\pi\mu\eta(K_m^2-1)(K_n^2-1)P^4} + \frac{2P^4+P^2+3K_m^2K_n^2}{\pi\mu\eta K_m K_n P^5}\left[\log\left(\frac{K_mK_n+P}{K_mK_n-P}\right) - i\pi\right], \tag{39}$$

$$Z_{m1} \sim \frac{8i(-1)^m K_m^2 e^{i\mu}}{\pi\mu^2\eta(K_m^2-1)^2K_n^2} + O(K_{mn}^{-5}), \tag{40}$$

$$Z_{n1} \sim \frac{8i(-1)^n K_n^2 e^{i\eta}}{\pi\mu\eta^2 K_m^2(K_n^2-1)^2} + O(K_{mn}^{-5}), \tag{41}$$

$$Z_{sd} \sim -\frac{8i(-1)^{m+n}K_m^2K_n^2}{\pi\mu\eta(K_m^2-\cos^2\theta_c)^2(K_n^2-\sin^2\theta_c)^2}\frac{e^{i\sqrt{\mu^2+\eta^2}}}{\sqrt{\mu^2+\eta^2}} + O(K_{mn}^{-5}). \tag{42}$$

The largest term is imaginary and $O(K_{mn}^{-1})$, while the components of the real part are, once again, $O(K_{mn}^{-4})$.

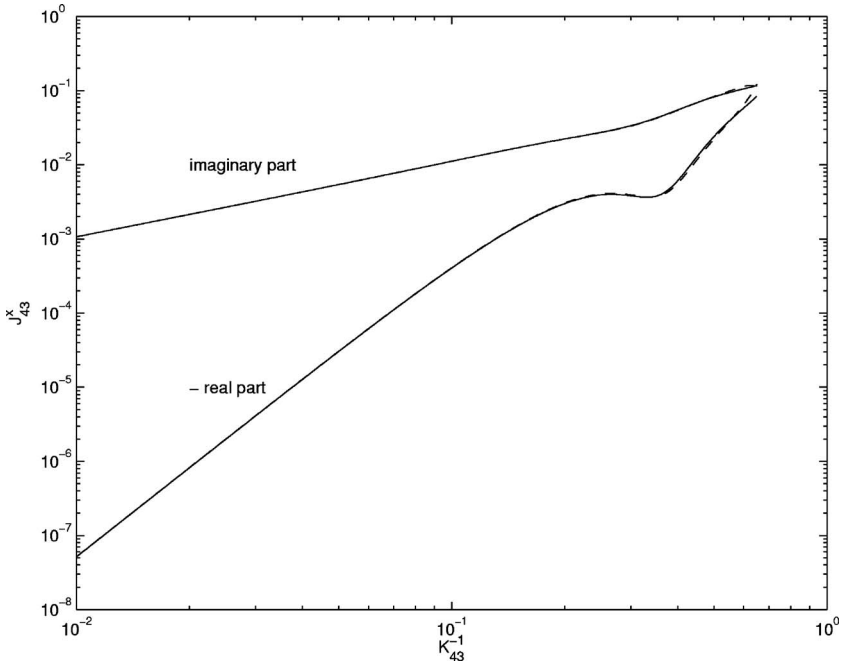


FIG. 5. Numerical (—) and asymptotic (- - -) results for the real and imaginary parts of the singly cross partial impedance J_{43}^x. Plate dimensions $a=1.5$ m, $b=1.0$ m.

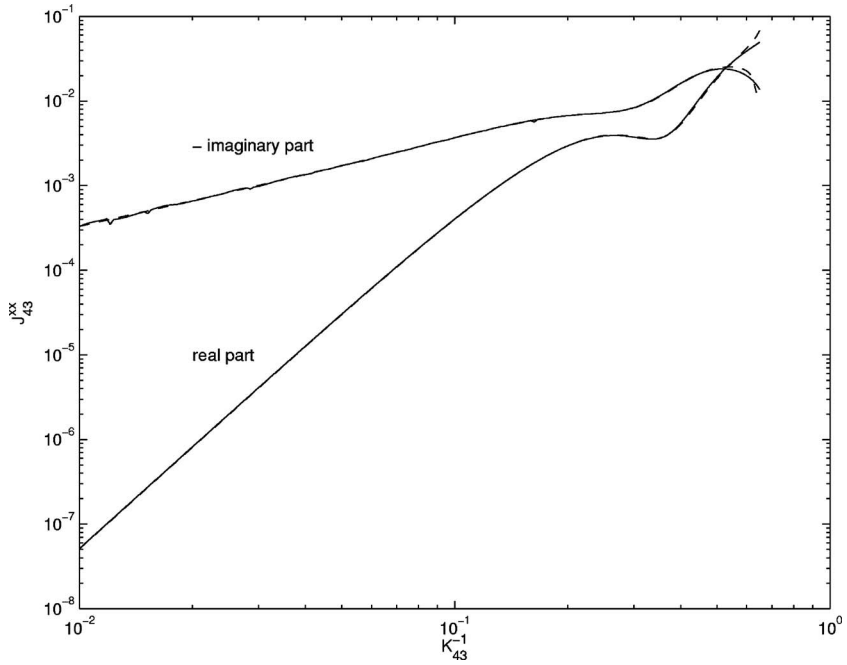


FIG. 6. Numerical (—) and asymptotic (- - -) results for the real and imaginary parts of the doubly cross partial impedance J_{43}^{xx}. Plate dimensions $a=1.5$ m, $b=1.0$ m.

F. Comparison with previous approximations

1. Davies' expressions

Davies (1971a) performed an *ad hoc* approximate analysis of the impedance integral, (3). His results, in the current notation, are

$$\Re(Z_{mnpq}) \simeq \frac{8}{\pi\mu\eta K_m K_n K_p K_q}\left[1-(-1)^m\frac{\sin\mu}{\mu} - (-1)^n\frac{\sin\eta}{\eta}+(-1)^{m+n}\frac{\sin\sqrt{\mu^2+\eta^2}}{\sqrt{\mu^2+\eta^2}}\right], \tag{43}$$

$$\Im(Z_{mnpq}) \simeq -\frac{\delta_{mp}\delta_{nq}}{K_{mn}}-\frac{4K_nK_q\delta_{mp}}{\pi\eta K_{mn}^2K_{mq}^2}-\frac{4K_mK_p\delta_{nq}}{\pi\mu K_{mn}^2K_{pn}^2}. \tag{44}$$

Noting that $K_mK_n \gg P$ for K_m, $K_n \gg 1$, and therefore that

$$\log\left(\frac{K_mK_n+P}{K_mK_n-P}\right) \sim \frac{2P}{K_mK_n}+O(K_{mn}^{-3}), \tag{45}$$

it is straightforward to show that the four terms in (43) are asymptotically equivalent to (respectively) Z_{ex}, Z_{m1}, Z_{n1}, and Z_{sd} [Eqs. (39)–(42)]. The analysis is somewhat more involved for the partial impedances, as Z_{mnpq} must be constructed according to (5) and (6), but gives the same result.

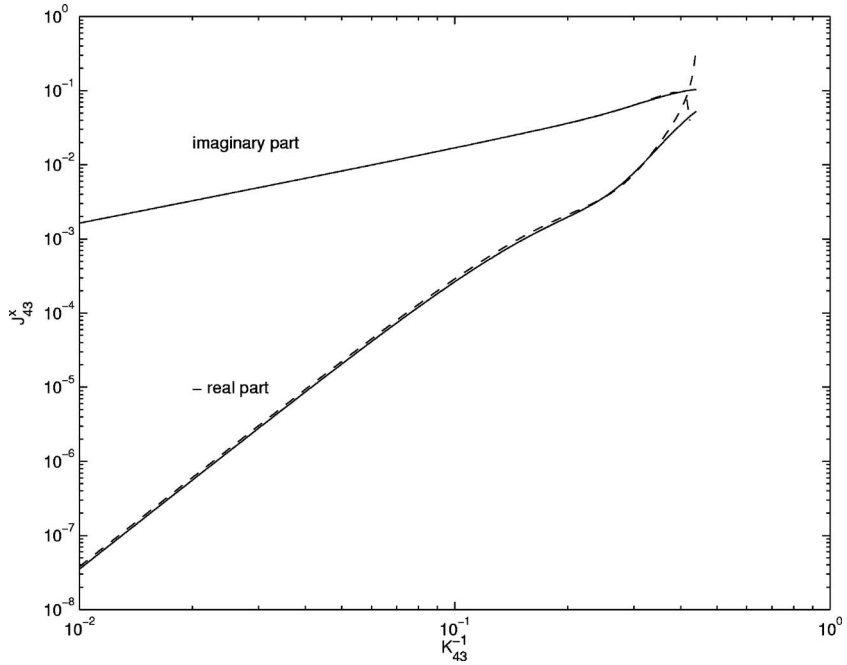


FIG. 7. Numerical (—) and asymptotic (- - -) results for the real and imaginary parts of the singly cross partial impedance J_{43}^{x}. Plate dimensions $a=1.0$ m, $b=1.5$ m.

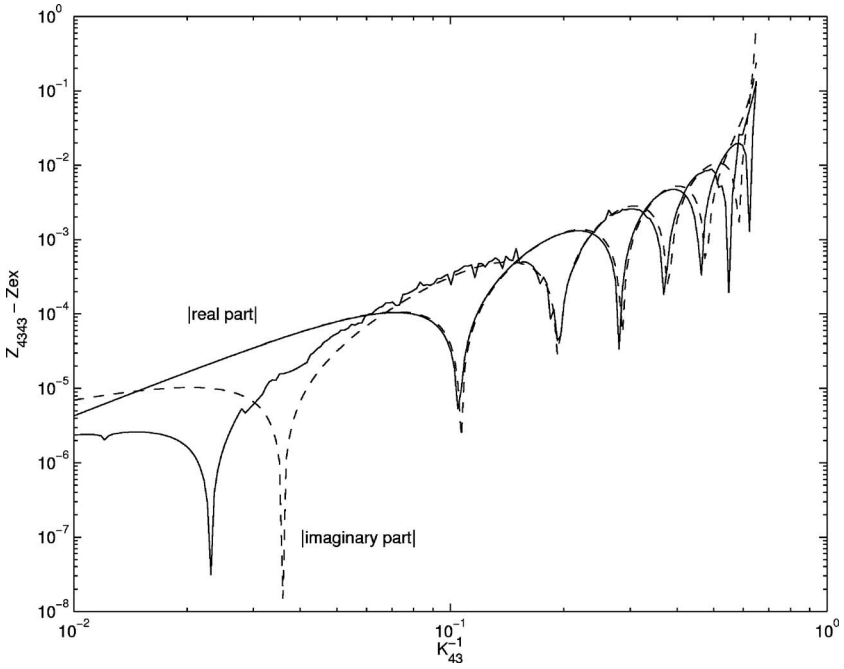


FIG. 8. Numerical (—) and asymptotic (- - -) results for the real and imaginary parts of the self-impedance component $Z_{4343}-Z_{ex}$. Plate dimensions $a=1.5$ m, $b=1.0$ m.

Davies' approximation for the imaginary part is, however, only satisfactory for the self-impedance, and then only in the form

$$\Im(Z_{mnmn}) \sim -\frac{i}{K_{mn}} + O(K_{mn}^{-2}), \tag{46}$$

which is equivalent to the leading-order contribution $-i/P$ from (39)–(42). Davies' terms at $O(K_{mn}^{-2})$ are clearly different from their counterparts in (39), which contain logarithms that cannot be expanded [since P, K_m, and K_n are all $O(K_{mn})$], and this error becomes crucial for the singly cross impedances [compare (44), (34), and (5)]. Finally, the approximation $\Im(Z_{mnpq}) \simeq 0$ when $m \neq p$, $n \neq q$ is unsatisfactory in comparison with the exact, $O(K_{mn}^{-3})$, contribution from (26).

2. Wallace's expressions

Wallace (1972) considered the real part of the self-impedance only, in a different asymptotic limit: $\mu,\ \eta \ll 1$. However, as this condition necessarily implies $K_m,\ K_n \gg 1$, it is of interest to compare the results. Wallace's expressions have recently been revisited by Li (2001), who found a correction for the case where both m and n are even. In our notation, Li gives

$$\Re(Z_{mnmn}) \sim \frac{32}{\pi\mu\eta K_m^2 K_n^2}\left[1 - \frac{\mu^2}{12}\left(1 - \frac{8}{m^2\pi^2}\right) - \frac{\eta^2}{12}\left(1 - \frac{8}{n^2\pi^2}\right)\right], \quad m,n \text{ odd}; \tag{47}$$

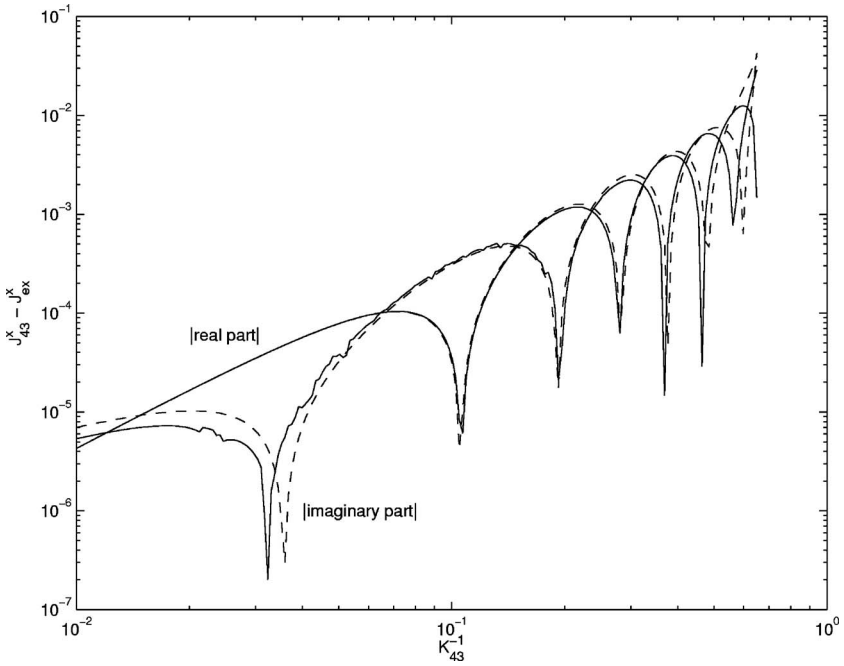


FIG. 9. Numerical (—) and asymptotic (- - -) results for the real and imaginary parts of the singly cross partial impedance component $J_{43}^x - J_{ex}^x$. Plate dimensions $a=1.5$ m, $b=1.0$ m.

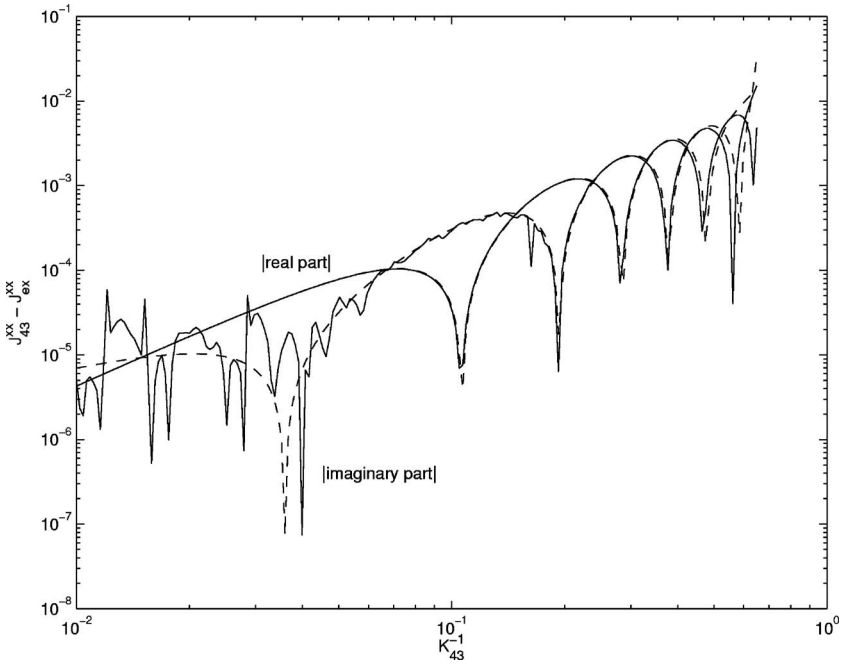


FIG. 10. Numerical (—) and asymptotic (- - -) results for the real and imaginary parts of the doubly cross partial impedance component $J^{xx}_{43}-J^{xx}_{ex}$. Plate dimensions $a=1.5$ m, $b=1.0$ m.

$$\Re(Z_{mnmn}) \sim \frac{8\eta}{3\pi\mu K_m^2 K_n^2}\left[1-\frac{\mu^2}{20}\left(1-\frac{8}{m^2\pi^2}\right) - \frac{\eta^2}{20}\left(1-\frac{24}{n^2\pi^2}\right)\right], \quad m \text{ odd, } n \text{ even}; \quad (48)$$

$$\Re(Z_{mnmn}) \sim \frac{2\mu\eta}{15\pi K_m^2 K_n^2}\left[1-\frac{\mu^2}{28}\left(1-\frac{24}{m^2\pi^2}\right) - \frac{\eta^2}{28}\left(1-\frac{24}{n^2\pi^2}\right)\right], \quad m,n \text{ even}. \quad (49)$$

Since the current expressions, (39)–(42), are equivalent, at leading order, to (43), they give

$$\Re(Z_{mnmn}) \sim \frac{8}{\pi\mu\eta K_m^2 K_n^2}[1-(-1)^m-(-1)^n+(-1)^{m+n} + O(\mu^2+\eta^2)] \quad (50)$$

as μ, $\eta \to 0$. Note that higher order terms cannot be obtained, as they depend on the unknown, $O(K_{mn}^{-5})$, coefficients in (40)–(42), which are extremely arduous to evaluate.

When m and n are odd, (50) agrees with (47) at leading order. However, if either (or both) of m, n are even, it gives zero instead of the higher-order expressions (48) and (49). This implies that practical implementations should revert to the Wallace equations for μ, $\eta \ll 1$, and this issue is considered further in the Appendix.

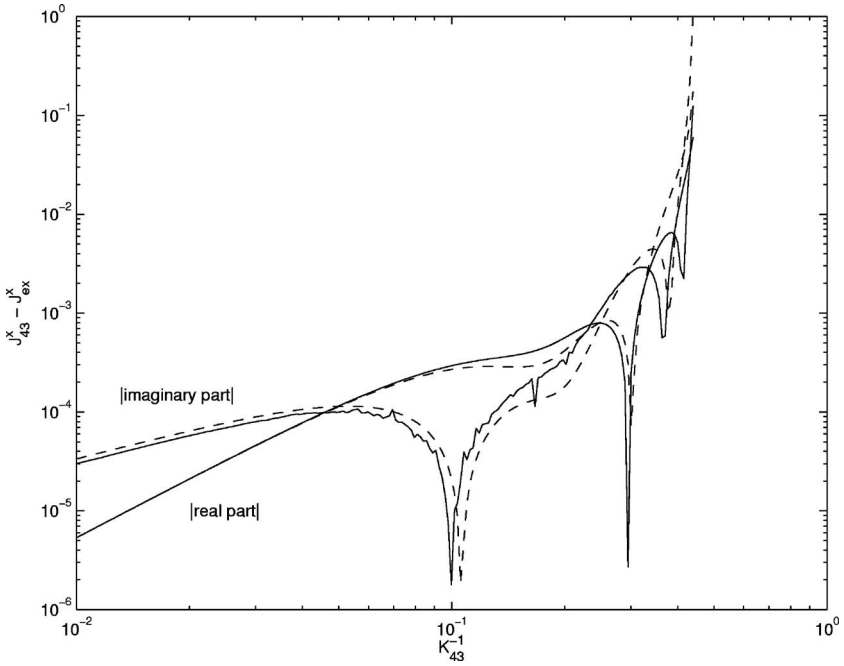


FIG. 11. Numerical (—) and asymptotic (- - -) results for the real and imaginary parts of the singly cross partial impedance component $J^{x}_{43}-J^{x}_{ex}$. Plate dimensions $a=1.0$ m, $b=1.5$ m.

V. VALIDATION OF THE ASYMPTOTIC EXPRESSIONS

A. Computational implementation

The asymptotic expressions can be straightforwardly and conveniently evaluated in the form of (19) and its equivalents, except that the exponentially small terms J^{xx}_{mo} and J^{xx}_{no} are never included. No special functions are required, unlike the high-frequency case.

Numerical evaluation of the impedance integrals proceeds first by splitting them into real and imaginary parts, which are then computed separately using automatic quadrature routines from the Numerical Algorithms Group (NAG) Fortran Library. Where infinite integration ranges are truncated, asymptotic corrections are employed to reduce the associated errors. For the partial impedances, where significant cancellation occurs, the range is split at the point where the integrand changes sign. This leads to improvements in speed and robustness, at a (potential) cost in loss of accuracy. The specified maximum error for each integration is 0.5%.

B. Results

In principle, one should test the asymptotic expressions for both $K_m < K_n$ and $K_m > K_n$, so that the behavior as both become smaller (with increasing frequency) can be observed. However, this is in fact only necessary for J^{x}_{mn}, due to the symmetry of the results for Z_{mnmn} and J^{xx}_{mn}. The main test case is thus chosen to be $m=4$, $n=3$, with $a=1.5$ m, $b=1.0$ m. For J^{x}_{mn}, a second test case is created by changing the plate dimensions to $a=1.0$ m, $b=1.5$ m.

Figure 4 shows the real and imaginary parts of the self-impedance, plotted against the dimensionless frequency K^{-1}_{mn}. Both show excellent agreement between numerical and asymptotic results up to around $K^{-1}_{mn}=0.5$, when the latter start to diverge significantly. (At this frequency, $K_m=1.33$, $K_n=1.49$.) Similar features are observed in Figs. 5 and 6, which show the corresponding results for the singly and doubly cross partial impedances, respectively.

Figure 7 shows the secondary case results for J^{x}_{43}. Here again, excellent agreement between numerics and asymptotics is evident at the lower frequencies, with divergence starting around $K^{-1}_{mn}=0.35$. At this frequency $K_m=2.56$, $K_n=1.28$.

Given the importance of the exact terms $(J^{xx}_{ex}, J^{x}_{ex}, Z_{ex})$ in the overall impedance estimates, it is also instructive to consider the remainders, i.e., $J^{xx}_{mn}-J^{xx}_{ex}$ and its equivalents. The comparisons between asymptotics and numerics for these quantities are given in Figs. 8–11. Noting the reduced levels compared to the exact terms, one concludes that good accuracy is achieved in the main case up to around $K^{-1}_{mn}=0.4$ $(K_m=1.66, K_n=1.87)$, and in the secondary case to about $K^{-1}_{mn}=0.3$ $(K_m=2.98, K_n=1.49)$. In the light of the large K_m, K_n assumption used to derive the approximations, the extent of the agreement is remarkable.

At very low frequencies, however, the imaginary asymptotic results in Figs. 8 and 9 show some inaccuracies, which are discussed further in the Appendix. Equally, the numerical results in Fig. 10 clearly exhibit the accuracy limit in the numerical calculation as do, to a lesser extent, the "glitches" in Figs. 8, 9, and 11. The level of these errors relative to the overall imaginary part results in Figs. 5–7 confirms that the loss in accuracy due to cancellation between separate numerical evaluations is not critical.

VI. CONCLUSIONS

This paper has presented asymptotic approximations for the modal acoustic impedances of a simply supported, rectangular plate. The analysis has been conducted under the assumption $K_m, K_n \gg 1$, with $\mu, \eta \sim O(1)$, and the resulting expressions have been validated against numerical solutions. This process has also shown that the asymptotic approximations remain accurate down to remarkably low (around 1.5) values of K_m and K_n, but their real parts must be replaced by alternative formulas if μ and η become small.

When compared with previous approximations, the expressions derived here are found to be asymptotically equivalent to Davies' (1971a) formula for $\Re(Z_{mnpq})$. In contrast, the same author's result for $\Im(Z_{mnpq})$ is shown to be correct only for the self-impedance, Z_{mnmn}, and then only at leading order. Finally, if one lets $\mu, \eta \to 0$ in the current expressions, one regains, to leading order, Wallace's (1972) result for $\Re(Z_{mnmn})$. Since, however, this term is only nonzero for both m and n odd, Wallace's formulas and their equivalents for the partial impedances (see the Appendix) are preferable in this regime.

NOMENCLATURE

a = plate length (x direction)
b = plate breadth (y direction)
c = speed of sound
H = Heaviside step function
$H^{(1)}_0$ = Hankel function
I_n, I_{nn} = inner integrals
J^{x}_{mn} = singly cross partial impedance (components J^{x}_{ex}, J^{x}_{mo}, J^{x}_{no}, J^{x}_{m1}, J^{x}_{n1}, J^{x}_{sd})
J^{xx}_{mn} = doubly cross partial impedance (components J^{xx}_{ex}, J^{xx}_{mo}, J^{xx}_{no}, J^{xx}_{m1}, J^{xx}_{n1}, J^{xx}_{sd})
$K_{m(p)}$ = dimensionless modal wave number $m(p)\pi/\mu$
K_{mn} = $\sqrt{K^2_m+K^2_n}$
$K_{n(q)}$ = dimensionless modal wave number $n(q)\pi/\eta$
K_x, K_y = dimensionless longitudinal and lateral wave numbers
K_0 = modified Bessel function
m, p = longitudinal mode numbers
n, q = lateral mode numbers
$p(x,y,z)$ = acoustic pressure field (harmonic component)
p_{mn} = acoustic pressure component in mode (m,n)
P = $\sqrt{K^2_m+K^2_n-1}$
t = time
$v(x,y)$ = plate normal velocity (harmonic component)
v_{mn} = plate velocity component in mode (m,n)
x, y, z = Cartesian coordinates
Z_{mnpq} = dimensionless acoustic impedance linking modes (m,n) and (p,q)
δ_{mp} = Kronecker delta symbol

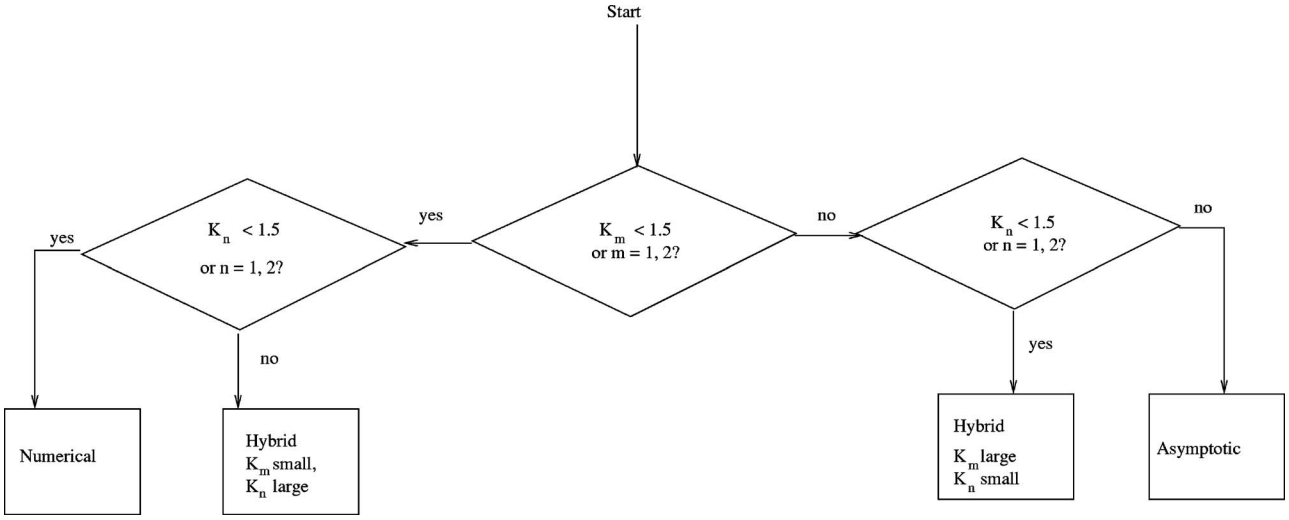


FIG. 12. Recommended evaluation methods for the modal impedances.

η = dimensionless plate breadth; $\eta=\omega b/c$
μ = dimensionless plate length; $\mu=\omega a/c$
ω = radian frequency
θ_c = $\tan^{-1}(b/a)$
ρ = fluid mean density
$\psi_{mn}(x,y)$ = shape of (m,n)th mode

APPENDIX: PRACTICAL IMPLEMENTATION OF THE EXPRESSIONS

1. Range of validity

The results presented in Sec. V B suggest that the asymptotic expressions are applicable as long as both K_m and K_n are greater than 1.5. Further testing confirms this hypothesis, with the exception of cases where either m or n is 1 or 2. Thus, in the following, $K_{m(n)}$ is deemed "small" if it is less than 1.5, or if $m(n)\leq 2$, and "large" otherwise. The asymptotic expressions are accurate as long as both K_m and K_n are large.

2. Hybrid evaluation of the impedances

In the situations where the approximations presented in this paper are useful (i.e., where many modes are significant contributors to the plate's vibration), there can be quite large numbers of modes with either K_m large, K_n small, or vice versa. The impedances for these cases are unsuitable for asymptotic approximation, but demanding to evaluate numerically, and an alternative, hybrid, approach has thus been implemented. Here the relevant impedance expression is arranged so that the inner integral corresponds to the direction in which the modal wave number is large. This integral is then approximated by the asymptotic results of Sec. III (or their equivalents, if K_m is large), and only the outer integral is estimated numerically. This approach has been validated against a fully numerical evaluation and has been found to perform entirely satisfactorily.

One thus has the option of asymptotic, hybrid, or numerical evaluation of the modal impedances. Figure 12 shows a flow chart summarizing the recommended choice, depending on the modal wave numbers.

3. Low frequencies

In Sec. IV F, it was shown that the real part of the modal self-impedance expression derived here is subject to significant cancellation when μ, $\eta \ll 1$. In this range, it is found that the leading order terms of Wallace's approximations give better agreement with numerics, and these expressions are thus used in preference. The point at which μ, η are sufficiently small to make this replacement is found to depend on whether m and n are even or odd. The recommended changeover points are

$$\mu^2+\eta^2<1.0, \quad m,n \text{ odd}; \tag{A1}$$

$$\mu^2+\eta^2<2.5, \quad \text{one of } m,n \text{ odd}; \tag{A2}$$

$$\mu^2+\eta^2<10.0, \quad m,n \text{ even}. \tag{A3}$$

The same boundaries are recommended for the partial impedances, for which the results corresponding to Wallace's are

$$\Re(J^x_{mn}) \sim -\frac{32}{\pi\mu\eta K_m^2 K_n^2}, \quad m,n \text{ odd}; \tag{A4}$$

$$\Re(J^x_{mn}) \sim -\frac{8\eta}{3\pi\mu K_m^2 K_n^2}, \quad m \text{ odd}, n \text{ even}; \tag{A5}$$

$$\Re(J^x_{mn}) \sim -\frac{8\mu}{3\pi\eta K_m^2 K_n^2}, \quad m \text{ even}, n \text{ odd}; \tag{A6}$$

$$\Re(J^x_{mn}) \sim -\frac{2\mu\eta}{15\pi K_m^2 K_n^2}, \quad m,n \text{ even}; \tag{A7}$$

$$\Re(J^{xx}_{mn}) \sim \frac{32}{\pi\mu\eta K_m^2 K_n^2}, \quad m,n \text{ odd}; \tag{A8}$$

$$\Re(J_{mn}^{xx}) \sim \frac{8\eta}{3\pi\mu K_m^2 K_n^2}, \quad m \text{ odd}, n \text{ even}; \tag{A9}$$

$$\Re(J_{mn}^{xx}) \sim \frac{8\mu}{3\pi\eta K_m^2 K_n^2}, \quad m \text{ even}, \ n \text{ odd}; \tag{A10}$$

$$\Re(J_{mn}^{xx}) \sim \frac{2\mu\eta}{15\pi K_m^2 K_n^2}, \quad m,n \text{ even}. \tag{A11}$$

The strong similarity of these results to the direct impedance expressions (47)–(49) is a reflection of Snyder and Tanaka (1995)'s finding that, in this parameter regime, the real part of any coupling impedance can be written as a weighted sum of the radiation efficiencies of the modes involved. Note, also, that the low frequency errors in the purely asymptotic component of the imaginary part (Figs. 8 and 9) could be investigated similarly. However, as these terms are much smaller than the exact component in this region, such a study is unnecessary.

Abramowitz, M., and Stegun, I. (**1970**). *Handbook of Mathematical Functions* (Dover, New York).

Bano, S., Marmey, R., Jourdan, L., and Guibergia, J.-P. (**1992**). "Etude théorique et expérimentale de la réponse vibro-acoustique d'une plaque couplée à une cavité en fluide lourd (Theoretical and experimental study of the vibro-acoustic response of a plate coupled to a cavity containing a heavy fluid)," J. Acoust. **5**, 99–124.

Chang, Y. M., and Leehey, P. (**1979**). "Acoustic impedance of rectangular panels," J. Sound Vib. **64**(2), 243–256.

Crighton, D. G., Dowling, A. P., Ffowcs Williams, J. E., Heckl, M., and Leppington, F. G. (**1992**). *Modern Methods in Analytical Acoustics: Lecture Notes* (Springer Verlag, New York).

Cunefare, K. A. (**1992**). "Effect of modal interaction on sound radiation from vibrating structures," AIAA J. **30**(12), 2819–2828.

Davies, H. G. (**1971a**). "Low frequency random excitation of water-loaded rectangular plates," J. Sound Vib. **15**(1), 107–126.

Davies, H. G. (**1971b**). "Sound from turbulent-boundary-layer-excited panels," J. Acoust. Soc. Am. **49**(3), Pt. 2, 878–889.

Gradshteyn, I. S., and Ryzhik, I. M. (**1994**). *Table of Integrals, Series, and Products* (Academic, London).

Graham, W. R. (**1995**). "High-frequency vibration and acoustic radiation of fluid-loaded plates," Philos. Trans. R. Soc. London, Ser. A **352**, 1–43.

Keltie, R. F., and Peng, H. (**1987**). "The effects of modal coupling on the acoustic power radiation from panels," ASME J. Vib., Acoust., Stress, Reliab. Des. **109**, 48–54.

Lee, H., and Singh, R. (**2005**). "Self and mutual radiation from flexural and radial modes of a thick annular disk," J. Sound Vib. **286**, 1032–1040.

Leppington, F. G., Broadbent, E. G., Heron, K. H., and Mead, S. M. (**1986**). "Resonant and non-resonant acoustic properties of elastic panels. I. The radiation problem," Proc. R. Soc. London, Ser. A **406**, 139–171.

Li, W. L. (**2001**). "An analytical solution for the self- and mutual radiation resistance of a rectangular plate," J. Sound Vib. **245**, 1–16.

Mkhitarov, R. A. (**1972**). "Interaction of the vibrational modes of a thin bounded plate in a liquid," Sov. Phys. Acoust. **18**(1), 123–126.

Pierce, A. D., Cleveland, R. O., and Zampolli, M. (**2002**). "Radiation impedance matrices for rectangular interfaces within rigid baffles: Calculation methodology and applications," J. Acoust. Soc. Am. **111**(2), 672–684.

Pope, L. D., and Leibovitz, R. C. (**1974**). "Intermodal coupling coefficients for a fluid-loaded rectangular plate," J. Acoust. Soc. Am. **56**(2), 408–415.

Snyder, S. D., and Tanaka, N. (**1995**). "Calculating total acoustic power output using modal radiation efficiencies," J. Acoust. Soc. Am. **97**(3), 1702–1709.

Sum, K. S., and Pan, J. (**2000**). "On acoustic and structural modal cross-couplings in plate-cavity systems," J. Acoust. Soc. Am. **107**(4), 2021–2038.

Wallace, C. E. (**1972**). "Radiation resistance of a rectangular panel," J. Acoust. Soc. Am. **51**(3), Pt. 2, 946–952.

Sound propagation above a porous road surface with extended reaction by boundary element method

Fabienne Anfosso-Lédée[a)]
Laboratoire Central des Ponts et Chaussées, route de Bouaye, BP 4129, 44341 Bouguenais cedex, France

Patrick Dangla
Laboratoire Central des Ponts et Chaussées, Laboratoire des Matériaux et Structures du Génie Civil, UMR 113, 2 allée Kepler, 77420 Champs sur Marne, France

Michel Bérengier
Laboratoire Central des Ponts et Chaussées, route de Bouaye, BP 4129, 44341 Bouguenais cedex, France

(Received 30 June 2006; revised 4 May 2007; accepted 5 May 2007)

Acoustic impedance of an absorbing interface is easily introduced in boundary element codes provided that a local reaction is assumed. But this assumption is not valid in the case of porous road surface. A two-domain approach was developed for the prediction of sound propagation above a porous layer that takes into account the sound propagation inside the porous material. The porous material is modeled by a homogeneous dissipative fluid medium. An alternative to this time consuming two-domain approach is proposed by using the grazing incidence approximate impedance in the traditional single-domain boundary element method (BEM). It can be checked that this value is numerically consistent with the surface impedance calculated at the interface from the pressure and surface velocity solutions of the two-domain approach. The single-domain BEM introducing this grazing incidence impedance is compared in terms of sound attenuation with analytical solutions and two-domain BEM. The comparison is also performed with the single-domain BEM using the normal incidence impedance, and reveals a much better accuracy for the prediction of sound propagation above a porous interface. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2749407]

PACS number(s): 43.28.En, 43.28.Js [KA] Pages: 731–736

I. INTRODUCTION

Boundary Element Method (BEM) has been extensively used over the past years for the prediction of environmental noise, especially for noise barrier calculation. Although it is more suitable for problems of finite size, the advantage of such a method in outdoor problems is the diversity of surface impedances and geometries that can be taken into account. Resolution techniques are well known in the case when a local type of reaction is assumed on the boundaries, i.e., the sound propagation inside the material forming the boundary can be neglected (see, for instance, many application cases in Ciskowski *et al.*, 1991). This assumption of local reaction is valid for many surfaces in road environment (grounds with earth or grass, noise barriers, dense road surfaces …), but fails in the case of porous road surfaces. Porous road pavements are extended reacting surfaces, and the boundary condition is more complex to be introduced in BEM. Watts (Watts *et al.*, 1999) used varying impedance values, separately calculated from an analytical expression requiring angles of incidence. The most important incident wave direction had to be assumed. A more robust approach is the multi-domain approach, in which the sound propagation in the air space and inside the porous medium is described by two integral equations, coupled by the expression of continuity of the sound pressure and particle velocity at the interface. Seybert (Seybert *et al.*, 1990), and Cheng (Cheng *et al.*, 1991) described the procedure for lossless media, and Wu (Wu *et al.*, 2003) studied the effect of bulk-reacting materials in mufflers with the same approach. Sarradj (Sarradj, 2003) used the multi-domain approach for the prediction of outdoor sound propagation, specifically to introduce porous road surfaces. He compared his results with a traditional single-domain approach with local reaction assumption. The author concluded that a significant difference exists when the absorption coefficient becomes high at medium and high frequencies (above 600 Hz), especially for higher source positions. The observed discrepancy between the two models can reach 5–10 dB at some frequencies.

In this paper, a similar two-domain approach is used for the prediction of sound propagation above a porous road surface with extended reaction. Although BEM is well suited to complex geometries, the basic case of sound propagation above a plane porous surface is first investigated. An alternative to this time consuming two-domain approach is proposed by using the grazing incidence approximate impedance in the traditional single-domain BEM. The results of both BEM approaches are compared with analytical solutions using sound ray theory, and to experimental results from the literature.

[a)]Author to whom correspondence should be addressed. Electronic mail: fabienne.anfosso@lcpc.fr

II. FORMULATION

A. Traditional two-dimensional BEM formulation for boundaries with local reaction

The sound propagation above the ground surface **S** is formulated in terms of an integral equation (assuming a $e^{-i\omega t}$ time dependence):

$$\varepsilon(M)p(M) + \int_S \left[\frac{\partial G_0(M,M_s)}{\partial n_s} - ik_0 \frac{Z_0}{Z_n} G_0(M,M_s) \right] \times p(M_s) dS(M_s) = p_{\text{inc}}(M), \quad (1)$$

where $p(M)$ is the acoustic pressure at a point M in the air space, n_s is the outward normal of the surface **S**, k_0 and Z_0 are, respectively, the wave number and the impedance of air, Z_n is the surface normal impedance, and p_{inc} is the incident sound pressure in free field. The problem can be three-dimensional, but the present study is restricted to a two-dimensional approach, as it is intended to be included in a general problem of large dimensions. In this case, the Green function G_0 is defined by

$$G_0(M,M') = \frac{i}{4} H_0^{(1)}(k_0 r), \quad (2)$$

where r is the distance between points M and M', and $H_0^{(1)}$ the Hankel function of first kind of order zero. More advanced Green functions have been proposed (Chandler-Wilde and Hothersall, 1995), in order to simplify the integral Eq. (1) for particular boundary conditions. But they are not adapted to the present case.

Furthermore, the present study will focus on the simple case of a flat, homogenous, and infinite boundary **S**, although it can be of any shape and made of various materials with different impedances. In these conditions, $\varepsilon(M)=1$ for $M \in \mathbf{V}$ and $\varepsilon(M)=0.5$ for $M \in \mathbf{S}$.

Usually, a local reaction is assumed on the boundary: the motion at point M_S in the direction normal to the surface depends on the acoustic pressure only at that point, and is independent of the motion of any other point of the surface (Morse *et al.*, 1968). In this case, at each surface point M_S, the surface impedance value Z_n only depends on the sound pressure at that point, $p(M_S)$. Analytical expressions of this surface impedance can be made. Thus for a layer of absorbing medium of characteristic impedance Z_p, wave number k_p and thickness e

$$Z_n = Z_p \coth(-ik_p e). \quad (3)$$

The integral Eq. (1) is traditionally solved numerically, leading to the resolution of a system of linear equations:

$$[\mathbf{H}]\{p\} = \{p_{\text{inc}}\}, \quad (4)$$

where $\{p\}$ is the vector of unknown nodal pressures, and $\{p_{\text{inc}}\}$ the nodal values of incident pressure. The details of this resolution can be found in (Anfosso-Lédée and Dangla, 2006).

B. BEM formulation for boundaries with extended reaction

1. Two-domain BEM approach

However, in the case of material with a rigid frame like porous road pavements, the assumption of local reaction is not valid (Sarradj, 2003). In other words, at each point M_S on the surface, the surface impedance Z_n depends on the sound pressure not only at M_S but also at other points of the boundary. For an absorbing layer of thickness e on a hard backing, the analytical expression of the surface impedance is (Li *et al.*, 1998, Bérengier *et al.*, 1997, Allard *et al.*, 2003):

$$Z_n = \frac{Z_p}{\chi} \coth(-ik_p e\chi), \quad (5)$$

where χ can be derived from the angle of incidence ϕ of the plane wave

$$\chi = \sqrt{1 - \left(\frac{k_0}{k_p}\right)^2 \sin^2 \phi}. \quad (6)$$

This geometrical approximation introducing plane waves and geometrical quantities such as angles of incidence is not suited to BEM formulation. Furthermore, the interaction between Z_n at point M_S and the sound pressure at other points of the boundary cannot be easily introduced in the integral equation. Most authors use the normal incidence approximation, corresponding to $\sin \phi = 0$, i.e., $\chi = 1$, and leading to an expression of surface impedance equivalent to the local reaction approximation (3).

A more accurate approach using the coupling of two propagation domains can be used and has been described in Seybert *et al.*, 1990 and Sarradj, 2003. In this approach, the sound propagation inside the porous medium is described by a boundary integral equation with unknowns p and $\partial p/\partial n$. It is coupled with the boundary integral equation of the sound propagation in the air, by the continuity of pressure and pressure gradient at the interface between both media. The numerical resolution by boundary elements leads to the set of linear equations in which the unknowns are the nodal values of sound pressure $\{p\}$ and pressure gradient $\{\partial p/\partial n\}$

$$\begin{cases} [\mathbf{H}]\{p\} - [\mathbf{G}]\left\{\dfrac{\partial p}{\partial n}\right\} = \{\mathbf{p_{inc}}\} \\ [\mathbf{H_p}]\{p\} - [\mathbf{G_p}]\left\{\dfrac{\partial p}{\partial n}\right\} = \{0\} \end{cases}. \quad (7)$$

The matrices $[\mathbf{H}]$ and $[\mathbf{G}]$ can be calculated from the Green function in the air G_0 and its normal gradient. The matrices $[\mathbf{H_p}]$ and $[\mathbf{G_p}]$ can be calculated from the Green function in the porous medium G_p and its normal gradient

$$G_p(M,M') = \frac{i}{4} H_0^{(1)}(k_p r), \quad (8)$$

where k_p is the complex wave number in the porous medium, and r the distance between points M and M'.

In the case of a layer of porous material bounded by an impervious substructure, typical of a porous road surface

layer, the Green function that integrates in its definition a homogeneous von Neumann condition on the sublayer is used

$$G_p(M,M') = \frac{i}{4}[H_0^{(1)}(k_p r) + H_0^{(1)}(k_p r')], \quad (9)$$

where r' is the distance between M and M'' the symmetrical point of M' with respect to the lower interface.

2. Model for sound propagation inside the porous medium

For the description of sound propagation inside a porous pavement, a rigid-frame model is used, assuming that the pore walls are nondeforming and motionless. This can be justified by the nature of the acoustic excitation, and by the much higher density and stiffness of the porous pavement material than the air, and is supported by many authors (Attenborough and Howorth, 1990; Hamet and Bérengier, 1993; Allard, 1993; Bérengier *et al.* 1997; Watts *et al.*, 1999; Lui and Li, 2004). A phenomenological model is used where the porous medium is considered as a homogeneous fluid, isotropic and dissipative due to viscosity and thermal exchanges between the air and the solid frame. The porous medium is fully described by the complex characteristic impedance Z_p and the complex wave number k_p as defined in Morse and Ingard, 1968, and later in Hamet and Bérengier, 1993

$$Z_p = Z_0 \sqrt{\frac{K}{\gamma}} \frac{1}{\Omega} \left(1 + i\frac{f_\mu}{f}\right)^{1/2} \left(1 - \frac{1 - 1/\gamma}{1 + if_\theta/f}\right)^{-1/2}, \quad (10)$$

$$k_p = k_0 \sqrt{K\gamma} \left(1 + i\frac{f_\mu}{f}\right)^{1/2} \left(1 - (1 - 1/\gamma)\frac{1}{1 + if_\theta/f}\right)^{1/2}, \quad (11)$$

where f_μ and f_θ are defined by

$$f_\theta = \frac{R_s}{2\pi\rho_0 P_r} \quad \text{and} \quad f_\mu = \frac{R_s \Omega}{2\pi\rho_0 K}. \quad (12)$$

Ω is the open porosity, R_s the air flow resistivity, K the tortuosity, γ is the specific heat ratio (1.4 in the case of an ideal gas) and Pr the Prandtl number (0.71 in the air).

III. GRAZING INCIDENCE APPROXIMATION FOR SURFACE IMPEDANCE

A. Surface grazing impedance

Obviously, the set of matrix Eqs. (7) in the two-domain approach requires longer computing time and more resources than the set of Eq. (4) in the traditional single-domain BEM, because it contains twice as many equations and unknowns. This is not suitable for parametric studies or for large size problems. An alternative is used by introducing the approximate surface impedance at totally grazing incidence. It derives from Eqs. (5) and (6) with an angle of incidence ϕ close to $\pi/2$, and thus

$$Z_{ap} = \frac{k_p Z_p}{k_0 \sqrt{\left(\frac{k_p}{k_0}\right)^2 - 1}} \coth\left[-ik_0 e \sqrt{\left(\frac{k_p}{k_0}\right)^2 - 1}\right]. \quad (13)$$

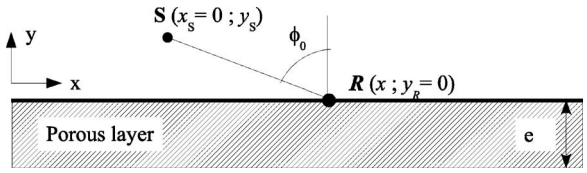


FIG. 1. Geometry of the interface for the investigation of surface impedance.

The consistency of this approximation is checked by comparing with the evaluation of the surface impedance derived from the two-domain BEM predictions. Actually, the resolution of Eq. (7) gives access not only to the sound pressure at the interface, but also to the normal pressure gradient. Thus, the surface impedance at the interface—i.e., the ratio Z_n/Z_0—can be calculated.

In the simulated case, a line source is located at x_S=0 above a porous layer characterized by a flow resistivity R_S =20 000 N m s^{-4}, a porosity Ω=0.15, a tortuosity K=4, and a layer thickness e=0.038 m (Fig. 1). Two different source heights are considered: y_S=0.3 m for a rather oblique inci-

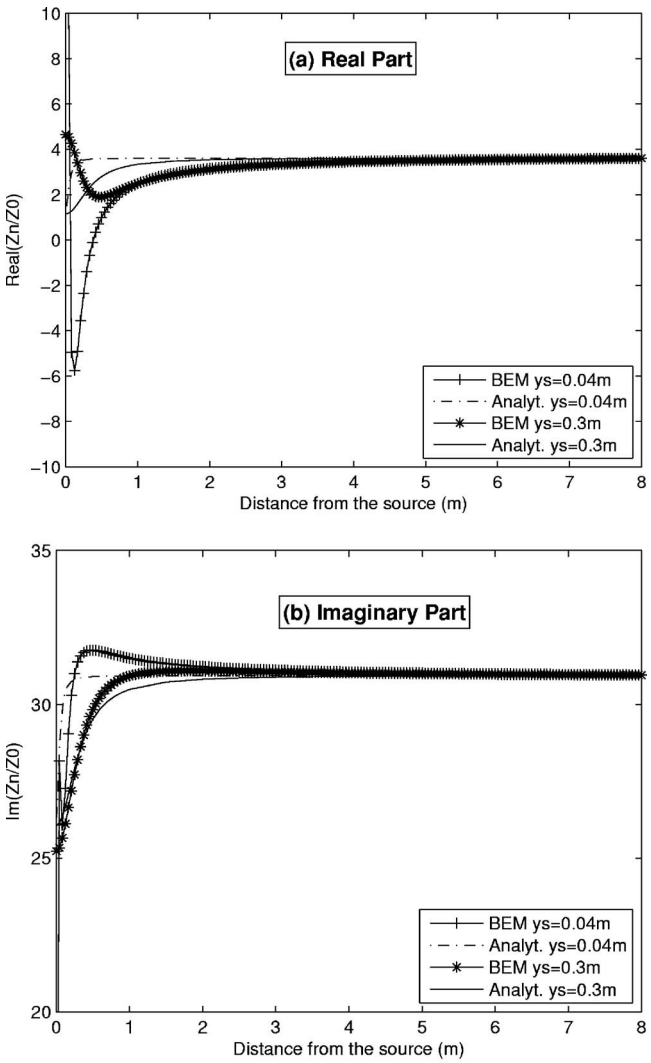


FIG. 2. Surface impedance along the interface at 250 Hz, predicted by BEM and analytical approximation for two source height (y_s=0.04 and 0.3 m); (a) real part, (b) imaginary part.

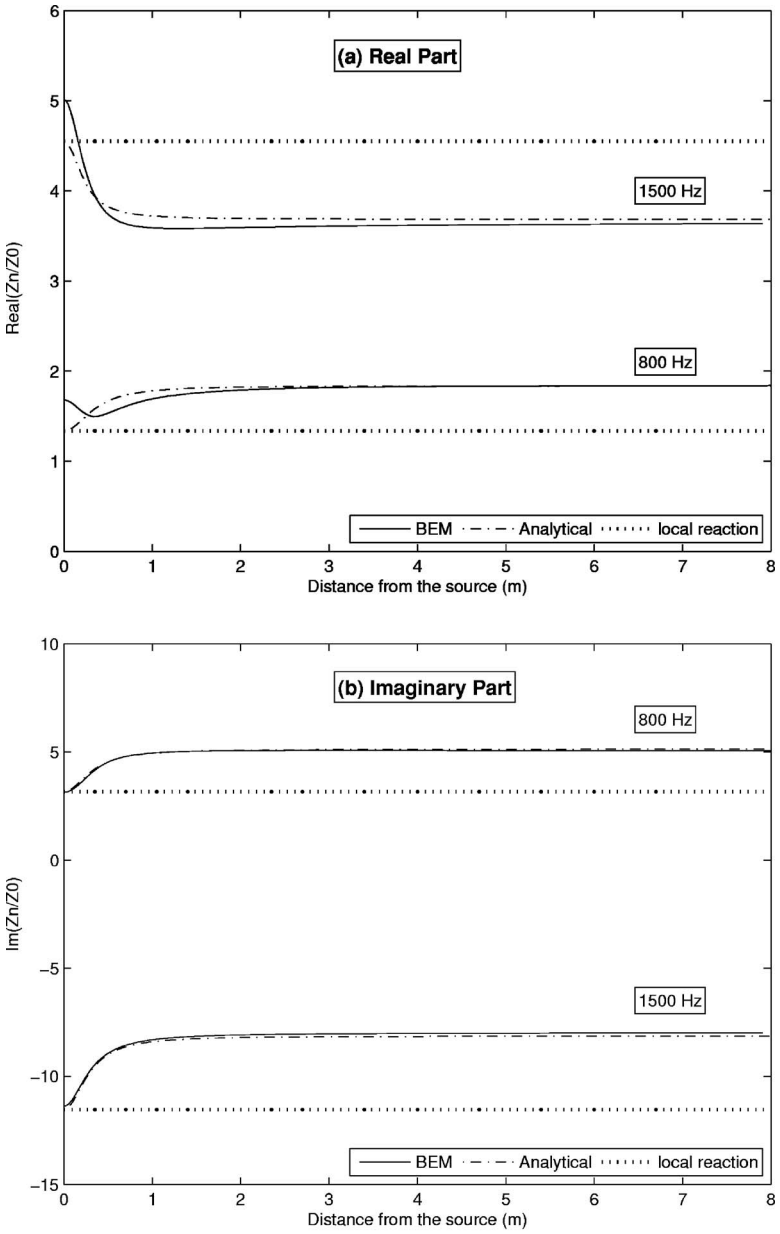


FIG. 3. Surface impedance along the interface at 800 and 1500 Hz, predicted by BEM and analytical approximation for source height H_s=0.3 m; (a) real part, (b) imaginary part.

dence, and y_S=0.04 m for a more grazing incidence. The surface impedance is calculated at points of abscissa x in the first 8 m away from the source, but the total interface modeled extends up to $x_{\max}$=10 m. The numerical values of the ratio Z_n/Z_0 at 250 Hz are presented in Figs. 2(a) and 2(b), respectively, for the real and imaginary parts, as a function of the distance x along the porous interface. The two-domain BEM predictions—in which no explicit boundary condition is introduced at the interface—are compared with the analytical solution of Eq. (5). The quantity χ defined in Eq. (6) is

$$\chi = \sqrt{1 - \left(\frac{k_0}{k_p}\right)^2 \left(\frac{x^2}{x^2 + y_S^2}\right)}. \tag{14}$$

It can be seen in Fig. 2 that far enough from the source, i.e., at a distance of about two wavelengths from the source, analytical and BEM models converge to the grazing incidence impedance. Very close to the source, discrepancies between analytical and BEM predictions are more important, especially in the case of the lower source (y_S=0.04 m). The failure of both models may explain this difference: the lowest source is $\lambda/34$ from the surface, the highest one $\lambda/4.5$. Very close to the source, BEM calculations fail due to the indetermination of the Green function at the source point, and the plane wave approximation in the analytical model is not fully valid.

Additional calculations of surface impedances for the highest source (y_S=0.3 m) are presented in Fig. 3 at higher frequencies: 800 and 1500 Hz. At these frequencies, the source zone is smaller, and again, analytical and numerical results converge rapidly to the same approximate value: the grazing incidence impedance as defined in Eq. (13). It is independent of the nature of the incident wave, and can be introduced as a constant impedance value in a traditional single-domain BEM model, in the same way as impedance with local reaction. The surface impedance with local reac-

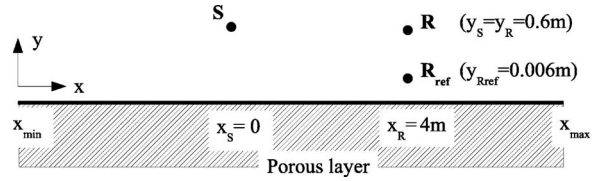


FIG. 4. Geometry of the simulated propagation case.

tion approximation is also plotted in Fig. 3. It is equivalent to the analytical extended reaction impedance calculated at normal incidence ($x=0$)

$$Z_{n0} = Z_p \coth(-ik_p e). \tag{15}$$

It can be seen that the grazing approximate value for surface impedance is a more relevant approximation for extended reaction surfaces, and that the error made on the surface impedance evaluation by considering a local reaction can be important, especially at low frequencies.

B. Effect of the approximation with grazing impedance on the prediction of sound propagation

The grazing surface impedance was introduced in the BEM code, and the effect on sound propagation predictions was investigated. In the first validation case the traditional single-domain BEM with grazing impedance was compared with the two-domain approach, and with analytical and experimental results published in Carpinello *et al.*, 2004. A sound source is located at $x_S=0$ m and $y_S=0.6$ m above a flat interface between air and a porous asphalt layer. Two receivers are located at $x_R=4$ m distance (Fig. 4). The first receiver **R** is 0.6 m high, the second one $\mathbf{R_{ref}}$ is 0.006 m high, standing for a reference microphone laying on the surface. The porous asphalt is characterized by a flow resistivity $R_S=5000$ N m s^{-4}, a porosity $\Omega=0.20$, a tortuosity $K=4.3$, and a layer thickness $e=0.04$ m. The analytical predictions were three dimensional, using sound ray and image source theory. An extended reaction was also considered at the interface, and the spherical reflection coefficient of the pavement was introduced. The characteristic impedance and wave number of the porous asphalt were described with the same phenomenological model as in BEM approach. The experimental results were obtained using an impulse technique. All results are expressed in terms of level differences between the two receiver positions: $\Delta L = L(\mathbf{R}) - L(\mathbf{R_{ref}})$. The BEM results are compared in Fig. 5 with analytical predictions and experimental results. A perfect agreement can be observed on the whole frequency range [100 Hz–4 kHz] between two-domain BEM predictions and single-domain BEM predictions with grazing incidence. Both correlate fairly well with analytical and experimental results. A small difference between BEM models and analytical model appears at the first interference gap that can be explained by a lack of accuracy for very low sound pressure levels, or a different frequency sampling in the two models. All predictions show a very good agreement with experimental results.

It is remarkable that two-dimensional BEM predictions and three-dimensional analytical predictions and experiments are in a good agreement when comparisons are made in terms of sound pressure level differences. However, when comparing absolute sound pressure levels, the difference of geometrical spreading (spherical for a point source in three dimensions and cylindrical for a line source in two dimensions) must be taken into account.

In the second simulated case, the comparison is made with another BEM approach in (Sarradj, 2003). Sarradj compared the sound pressure level of a single-domain BEM with a local reaction model and the two-domain approach, for a propagation case above a porous absorber strip embedded in a rigid baffle. The line source is on the median plane of a 10-m-wide baffle, at different heights above it. The receiver is located at 7.5 m distance, 1.2 m above the rigid baffle (Fig. 6). The porous road surface is characterized by a flow resistivity $R_S=8000$ N m s^{-4}, a porosity $\Omega=0.24$, and a tortuosity $K=5$, and the layer thickness is $e=0.03$ m. The author compared a two-domain BEM similar to the one described above and a traditional single-domain BEM, using local reaction surface impedance, i.e., Z_{n0} as defined in Eq. (15). The same calculations have been repeated for two

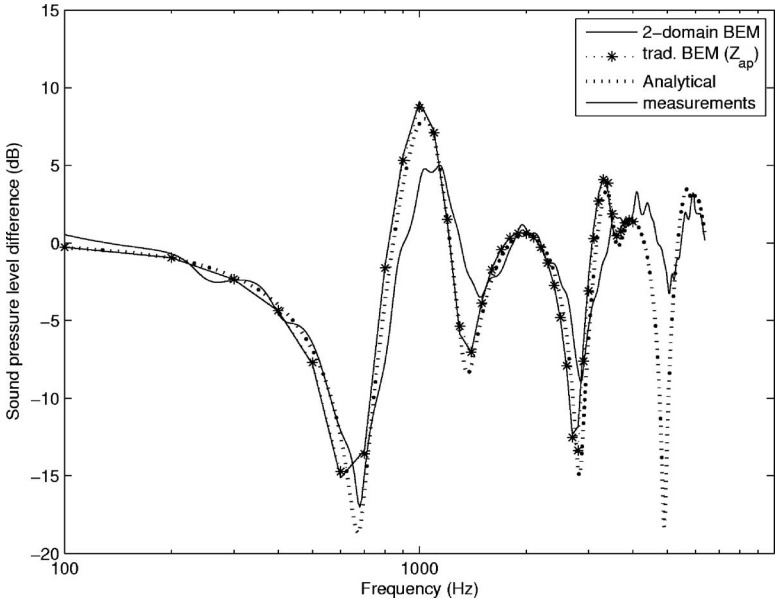


FIG. 5. Sound level difference ($L_R - L_{\mathrm{ref}}$) above a porous pavement predicted by two-domain BEM, single-domain BEM with grazing impedance, and compared with analytical and experimental results in (Carpinello *et al.*, 2004).

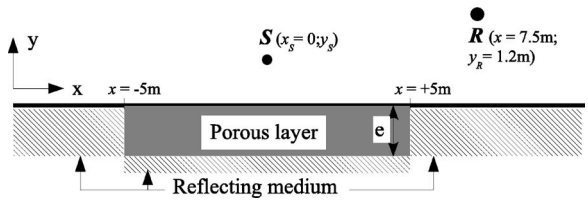


FIG. 6. Description of the modeled propagation case (identical to Sarradj, 2003).

source heights, y_S=0.05 m and y_S=0.2 m. In the two-domain approach, the lateral limits of the porous layer, at the junction between the porous medium and the rigid baffle, must be considered and discretized. In the single-domain approach, this junction is only modeled as a surface impedance jump. The calculations were also compared with the single-domain approach using the approximate surface impedance Z_{ap} derived in Eq. (13). The results are presented in Fig. 7 in terms of sound pressure level difference between two-domain and traditional single-domain BEM. The curves for the cases considering surface impedance Z_{n0} (local reaction assumption), are very similar to the one in Sarradj, 2003, Fig. 9: the difference between the two BEM approaches is negligible at low frequencies but is important at higher frequencies when sound absorption in the porous medium is significant. As observed in Sarradj, 2003, the difference is more important for a higher source position, i.e., for larger grazing incidence. However, it is remarkable in Fig. 7, that the use of the approximate impedance Z_{ap} in the traditional BEM approach is a much more correct approximation than using the normal incidence one Z_{n0}: the difference with the two-domain BEM is drastically reduced on the whole frequency range. This approximate impedance Z_{ap} is thus an efficient parameter to be introduced in the BEM model for the surface impedance effect of porous interfaces, requiring much less computation capacities than the two-domain approach.

IV. CONCLUSIONS

Sound propagation above a porous road surface can be modeled accurately by a two-domain BEM, based on the resolution of two coupled integral equations describing sound propagation, respectively, in air and inside porous medium. But an alternative to this time consuming resolution scheme was proposed by using an approximate surface impedance, constant along the interface, that can be introduced in a traditional single-domain BEM. This approximate value is the grazing incidence impedance, and it was shown that it is consistent with the exact surface impedance calculated from the two-domain BEM results. The use of this approximate surface impedance provides a smart way to introduce extended reaction-type of surface impedances in a classical BEM, without additional numerical cost, but with a better accuracy than assuming a local reaction. Thus porous interfaces can be introduced in a more general acoustic problem, with mixed impedance properties, for instance, to predict the effect of porous surfaces on tire horn effect reduction, or when used in conjunction with noise barriers.

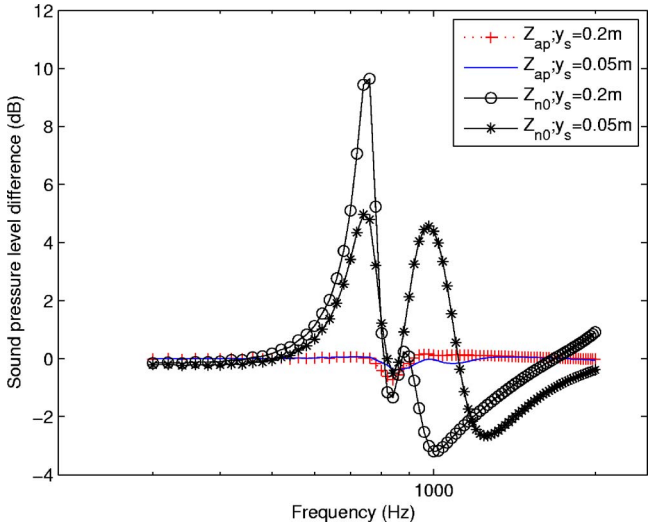


FIG. 7. (Color online) Sound level difference between two-domain BEM and traditional single-domain BEM simulations using approximate impedance Z_{ap} or normal incidence impedance Z_{n0}; two source heights (H_s=0.2 and 0.05 m).

Allard, J. F. (**1993**). *Propagation of Sound in Porous Media, Modeling Sound Absorbing Materials* (Chapman and Hall, London).

Allard, J. F., Gareton, V., Henry, M., Jansens, G., and Lauriks, W. (**2003**). "Impedance evaluation from pressure measurements near grazing incidence for non-locally reacting porous layers," Acust. Acta Acust. **89**, 595–603.

Anfosso-Lédée, F., and Dangla, P. (**2006**). "Sound propagation above a porous road surface by Boundary Element Method," J. Road Mater. Pavement Des. **7**(3), 289–312.

Attenborough, K., and Howorth, C. (**1990**). "Models for the acoustic characteristics of porous road surfaces," *International Tire/Road Noise Conference*, Goteborg, Sweden, 177–191.

Bérengier, M. C., Stinson, M. R., Daigle, G. A., and Hamet, J. F. (**1997**). "Porous road pavements: Acoustical characterization and propagation effects," J. Acoust. Soc. Am. **101**(1), 155–162.

Carpinello, S., L'Hermite, P., Bérengier, M., and Licitra, G. (**2004**). "A new method to measure the acoustic surface impedance outdoors," Radiat. Prot. Dosim. **111**(4), 363–367.

Chandler-Wilde, S. N., and Hothersall D. C. (**1995**). "Efficient calculation of the Green's function for acoustic propagation above a homogeneous impedance plane," J. Sound Vib. **180**, 705–724.

Cheng, C., Seybert, A., and Wu, T. (**1991**). "A multi-domain boundary element solution for silence and muffler performance prediction," J. Sound Vib. **151**, 119–129.

Ciskowski, R. D., and Brebbia, C. A. (**1991**). *Boundary Element Methods in Acoustics* (Computational Mechanics, and Elsevier Applied Science).

Hamet, J. F., and Bérengier, M. (**1993**). "Acoustical characteristics of porous pavements: A new phenomenological model," *Proceedings of Internoise'93*, Leuven, Belgium, pp. 641–646.

Li, K. M., Waters-Fuller, T., and Attenborough, K. (**1998**). "Sound propagation from a point source over extended-reaction ground," J. Acoust. Soc. Am. **104**(2), 679–685.

Lui, W. K., and Li, K. M. (**2004**). "A theoretical study for the propagation of rolling noise over a porous road pavement," J. Acoust. Soc. Am. **116**(1), 313–322.

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw–Hill, New York).

Sarradj, E. (**2003**). "Multi-domain boundary element method for sound fields in and around porous absorbers," Acust. Acta Acust. **89**, 21–27.

Seybert, A., Cheng, C., and Wu, T. (**1990**). "The solution of coupled interior/exterior acoustic problems using the boundary element method," J. Acoust. Soc. Am. **88**, 1612–1618.

Watts, G. R., Chandler-Wilde, S. N., and Morgan, P. A. (**1999**). "The combined effects of porous asphalt surfacing and barriers on traffic noise," Appl. Acoust. **58**, 351–377.

Wu, T. W., Cheng, C. Y. R., and Tao, Z. (**2003**). "Boundary element analysis of packed silencers with protective cloth and embedded thin surfaces," J. Sound Vib. **261**, 1–15.

Monitoring near-shore shingle transport under waves using a passive acoustic technique

T. Mason[a)]
Channel Coastal Observatory, National Oceanography Centre, European Way, Southampton, SO14 3ZH, United Kingdom

D. Priestley
Maritime Technology Division, Britannia Royal Naval College, Dartmouth, Devon, TQ6 0HJ, United Kingdom

D. E. Reeve
Centre for Coastal Dynamics and Engineering, School of Civil Engineering, University of Plymouth, Drake Circus, Plymouth, Devon PL4 8AA, United Kingdom

(Received 23 December 2005; revised 9 May 2007; accepted 11 May 2007)

Passive acoustic techniques have been used to measure shingle (gravel) sediment transport in very shallow water, near the wave breaking zone on a beach. The experiments were conducted at 1:1 scale in the Large Wave Flume, Große Wellen Kanal (GWK) at Hannover, Germany. The frequency spectrum induced by shingle mobilized under breaking waves can be distinguished from other ambient noise, and is found to be independent of water depth and wave conditions. The inverse relationship between centroid frequency and representative grain size is shown to remain valid in shallow water wave conditions. Individual phases of onshore and offshore transport can be identified. Analysis of the acoustic frequency spectrum provides insight into the mechanics of phase-resolved shingle transport.
 [DOI: 10.1121/1.2747196]

PACS number(s): 43.28.We [KGF] Pages: 737–746

I. INTRODUCTION

One of the most important reasons for monitoring beaches is to determine how much beach volume is needed either to prevent coastal flooding (by waves overtopping a beach or structure) or to prevent scour and subsequent failure of the foundations of seawalls. Shingle (gravel) is a particularly good sea defense, since the wave energy is dissipated in moving sediment and by infiltration into the porous beach, as illustrated by Andrews,[1] who documented a storm event at Seaton, South Devon, UK, where seawalls along the defended section of beach were undermined, while an adjacent stretch of natural shingle beach withstood the storm with very little damage. The vast majority of near-shore beach research, however, has been concerned with sandy beaches, although environmental conditions on shingle beaches are very different; shingle beaches being typically steep ($\sim$1:8), with significant three-dimensional flow through the beach due to the high permeability of the sediment. Wave breaking is confined to an energetic, narrow band of plunging breakers, close to the shore.

A recent review of 14 transport equations potentially suitable for coarse-grained beaches found that nearly all over-predicted the transport rate with factors varying up to five times the measured[2] due, in most cases, to the lack of grain size-dependent parameters. The review included formulas based on energetics[3] and force-balance[4] principles, as well as a range of equations derived by dimensional analysis (e.g., the Delft longshore equation for random waves[5]). The difficulty with the energy flux methods (which relate the transport rate to the wave power) is that the required coefficient, K, for shingle beaches is not only different from that used for sand beaches, but also shows considerable variability. This may be due, in part, to the paucity of shingle field experiments, as well as a lack of measured high energy events over which to calibrate the CERC-type formulas, but also suggests that the K value encompasses other site-specific features such as tidal currents. Damgård and Soulsby's[4] formulation was specifically derived for shingle sediment and related the sediment transport to the bed shear stresses, adapted for combined waves and currents, but the results over-predicted by a factor of 12, indicating that one or more of the parameters in this complex equation is not properly represented.

Cross shore, the emphasis is to predict the profile response to wave forcing, since the behavior is very different than that of sand beaches.[6] Many attempts have been made to integrate cross and longshore transport mechanisms, over a variety of time scales[7–11] but, as yet, no consensus has developed on a reliable and robust method for predicting the evolution of a gravel beach. Yet obtaining reliable estimates of sediment transport on shingle and mixed sand/shingle beaches is of particular engineering importance, particularly in the United Kingdom where over one-third of the coastline comprises shingle beaches and where beach management plans increasingly involve replenishment with shingle or mixed sediments.

[a)]Electronic mail: tem@noc.soton.ac.uk

Techniques for field measurement of shingle transport have not kept pace with those for sand, and technological improvements such as optical backscatter sensors or acoustic backscatter sensors have no real equivalent for shingle sized sediment. A potential development is the use of an acoustic doppler current Profiler to measure gravel bedload in rivers,[12] but very few instruments can withstand the harsh environment of fast, reversing flows, plunging breakers, and highly mobile shingle. The result is a notable lack of sediment transport measurements on shingle beaches, and those that do exist are mostly at the resolution of a tidal cycle. Indeed, long-term impoundment against a jetty is still considered one of the most reliable methods of estimating a sediment transport rate.[2] Alternative techniques such as painted, aluminum, or even, latterly, electronic pebble tracers[13] have serious drawbacks, notably, representation, being labor-intensive and providing only an estimate of the depth of disturbance on a gravel beach.[14]

In summary, both the modeling and the measurements (field and laboratory) on shingle beaches remain primitive in comparison to sandy beaches. Given the current difficulty of obtaining reliable sediment transport rates on shingle beaches, it was decided to revive the principle of using underwater sound to infer sediment transport and to investigate its application to shingle movement in the surf zone.

Experiments were carried out in conjunction with the Gravel and Mixed Beach Project at the Große Wellen Kanal (GWK), Coastal Research Centre, Hannover, Germany, in 2002. The research was conducted at 1:1 scale, thereby avoiding the main drawbacks of both field experiments (longshore or tidal currents, unpredictable wave climate, longshore sediment transport) and scaled laboratory experiments, since shingle sized sediment cannot be scaled properly for both sediment size and hydraulic conductivity. Further details about experimental aims of the Gravel and Mixed Beach Project are given in Blanco *et al.*.[15] In Sec. II of this paper, the application of measurement of shingle-generated noise in the marine environment is reviewed. Section III describes the experimental setup and potential limitations of the equipment used. Results are given in Sec. IV, with discussion and concluding remarks in Secs. V and VI, respectively.

II. PREVIOUS WORK

Sediment particles colliding with each other undergo a very rapid change in velocity resulting in a pulse of acoustic energy, which is transmitted into the water—the process known as rigid body radiation. In the marine environment, the appropriate acoustic generating mechanisms are bedload (rolling or sliding particles with persistent intergranular forcing) and saltation (intermittent intergranular forcing), where sediment particles are partially entrained by oscillatory or mean currents, or by wave breaking processes, but are too large to remain in suspension for half a wave cycle. The acoustic energy spectrum induced by mobile sediment can be used to infer information about the sediment in transport, such as the particle size and quantity of sediment in motion. The technique of identifying sediment size from the acoustic spectrum is limited to sediments larger than sand ($D_{50} > 2$ mm) since for smaller sediments the signal to noise ratio is too low, whilst for particles larger than 100 mm the spectrum is less easily distinguished from other ambient noise.

In a series of papers in the 1980s, Thorne calculated the theoretical acoustic resonance frequency spectrum of sediment particles and compared it with laboratory measurements of the acoustic spectrum generated by unimodal ballotini rotating in a drum. He confirmed that the centroid frequency of the noise was inversely proportional to the mean grain diameter, D_{50}, of the sediment and that the shape of the spectrum was not influenced by the mass of sediment in motion.[16] Thorne[17] extended the range of sediment sizes to include marine gravels up to 25 mm and found that, although the spectral peak was less well-defined than with ballotini (probably due to the spread of particle sizes and their nonsphericity), sediment size remained the controlling parameter for the shape of the spectrum and that the equivalent particle diameter, D, was related to the centroid frequency of the spectrum, f_c, by

$$f_c = \frac{209}{D^{0.88}}. \tag{1}$$

The centroid frequency was calculated by taking the frequency range over which significant spectral pressure levels were obtained and computing the centroid of the integral, defined as

$$\int_{f_1}^{f_c} P(f)df = \int_{f_c}^{f_2} P(f)df, \tag{2}$$

where $P(f)$ is the spectral pressure level (SPL); the significant region is given by $P(f) > 0.1P(f)_m$, where Pf_m is the maximum SPL of the spectrum, f_1 and f_2 the frequencies where $P(f) > 0.1P(f)_m$, and f_c is the centroid frequency. Earlier work with marine gravels related the particle size diameter to the peak frequency of the spectrum.[18]

Thorne[17] also reported a linear relationship between acoustic intensity and the mass of sediment in transport; this was later validated by field experiments in 15 m water depth in the western Solent[19–21] where the transport rates predicted by the hydrophone were compared to visual estimates from an underwater camera.

The measured ambient noise field near the surf zone contains acoustic energy from a number of sources, of which the two largest might be expected to be the noise of coarse-grained sediments in transport and the hydrodynamic noise, which is generated by many sources including bubble oscillations, air entrainment, and turbulence.[22,23] Richards *et al.*[24] suggested that wave-induced shingle noise was the principal source of acoustic energy near coarse-grained coasts and later, in one of the few shallow water acoustic field experiments near a gravel beach, Jones and Richards[25] set out to quantify the relative contributions of mobile shingle-induced noise and bubble (surf) noise. An array of four hydrophones was located outside the surf zone on a steep, gravel beach in about 5 m water depth. They expected bubble resonance to be the dominant source of acoustic energy but instead found that the acoustic spectrum could be attributed to individual

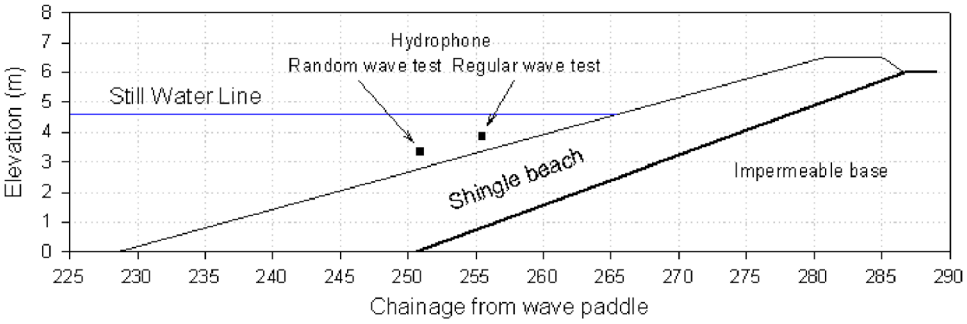


FIG. 1. (Color online) Schematic of GWK flume showing location of instrumentation.

bursts of sediment transport, interspersed with quiescent phases. Jones and Richards attributed the relative lack of hydrodynamic noise to strong absorption and scattering by the individual bubbles and bubble plumes. Acoustic absorption by high concentration of bubbles has been estimated at 50 dB m^{-1} by Deane,[26] based on measurements of spilling breakers. Although no detailed field measurements have been made for plunging breakers, it is likely that absorption and scattering is even higher, since plunging is concentrated over a narrow width of surf, additional air is entrapped by the leading edge, and the water is disturbed through the whole water column right to bed level. Other potential sources of noise are discussed in Sec. III.

III. EXPERIMENTAL SETUP

The GWK flume is 342 m long, 7 m deep, and 5 m wide with a permanent, impermeable slope of 1:6 at the "beach" end. The wave paddle can generate random waves of up to 2 m height. 25 capacitance wave gauges were placed along the side of the flume to measure the water surface elevation (Fig. 1). There is also a mobile instrument carriage and gantry which can be placed anywhere along the flume (Fig. 2).

The hydrophone used in the experiments was an omnidirectional Brüel & Kjær Type 8105, which had been calibrated by the manufacturer shortly before deployment in the GWK. The 8105 is a small, spherical instrument, with a piezoelectric ceramic sensor, bonded on to sound-transparent polychloroprene rubber. Hence, it was sufficiently rugged to withstand the hostile environment of breaking waves and mobile shingle. The dynamic range of the 8105 (respectively, 250 Hz) is 50 to 15 000 Hz, +3.5 to −10.0 dB. Receiving voltage sensitivity was −206.3 dB (±0.25 dB) reference to 1 V/μPa. It is omnidirectional over 360° across its full frequency range and its frequency response is flat across the range of frequencies of interest to this research (0.5–12 kHz). The hydrophone was connected via a 10 m shielded cable to a conditioning amplifier Type 2650. Transducer sensitivity was set to 4840, as per the calibration, and the transducer range to 0.1 V/unit. A hardware band pass filter was applied to the input signal, to remove frequencies lower than 0.3 kHz and higher than 30 kHz. The signal output from the conditioning amplifier was digitized at 48 kHz and recorded onto high quality tape by a Sony digital video recorder type TRV-30E. The hydrophone was mounted on the mobile gantry, equidistant from the sides of the flume and at a height of 0.4 m above the bed as shown in Fig. 3.

The hydrophone was held in place using a fixed mounting, clamped 0.05 m above the sensing element. Also mounted on the mobile gantry was a vertical array of acoustic Doppler velocimeters (ADV) to measure wave-induced currents at 0.3, 0.5, and 0.7 m above the bed, and a capacitance wave gauge. These were logged simultaneously at

FIG. 2. (Color online) GWK viewed from beach end, showing mobile gantry with instruments submerged (the roller was used to profile the beach in between wave tests).

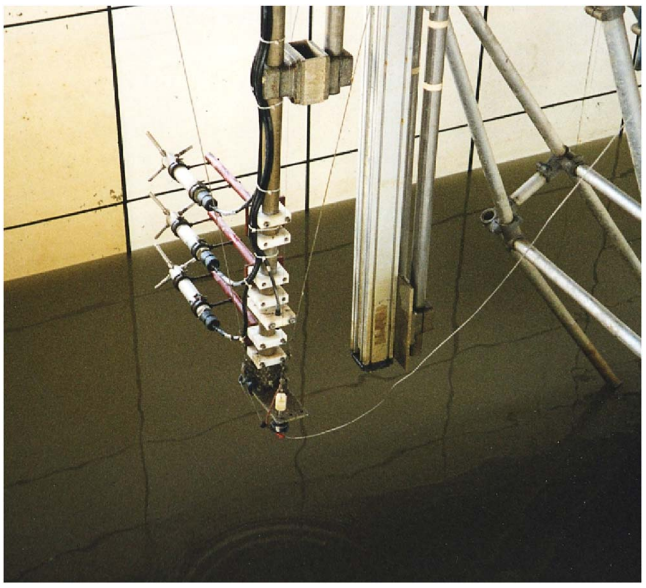


FIG. 3. (Color online) Hydrophone and ADVs mounted on mobile gantry (the hydrophone is fixed in the third block from the bottom).

60 Hz. An early aim had been to deploy an underwater video camera for simultaneous recordings of the acoustic signal and a visual record of sediment in transport in order to confirm transport directions and, possibly, transport rates, but the underwater visibility in the flume was too poor even for a low-light camera.

Potential limitations of the method are the operating capabilities of the measuring equipment itself and contamination of the acoustic signal by unwanted ambient noise. The principal aim of the experiments was to measure shingle put into motion by shoaling and/or breaking waves. The generally steep nature of shingle beaches results in a narrow surf/swash zone, with a single line of energetic, plunging breakers and therefore most sediment transport is within or just seawards of this surf zone. This is in direct contrast to waves on a flatter, sandy beach where energy is generally dissipated over a wide surf zone with several lines of spilling breakers. Furthermore, since the beach is steeply shelving, it is possible for the hydrophone to be out of the water for brief periods. Accordingly, the mounting arrangement for the hydrophone was a compromise between security and ideal free-stream conditions.

The operating frequency of the ADVs was 10 MHz and therefore well outside the receiving range of the hydrophone. Potential sources of acoustic contamination include machinery noise from the wave paddle and hydrodynamic noise. The hydrophone recordings for every test started at least 30 s before the paddle started moving, in order to examine the acoustic signature of the paddle and to measure ambient noise levels within the flume. The small diameter (0.022 m), sphericity and rubber encasement, small mounting block, and siting all serve to reduce flow noise to a minimum. Nonetheless, the solid vertical structure of the gantry and the adjacent ADVs meant that some flow noise was inevitable, both induced by the structure itself and generated elsewhere and advected past the hydrophone. Deane[22] demonstrated, however, that flow noise around a hydrophone is generally restricted to frequencies below 50 Hz, whilst machinery noise is usually confined to frequencies around or below 500 Hz. The wave paddle was 340 m from the hydrophone at its closest. Examination of all acoustic recordings showed that onset of a discernible acoustic energy record coincided with wave motion at the hydrophone and with negligible acoustic intensity before that. Accordingly, both flow and machinery noise are below both the lowest frequency for which the acoustics method is applicable ($\sim$1 kHz) and are also below the window between wave noise and sediment noise, which lies between 500–1500 Hz, so neither source is likely to contaminate the recorded acoustic signal in the frequency range of interest.

A distinct advantage of conducting acoustic experiments in the GWK flume was freedom from contamination by traditional sources of ambient noise in the open ocean; namely shipping noise, precipitation, ice, or biological noise. The final potential contaminant of the acoustic signal is noise generated by breaking waves, which in the near-shore region is proportional approximately to the square of the wave height[23] and, for spilling breakers on a sand substrate, is at a frequency of about 500 Hz, with negligible intensities at frequencies higher than 1 kHz; indeed field experiments have found very little acoustic energy from any source between about 500–1500 Hz in the surf zone.[27] Consequently, it is considered that: (i) the acoustic record represents shingle-sized sediment in transport; (ii) the acoustic signal recorded by the hydrophone can be regarded as free from contamination at the frequencies of interest, between 0.5 and 12 kHz; and (iii) that the noise generated by moving shingle is clearly distinguishable from other sources of ambient noise.

Two different beach types were used for the experiments: (1)sieved gravel between 16 and 32 mm with a median diameter, D_{50}, of 21 mm and (2) mixed sand and gravel; a bimodal sediment where the gravel was mixed with about 30% sand of D_{50}=300 μm. Both beach types were placed at an initial slope of 1:8 (typical of natural shingle beaches), with a minimum depth of 2 m of sediment, in order that the pattern of groundwater flow within the beach should be properly represented. A series of both random (JONSWAP type spectra[28]) and regular waves was run and the beach was profiled between each test.

The wave generation program was run until the desired number of waves was achieved. Long sequences were used to ensure that a wide spectrum of wave heights and periods were generated in each test. The generated waves were validated using 25 capacitance wave gauges deployed along the wave channel. The surf zone consisted of a single line of plunging breakers. The cross-shore position of the instrumented mobile gantry was moved after each wave test, which meant that acoustic recordings could be made during identical wave conditions, but at varying distances from the still water line (and, by implication, from the wave breaking position). The digital video recorder was positioned to obtain a simultaneous video recording of the waves passing over the hydrophone so that the exact moment and position of wave breaking could be obtained. Recordings started before the wave paddle was switched on and each recording lasted approximately 5 min. In total, over 9 h of recordings were made.

Two minute long subsections of the acoustic recordings were extracted starting 10 s before the first observed wave motion. Since the incipient motion of the water directly above the hydrophone was indicated in the video record by small bubbles moving on the water surface, the acoustic records were synchronised with the wave and current data files with a precision of approximately 0.25 s.

IV. RESULTS

The results from two representative shingle beach experiments, one for regular waves and one for random waves, are presented here.

A. Test one-Regular waves

In this case, regular waves of 6.5 s period and 1.0 m wave height were generated. The hydrophone was deployed approximately 2 m seaward of the breaker zone and (10 m seaward of the still water line), in about 1 m water depth. Figure 4 shows the initial portion of the recorded ambient noise spectra associated with the passage of two waves. Data

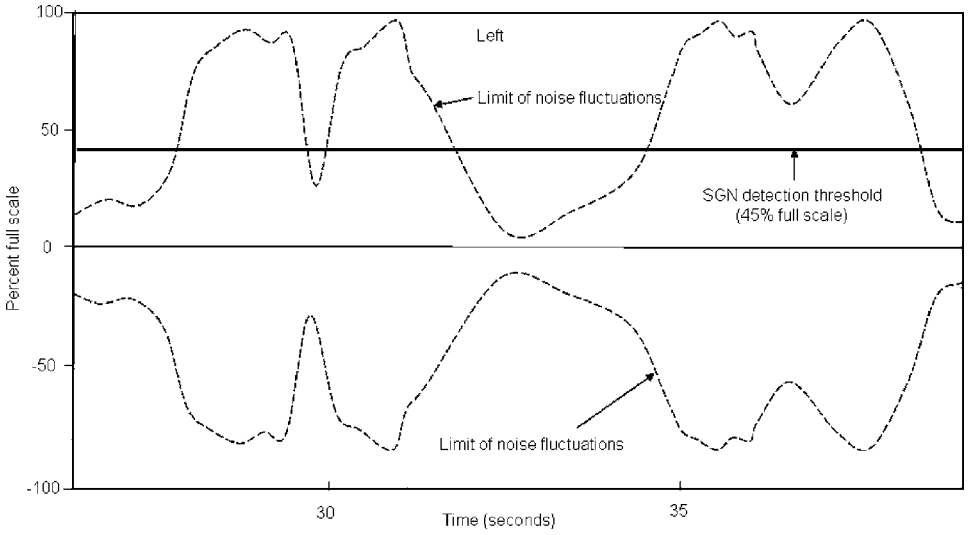


FIG. 4. Mobile shingle detection threshold.

are normalized relative to the maximum signal which can be received by the PC sound card. A frequency integrated noise threshold was set at 45% of the full scale value, which is the equivalent of a recognition differential of 1 dB above the background noise; signals exceeding this threshold were regarded as wave-induced shingle noise. For the purpose of sediment transport research, the time-frequency domain is more useful and Fig. 5 shows a spectrogram for nine waves, display-averaged at 0.1 s. The darker shading indicates the higher intensity signal levels (the pale horizontal lines are the result of electrical pickup from the cable connecting the hydrophone to the video recorder).

The spectra are dominated by discrete events of approximately 4 s duration, which have relative amplitudes up to 20 dB above the background noise level. Apart from the first event, these discrete signals occur in pairs but with little or no intervening time interval. The interval between the start of consecutive paired events is approximately 6.5 s, i.e.,. correlates very well with the wave period, indicating that the signals are due either to the hydrodynamic noise associated with wave breaking, or due to the noise generated by intergranular shingle collision. The frequency component of the events ($\sim$1.5–12 kHz) is consistent with that of noise resulting from coarse shingle bedload transport[18,29] rather than the noise generated by breaking waves, which generally occurs at frequencies between 100 and 800 Hz.[27] The shape and frequency content of the events is also inconsistent with that previously described for breaking waves by Bass and Hay. The sudden increases in relative signal amplitude occur across almost the full range of frequencies and persist for between 1 and 4 s, before decaying almost as rapidly as they appear. They are more or less symmetrical in shape (see Fig. 6) where overlays 1, 2, and 3 are the averaged spectra for the offshore, onshore, and quiescent segments, respectively, of one paired event. In contrast, Bass and Hay[27] describe marked asymmetry in the spectral form of breaking wave noise events. Thus, it would appear from the frequency composition and periodicity of these discrete events that they are attributable to the noise that results from shingle transport under the waves generated in the channel.

The calculated centroid frequencies (using the inversion method of Thorne[17]) for the discrete events are shown in Fig. 7. There is a gradual coarsening of the mobile grain population over the first five to six wave cycles; equivalent grain diameters are approximately 20 mm for the first four waves, rising to between 25 and 30 mm and remaining consistent thereafter. This suggests initial selective mobility in the rela-

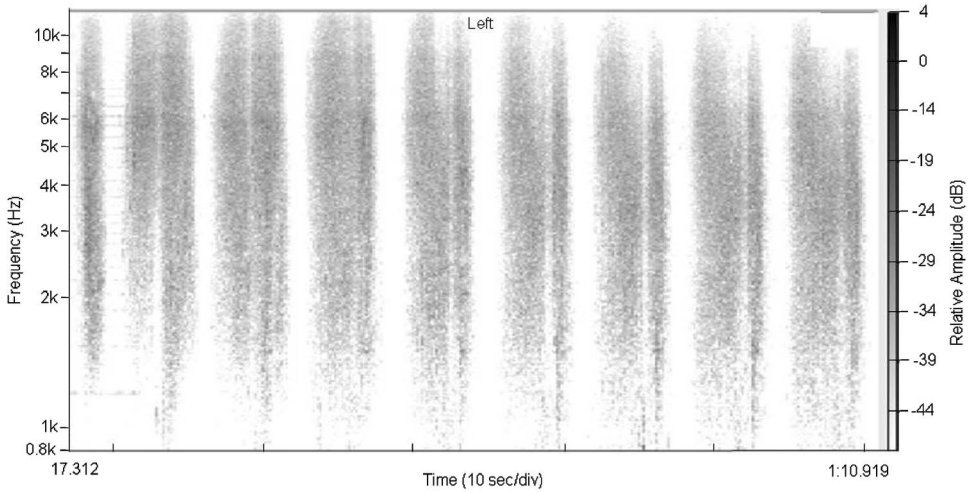


FIG. 5. Spectrogram for regular waves.

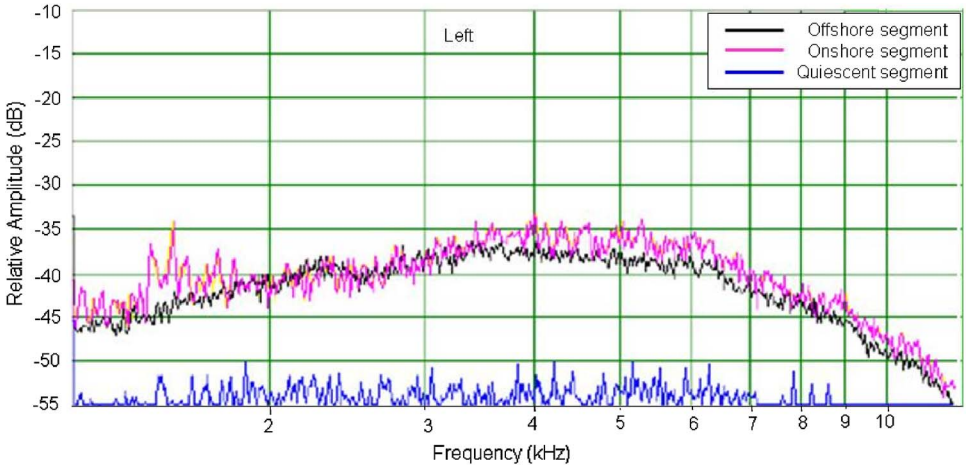


FIG. 6. (Color online) Averaged acoustic spectra for offshore, onshore, and quiescent phases of a paired transport event.

tively low current velocities early in the record as the wave heights increase toward their target height of 1 m. Thereafter, all grains are consistently mobile.

The grain diameters identified as being mobile using the acoustic inversion technique match very closely the grain sizes of the sediment in the flume, as illustrated in Fig. 8, which shows the grain size distribution of three separate samples of the beach sediment, indicating the grain size range from 10 to 32 mm, with a median diameter (D_{50}) of 21 mm.

The acoustically estimated mobile grain diameter range of 20–29 mm would appear to be an indicator of the grain size range D_{50}–D_{90}, i.e., of the coarser element of the grains present at the site. This is as expected, since the collision of the coarsest grains would dominate the acoustic spectrum and finer grains could be expected to settle into the interstices between the coarser grains, and hence contribute less to the acoustic record.

Onshore events initially have a duration of approximately 2 s, but once the wave train becomes properly established (after the first two or three wave cycles) the onshore duration is approximately 1 s and remains consistent. In contrast, offshore events initially have a duration of approximately 2 s, increasing to approximately 4 s. As a consequence, a clear asymmetry is seen in the duration of the onshore and offshore transport events. The switch from a relatively long offshore transport phase to a shorter onshore phase occurs almost instantaneously. There is, however, a consistent interval of approximately 1.5 s between the cessation of the onshore transport and the commencement of the offshore transport.

After the first few waves, the intensity levels and frequency content associated with the discrete events vary little with time, suggesting that the diameters of the mobile grain population and the transport rate become consistent. This indicates that all the bed is mobile—there is no selective mobility, as the threshold velocity for the bed as a whole has been exceeded.

B. Test two: Random waves

For random wave conditions, a JONSWAP-type spectrum with a peak period of 5.2 s and 1.0 m significant wave height (H_s) was used. The results presented here are for the hydrophone deployed 4 m further offshore than for the regular wave test, where the water depth was approximately 1.6. m. Figure 9 shows a segment of the spectrogram covering the start-up phase. In this case, unlike the regular wave

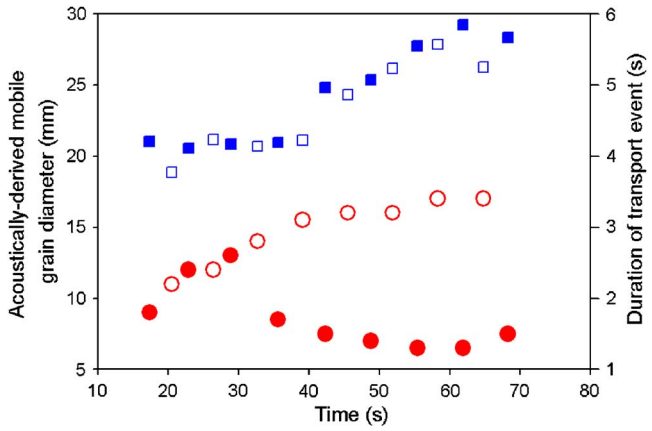


FIG. 7. (Color online) Mobile grain diameter derived acoustically for onshore (closed square) and offshore (open square) transport events, together with the duration of each onshore (closed circle) and offshore (open circle) events, under regular waves.

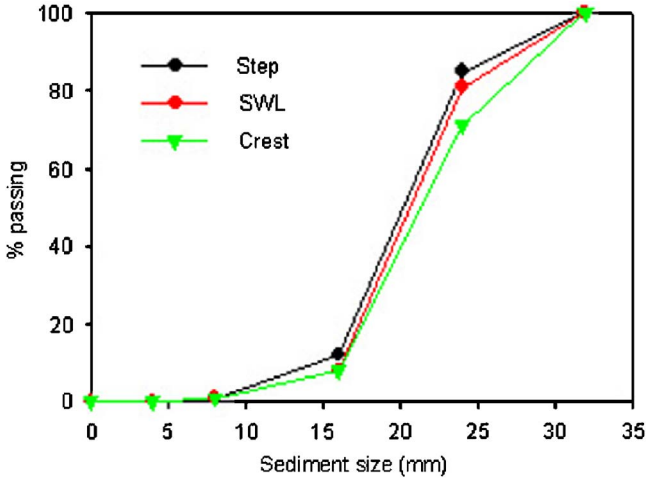


FIG. 8. (Color online) Grain size distributions of the beach sediment at start of tests, taken at the beach step, the intersection of the beach and the still water line and at the beach crest.

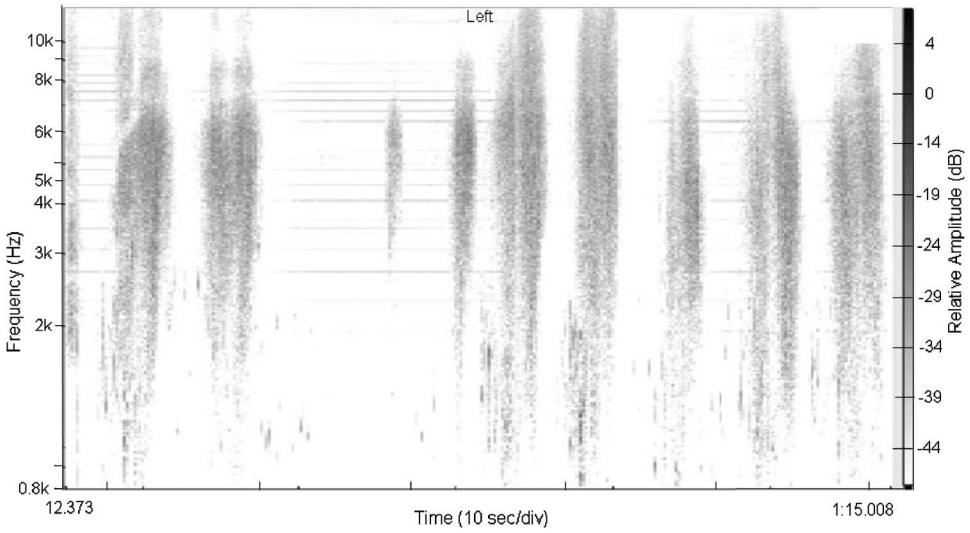


FIG. 9. Spectrogram for random waves.

conditions, the spectrogram shows transport events that are "paired" but irregularly spaced in a temporal sense, particularly after the first few waves when the wave record becomes truly random in nature. There is also evidence in this record of an isolated onshore transport event occurring with no paired offshore transport event.

Some, but not all, onshore events have higher acoustic intensities, implying greater bedload transport rates. In common with the regular wave record, there is no cessation of transport at the end of the offshore transport phase, but a gap of approximately 1 s again occurs following the onshore phase. The frequency component, however, shows marked variability so, employing the acoustic inversion technique, this implies significant variation in the diameter of the mobile grain population. The variation in the computed mobile grain diameter and event duration as a function of time can be seen in Fig. 10, where equivalent grain diameters range from approximately 18 to 33 mm, which is again within the grain size range of the shingle deployed in the flume.

V. DISCUSSION

The aim of the acoustic experiments in the GWK was to determine whether the technique of using acoustic measurements to represent shingle transport was valid in shallow, oscillatory flow near the surf zone. The results showed that the periodicity of the acoustic events is identical to that of incident wave period; hence, the source must be hydrodynamic noise or mobile shingle-induced noise. The frequency of the recorded noise is exactly that expected from earlier experiments[17] for the grain diameters used in the experiment. The method has been validated previously in the field for shingle under tidal (unidirectional) currents using underwater cameras,[20,21] and these experiments confirm that the mobile shingle-induced noise frequency is the same in the surf zone and thus shingle transport noise is independent of wave conditions or water depth.

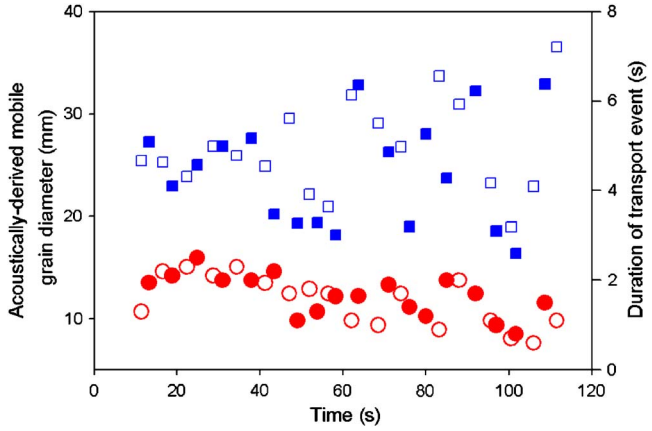


FIG. 10. (Color online) Mobile grain diameter derived acoustically for onshore (closed square) and offshore (open square) transport events, together with the duration of each onshore (closed circle) and offshore (open circle) events, under random waves.

Before further deductions can be drawn from the acoustic recordings, two important facets of this research must be established, namely just where the noise recorded by the hydrophone is being generated and to assess the footprint of the hydrophone. Wave noise occupies a different frequency band to that generated by shingle in transport, being strongest below 1 kHz. Despite all recordings being made within a few meters of the breaker zone, none showed the presence of wave noise, in agreement with the findings of Jones and Richards.[25] The lack of noise from the breaker zone indicates the existence of an acoustic barrier landward of the hydrophone. This can be attributed both to very high propagation losses due to strong absorption and scattering by bubbles within the plunging breaker region,[30] but also to the sharply positive sound speed gradient induced by the lower density of the highly aerated near surface water. Both factors limit the detection radius of the hydrophone and suggest that the bulk of the noise received at the hydrophone will be locally derived.[31] Although most transport is likely to occur landward of the breaker zone, significant shingle transport can take place outside the surf zone, particularly in these energetic and rapidly shoaling wave conditions. The depth at which sediment mobilization will occur can be calculated using the maximum oscillatory current, U_{max}, in shallow water, and the critical oscillatory current under waves, U_{wcr}, of 0.8 ms^{-1} for sediment D_{50} of 21 mm,[32] indicating that sediment will be mobilized at depths shallower than 3.5 m—each wave progressively mobilizing sediment shore-

ward, with the result that shingle-induced noise will be generated consistently across the profile shoreward of the critical depth. If a fixed-point hydrophone were capturing noise over a wide area or, indeed, from a transport event far from the hydrophone, there should be a near-continuous signal, fluctuating with time. This is not the case, however, since the hydrophone recorded discrete signals of short duration, interspersed with periods of no measurable energy. Further evidence for the local nature of the noise generating area can be provided by estimating the implied detection radius of an omnidirectional hydrophone, by calculating the shoreward propagation of a pulse of shoaling/breaking wave-induced shingle noise,[27] which should equate to the wave phase speed. Using the shallow water approximations of linear wave theory, and assuming that sediment threshold is half the maximum velocity, a wave of 5 s in 1 m water depth will generate a mobilized patch of sediment of about 4 m, which will pass the hydrophone in ~1.25 s. This equates closely to the discrete ~1.5 s transport bursts identified from the acoustic recordings and lends weight to the localized nature of the signal being received by the hydrophone. Furthermore, if the noise were being generated at some distance from the hydrophone, the peak intensity would be phase-shifted from the current. For example, the wavelength of a 5.1 s wave in 1 m water depth is approximately 16 m, hence the 6 m horizontal difference between the swash zone and the offshore hydrophone represents a phase difference of approximately 1.9 s. No such phase lag is observed; the synchronized acoustic/ADV records confirm that, at both hydrophone locations, the peak signal is in phase with the maximum current velocity. These lines of evidence, taken together, indicate strongly that the signal recorded by the hydrophone results from shingle being transported beneath or very close to the hydrophone, although future experiments will be required using an array of hydrophones to confirm the contributing area to the acoustic spectrum.

The synchronization of the acoustic and ADV records make it possible to identify where in the current record transport commenced and finished, with a temporal resolution, of 0.25 s or less, as shown for the regular wave test (Fig. 11), which shows the velocity time series with filled patches superimposed representing the duration of onshore and offshore transport bursts as derived from the acoustic record.

The synchronization also makes it possible to identify the threshold velocity associated with each acoustic event. In the regular wave test, a pattern develops after the first few waves where in the onshore direction the shingle starts to move with near zero current, i.e., it is mobilized, and almost immediately the current reverses from offshore to onshore; there is virtually continuous mobility of shingle at this point in the wave cycle. In an offshore direction, the threshold velocity remains consistent at approximately 0.2 m s^{-1}. At the completion of the offshore transport phase, the sediment is subject to rapid acceleration and is immediately mobilized in the onshore direction. In contrast, there is a time gap following cessation of onshore transport; here it is hypothesized that the sediment settles back into position until it is eventually mobilized when velocities exceed 0.2 m s^{-1} by the less rapidly accelerating offshore current. A similar pattern of

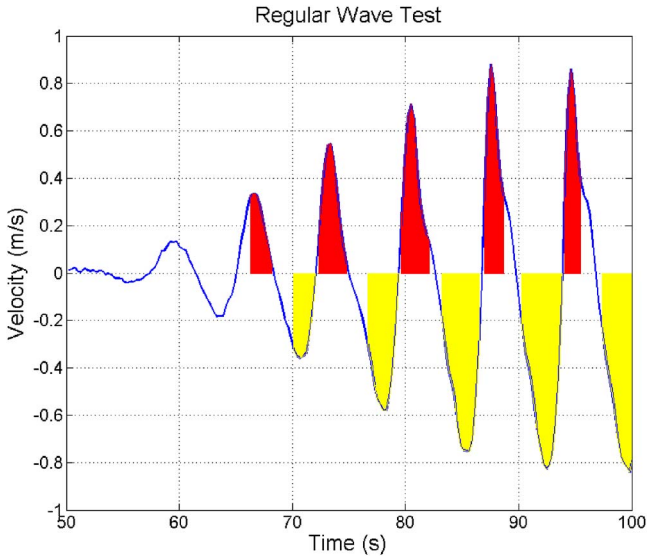


FIG. 11. (Color online) Current velocities measured by near-bed ADV, with superimposed periods of onshore and offshore shingle transport under regular waves (duration of transport is derived from the acoustic records).

threshold velocities is observed in the random wave record. Offshore events have threshold velocities ranging from 0.1 to 0.5 m s^{-1}. In one instance an offshore current that peaks at 0.21 m s^{-1} has no associated signal, indicating that it fails to mobilize the bed sediment. This demonstrates clearly the potential of the acoustic technique to establish threshold velocities for individual onshore or offshore events within a wave cycle.

These results so far serve to extend the realm of validity of the technique of using the acoustic signal as an indicator of shingle transport into shallow water, including a highly energetic surf/surf zone, where it is very difficult to measure in any other way. But the most fundamental leap forward provided by these experiments is the light shed on thresholds of sediment motion and in establishing different phases of transport within a single wave. The acoustic method allows, for the first time, very precise identification of the threshold current velocity required to mobilize shingle within individual phases of a wave cycle. Values of the Shields parameter were derived, but early analysis of the results indicates that the Shields curve (modified for coastal environments[33]) over-predicts the threshold velocities. Ongoing work is examining the possibility that spikes in the acceleration time series may be more informative in triggering transport events than either the measured velocity or the Shields parameter.

The acoustic signal also has the potential to identify sediment sorting processes, since the spectrum can resolve transport with very high time resolution. Analysis so far has indicated that the composition of the acoustic signal is slightly different for onshore and offshore phases—the centroid frequency is similar, but the duration is different, as is the short phase of progressive mobility of larger grains which is most usually present on offshore events only. This suggests that sediment sorting processes are confined to only a very small part of the wave phase, and that for the majority of the wave cycle, the concept of equimobility applies. This has important implications for phase-resolved shingle trans-

port models which model different sediment size fractions individually and then recombine to predict a wave-by-wave sediment transport rate. Nevertheless, some sediment sorting was found to occur during the shingle beach tests, and therefore the facility to validate the proportion of the wave when different fractions are mobile will be of use for mixed sediment modeling.

The present work has focused on establishing the potential for using acoustics as an indicator of shingle transport in the surf zone, based primarily on the frequency spectrum. Since the intensity of the recordings is relative rather than absolute, no attempt has been made to establish a sediment transport rate (a bulk rate and onshore/offshore rates); neither was there independent validation from another method of measuring sediment transport, other than at the coarse resolution of measuring the beach profile after each wave test. Proposals have been provisionally accepted for further experimental work in a large wave facility to confirm and extend these findings, using directional arrays with a narrowly defined footprint to quantify the noise contributing area, together with an integrated underwater camera to record the timing, location, and direction of sediment transport. Sediment transport rates will also be derived by estimating the transport in each graticule of a grid superimposed on the underwater video image and integrated through time, as conducted by Thorne[20] but integrated over short time periods appropriate for oscillatory flow in shallow water.

VI. CONCLUDING REMARKS

Previous laboratory experiments investigating shingle transport have been hampered by the physical impossibility of simultaneously scaling properly for hydrodynamic behavior and hydraulic conductivity. The 1:1 experiments conducted in the GWK provide confidence in the realistic behavior of shingle in transport under wave action within carefully measured and typical field conditions. The acoustic frequency and spectral shape of shingle in transport in the surf zone has been found to be the same as in deep water and is, therefore, proven to be independent of wave conditions and water depth. It has been demonstrated from the duration of the signal, its frequency content, and spectral shape, that it is possible to use passive acoustics to monitor sediment movement on shingle beaches under a range of wave conditions. Furthermore, the technique can be applied reliably in the high-noise environment of the surf zone, even very close to wave breaking.

The acoustics method has been shown to have the capability to embrace very rapid changes in sediment mobility and, therefore, can be utilized in bidirectional as well as unidirectional flows. Indeed, it can offer insight into the mechanics of phase-resolved shingle transport which is unobtainable from any other method at present. The extremely good match between the acoustic signal and the hydrodynamics means that the acoustic analysis can be used as a representative *in situ* measurement of sediment transport in what can be a very harsh and hazardous environment. Potential extensions to the analysis include establishing the threshold of motion, both for given sediment sizes and for the condition of equimobility, and to compare the threshold orbital velocities obtained from the acoustic measurements with the few theoretical equations which extend to shingle-sized sediment, e.g., Soulsby[33] and Komar and Miller.[34] It is also anticipated that the acoustic recordings may be used to provide an estimate of transport rates for individual onshore and offshore events.

ACKNOWLEDGMENTS

The research project was funded by the UK Department of Environment, Food and Rural Affairs, under Commission FD1901 (Development of Predictive Tools and Design Guidance for Mixed Beaches—Stage 2). The experiments in the Large Wave Channel (GWK) of the Coastal Research Centre (FZK) were supported by the European Community under the Access to Research Infrastructures, Human Potential Programme (contract HPRI-CT-1999-00101). The authors acknowledge, with thanks, the assistance provided by the staff of the GWK flume (Hr, Joachim Grüne, Dr. Uwe Sparboom, Reinold Schmidt-Koppenhagen, Wolfgang Malewski, Dieter Junge, Günter Bergmann, and Kai Irschik) and the project team in the UK (Andrew Bradbury, Belen Blanco, and Tom Coates) for inviting them to take part in the experiments. The assistance of Maurice McCabe and Nick Smith during the experimental phase and of Brian Parker from BRNC was also much appreciated.

[1]J. Andrews, "Coast protection at sidmouth," in Proceedings of the 29th International Conference on Coastal Engineering, American Society of Civil Engineers, 2004, pp. 2943–2953.
[2]E. Van Wellen, A. J. Chadwick, and T. Mason, "A review and assessment of longshore sediment transport equations for coarse grained beaches." Coastal Eng. **40**, 243–275 (2000).
[3]J. A. Bailard, "An energetics total load sediment transport model for a plane sloping beach," J. Geophys. Res. **86**, 10938–10954 (1981).
[4]J. S. Damgård and R. L. Soulsby, "Longshore bedload transport," in Proceedings of the 25th Internatioanl Conference on Coastal Engineering, American Society of Civil Engineers, 1996, pp. 3614–3627.
[5]E. Van Hijum and K. W. Pilarczyk, "Equilibrium profile and longshore transport of coarse material under regular and irregular wave attack," Delft Hydraulics Laboratory, Delft, The Netherlands, Publication No. 274, 1982.
[6]M. C. Quick and P. Dyksterhuis, "Cross-shore transport for beaches of mixed sand and gravel," in the International Symposium on Symposium: Waves and Physical Modelling, Vancouver, BC, Canadian Society of Civil Engineers, 1994, pp. 1443–1452.
[7]R. B. Nairn and H. N. Southgate, "Deterministic profile modelling of nearshore processes. 2: Sediment transport and beach profile development," Coastal Eng. **19**, 57–96 (1993).
[8]J. A. Roelvink and I. H. Broker, "Cross-shore profile models," Coastal Eng. **21**, 163–191 (1993).
[9]K. A. Rakha, R. Deigaard, and I. Brøker, "A phase-resolving cross shore sediment transport model for beach profile evolution," Coastal Eng. **31**, 231–261 (1997).
[10]J. Lawrence, A. J. Chadwick, and C. A. Fleming, "A phase resolving model of sediment transport on coarse grained beaches" in Proceedings of the International Conference on Coastal Engineering, American Society of Civil Engineers, 2000, pp. 624–636.
[11]L. C. Van Rijn, A. G. Davies, J. van de Graaff, and J. S. Ribberink, *Sediment Transport Modelling in Marine Coastal Environment* (Aqua, Amsterdam, 2001).
[12]C. D. Rennie and P. V. Villard, "Site specificity of bed load measurement using and acoustic Doppler current profiler," J. Geophys. Res. **109**, 3003–3018 (2004).
[13]G. Voulgaris, M. Workman, and M. B. Collins, "Measurement techniques of shingle Transport in the nearshore zone," J. Coastal Res. **15**, 1030–1039 (1999).
[14]T. Mason and T. T. Coates, "Sediment transport processes on mixed

beaches: A review for shoreline management," J. Coastal Res. **17**, 645–657 (2001).
[15]B. Blanco, T. T. Coates, P. Holmes, A. J. Chadwick, A. Bradbury, T. E. Baldock, A. Pedrozo-Acuña, J. Lawrence and J. Grüne, "Large-scale experiments on gravel and mixed beaches: Experimental procedure, data documentation and initial results," Coastal Eng. **53**, 349–362 (2006).
[16]P. D. Thorne, "The measurement of acoustic noise generated by moving artificial sediments," J. Acoust. Soc. Am. **78**, 1013–1023 (1985).
[17]P. D. Thorne, "Laboratory and marine measurements on the acoustic detection of sediment transport," J. Acoust. Soc. Am. **80**, 899–910 (1986).
[18]N. W. Millard, "Noise generated by moving sediments," Institute of Acoustics, Proceedings of the Conference on Recent Developments in Underwater Acoustics, AUWE, Portland, 31 March and 1 April 1976, Paper 3.5.
[19]P. D. Thorne, "An intercomparison between visual and acoustic detection of seabed gravel movement," Mar. Geol. **72**, 11–31 (1986).
[20]P. D. Thorne, J. J. Williams, and A. D. Heathershaw, "In situ measurements of marine gravel threshold and transport," Sedimentology **36**, 61–74 (1989).
[21]J. J. Williams, P. D. Thorne, and A. D. Heathershaw, "Comparisons between acoustic measurements and predictions of the bedload transport of marine gravels," Sedimentology **36**, 973–979 (1989).
[22]G. B. Deane, "Sound generation and air entrainment by breaking waves in the surf zone," J. Acoust. Soc. Am. **102**, 2671–2689 (1997).
[23]G. C. Lauchle and A. R. Jones, "Flow-induced self-noise on a spherical acoustic velocity sensor," Joint 133rd Meeting of the Acoust. Soc. Am. and Noise-Con 97, State College, PA [J. Acoust. Soc. Am. **101**, Pt. 2, 3053 (1997)].
[24]S. D. Richards, A. D. Heathershaw, and P. D. Thorne, "The effect of suspended particulate matter on sound attenuation in seawater," J. Acoust. Soc. Am. **100**, 1447–1450 (1996).
[25]S. A. S. Jones and S. D. Richards, "Ambient noise measurements in the surf zone," Proc. Inst. Acoustics **23**, 335–341 (2001).
[26]G. B. Deane, "A model for the horizontal directionality of breaking waves in the surf zone," J. Acoust. Soc. Am. **107**, 177–191 (2000).
[27]S. A. Bass and A. E. Hay, "Ambient noise in the natural surf zone," IEEE J. Ocean. Eng. **22**, 411–424 (1997).
[28]K. Hasselmann, T. P. Barnett, H. Bouws, H. Carlson, D. E. Cartright, K. Enke, J. A. Ewing, H. Gineapp, D. E. Hasselmann, P. Kruseman, A. Meerburg, P. Muller, D. J. Olbers, K. Richter, W. Sell, and H. Walden, "Measurements of wind-wave growth and swell decay during the joint North Sea wave project (JONSWAP)," Dtsch. Hydrogr. Z. **12**, 8–95 (1973).
[29]J. Hardisty, "Monitoring and modelling sediment transport at turbulent frequencies," in *Turbulence: Perspectives on Flow and Sediment Transport*, edited by N. J. Clifford, J. R. French, and J. Hardisty (Wiley, New York, 1993), pp. 35–59.
[30]G. B. Deane, "A model for the horizontal directionality of breaking waves in the surf zone," J. Acoust. Soc. Am. **107**, 177–191 (2000).
[31]O. B. Wilson, M. S. Stewart, J. J. Wilson, and R. H. Bourke, "Noise source level density due to surf. I. Monterey Bay, CA," IEEE J. Ocean. Eng. **22**, 425–433 (1997).
[32]R. Soulsby, *The Dynamics of Marine Sands: A Manual for Practical Applications* (Telford, London, 1997), 256 pp.
[33]R. L. Soulsby and R. J. Whitehouse, "Threshold of sediment motion in coastal environments," Proceedings of the Pacific Coasts and Ports '97 Conference, University of Christchurch, New Zealand, 1997, pp. 149–154.
[34]P. A. Komar and M. C. Miller, "The threshold of sediment motion under oscillatory water waves," J. Sediment. Petrol. **43**, 101–110 (1974).

Frequency dependence and intensity fluctuations due to shallow water internal waves

Mohsen Badiey[a)]
College of Marine and Earth Studies, University of Delaware, Newark, Delaware 19716

Boris G. Katsnelson[b)]
Voronezh State University, Universitetskaya Sq.1, Voronezh 394006, Russia

James F. Lynch[c)]
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Serguey Pereselkov
Voronezh State University, Universitetskaya Sq. 1, Voronezh 394006, Russia

(Received 27 July 2006; revised 16 February 2007; accepted 8 March 2007)

A theory and experimental results for sound propagation through an anisotropic shallow water environment are presented to examine the frequency dependence of the scintillation index in the presence of internal waves. The theory of horizontal rays and vertical modes is used to establish the azimutal and frequency behavior of the sound intensity fluctuations, specifically for shallow water broadband acoustic signals propagating through internal waves. This theory is then used to examine the frequency dependent, anisotropic acoustic field measured during the SWARM'95 experiment. The frequency dependent modal scintillation index is described for the frequency range of 30–200 Hz on the New Jersey continental shelf. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2722052]

PACS number(s): 43.30.Bp, 43.30.Dr, 43.30.Es, 43.30.Zk [WMC] Pages: 747–760

I. INTRODUCTION

The study of broadband acoustic wave propagation in shallow water in the presence of inhomogeneities has attracted increasing attention in recent years, with one of the more important sub-topics being the study of the intensity fluctuations of the acoustic field due to nonlinear internal waves.[1–6] In particular, the frequency dependence of the variability of sound propagation in shallow water is of interest and is closely related to the spatial and temporal variability of physical oceanographic entities such as internal waves. We have recently presented studies[6] showing the frequency dependence of the refraction of horizontal rays, that correspond to specific acoustic normal modes, using the Weinberg-Burridge formalism.[7] In our previous theoretical[8–11] and experimental[2] papers, we showed that for an acoustic track placed approximately parallel to the wave fronts of internal solitary waves (ISWs), significant horizontal refraction takes place, leading to rather remarkable intensity fluctuations due to the "focusing and defocusing" of the horizontal rays (HRs) during the passage of the internal wave solitons.

In the present paper, we consider this frequency dependence in more detail, both theoretically and using further experimental data from the SWARM'95 experiment.[3] In particular, we consider the frequency dependence of the fluctuations of the modal intensity as a function of geotime. It was noted[6] that the shape of the amplitude dependence on frequency for different modes repeats the shape of the frequency dependence of the refraction index in the horizontal plane for a given mode number. Because this behavior is intuitively clear, our previous paper did not present a formal explanation of this phenomenon. In this paper we give an explanation in ray language, and point explicitly to the areas in the horizontal plane where we can observe such behavior. Additionally, we compare theory with experimental results obtained in the SWARM'95 experiment for different positions of the sources, frequency bands, and time periods.

We start by considering the properties of sound radiated by a point source in the horizontal plane, using the Weinberg-Burridge "horizontal rays and vertical modes" formalism.[8] To do this, we construct a diagram (Fig. 1), where, in the horizontal (XY) plane, the wave fronts of the internal solitary waves (ISWs) are placed parallel to the X axis, so that the ISW train propagates along the Y axis. We place the acoustic source at the origin. We will be interested in the dependence of the propagation on the horizontal angle χ between the X axis and the direction of the acoustic track. We then divide the horizontal plane into four quadrants. Because we are only interested in forward scattering, and due to the left-right symmetry of the forward scattering, we need consider only the first quadrant as shown in Fig. 1. The sectors that we divide the first quadrant into are denoted by different radial lines, defining an angular region between the direction of the internal wave front and a given acoustic track. They are also marked by different acronyms. We now explain the meaning of these sectors and their acronyms.

[a)]Author to whom correspondence should be addressed. Electronic mail: badiey@udel.edu
[b)]Electronic mail: katz@phys.vsu.ru
[c)]Electronic mail: jlynch@whoi.edu

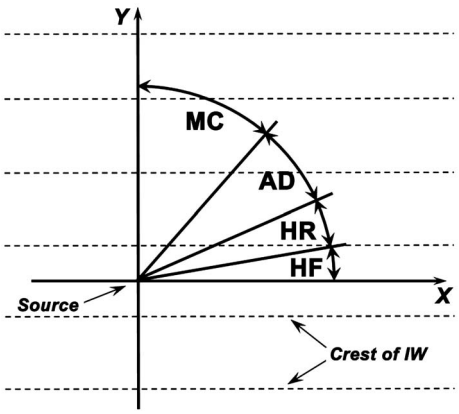


FIG. 1. Schematic diagram showing the regions of acoustic phenomenon occurring in relation to the propagation track relative to the propagation front of the internal waves.

The large sector denoted by MC is the angular region where mode coupling is expected to play a dominant role as a mechanism of sound fluctuations. In this region, the sound signals cross numerous different crests of the ISW train. A simple estimate of the amount of mode coupling one sees in this sector is determined by ratio of the mode cycle distance L (which is roughly the ray cycle distance, plus a small correction for beam displacement) to the quasi-period Λ of the ISW, taking into account the changing projection of this length with horizontal angle. We note that if $L \sin \chi / \Lambda \ll 1$, there is no appreciable mode coupling, whereas if $L \sin \chi / \Lambda \geq 1$, we will have substantial coupling. For typical conditions, $L \leq 1$ km and $\Lambda \sim 600-700$ m, so for angles $\chi < 45°$ we can suppose that mode coupling is small.

The large sector denoted by AD refers to the region where the propagation is mainly adiabatic. In this region, little mode coupling or horizontal refraction takes place.

The third sector we identify is where the effects of horizontal refraction are significant. This sector is divided into two subsectors. They are referred to as the HF (horizontal focusing) region, where the focusing of the horizontal (modal) rays can occur,[10] and the HR (horizontal refraction) sector outside of this region, where horizontal refraction still occurs (and is comparatively small) but focusing does not.

It will be shown that in the aforementioned sectors, sound fluctuations have different behavior. For example, in sector AD, the intensity fluctuations, which are caused by the vertical broadening and narrowing of the effective channel by perturbations of the thermocline level, are of the order of 1.5–2 dB. Value of fluctuations in sectors HR and HF are substantially different, as will be discussed later.

In order to clarify the above in terms of horizontal ray-vertical mode theory, we next examine the behavior of rays in the HF and HR sectors in Fig. 2. The XY plane is shown again with the internal wave crests propagating in the Y direction. Two regions are identified. A focusing ray is shown in the HF region. In the HR region, three ray rubes are shown having variable ray tube cross sections δA. For a homogeneous environment, the rays are the same at all the angles (as shown by the solid line, R1). If we consider a ray tube with cross section δA when an acoustic source is at the peak of an internal wave (corresponding to the maximum displacement

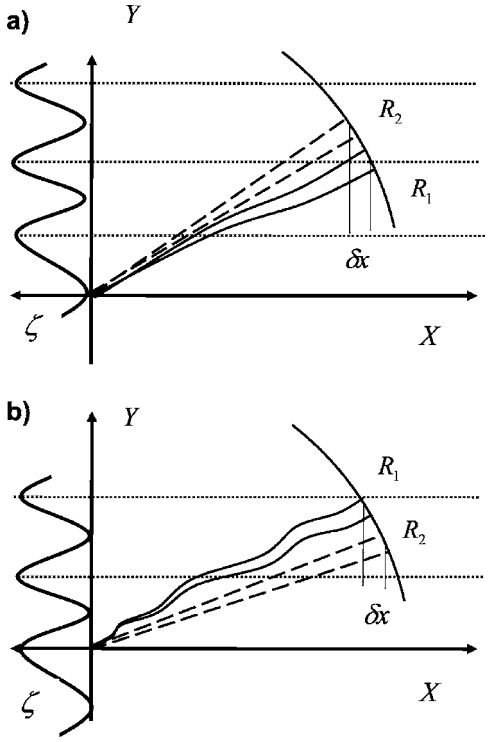


FIG. 2. Schematic diagram showing the behavior of the horizontal ray tubes for sectors HR for different positions of the IW. Positions of ISW, as well as horizontal rays corresponding to "focusing" and "defocusing" situations, are shown by solid lines in (a) and (b) panels. Dotted lines denote crests of internal waves; dashed lines denote horizontal rays in the absence of IW.

of the water layer toward the bottom), we have defocusing, whereas when an acoustic source is in a trough, we have focusing. These two situations can be described in terms of horizontal rays and their ratio can be estimated. The difference between focusing and defocusing is clearly seen in the fluctuations in the data. This was also repeated in our earlier paper.[10,11]

In this paper, a more detailed examination of the ray approximation for horizontal refraction will be presented. In addition, we will perform this analysis with an emphasis on frequency dependence. In our previous paper, we showed fluctuations of the acoustic intensity for a source having a frequency bandwidth of 30 to 160 Hz. Here we will show a wider frequency range, using additional data from an LFM signal that ranged from 50 to 200 Hz.

This paper is organized as follows. In Sec. II the theory of frequency dependent intensity fluctuations due to shallow water internal waves is presented. Then a model is presented for a shallow water channel containing internal waves with characteristics that generally correspond to the SWARM experiment wave field.[1–3] Next, the experimental data from SWARM'95 are described in detail (Sec. III) for the period of time where the frequency dependence is observed. This is followed by the analysis of these data. Finally, a summary and conclusion section is provided with recommendations for future work.

II. THEORY OF FLUCTUATIONS AND FREQUENCY DEPENDENCE

In establishing the theoretical model of sound propagation in shallow water that is used here, we follow the theory

developed by Ref. 10. We first consider the sound radiated by a point source in the presence of internal solitons in the horizontal plane, using the Weinberg-Burridge horizontal rays/vertical modes approach. We define that the wave fronts of the ISW train are directed along the X axis (Fig. 2). The ISWs are best described by their surfaces of constant density as a function of both spatial coordinates and time. In one popular approximation, the density surface value can be represented as a product of the first gravitational mode of the ISW and an envelope function:

$$\zeta(\boldsymbol{r},z,T)=\Phi(z)\zeta_s(\boldsymbol{r},T), \tag{1}$$

where $\Phi(z)$ is the first gravitational mode, normalized by $\max[\Phi(z)]=1$, and $\zeta_s(\boldsymbol{r},T)$ is the envelope of the internal waves (the displacement of the isodensity surface at the depth where Φ has a maximum). We will denote the time describing the motion of the solitons as "slow time" and use a capital T for it, in contrast to the "fast time" describing the sound pulse arrival, denoted by t. Horizontal position is denoted by $\boldsymbol{r}=(x,y)$. The solitons provide a sound speed profile perturbation free given by

$$\delta c = Qc_0(z)N^2(z)\zeta_s(\boldsymbol{r},z,T), \tag{2}$$

where $c_0(z)$ is the unperturbed sound speed profile, $N(z)$ is the buoyancy frequency, and the coefficient Q is a temperature dependent quantity having the value 2.4 m/s at 10 °C and 1.1 for 21 °C.

Our analysis will assume the sound field to be due to a broadband source with spectrum $S(\omega)$, placed at the point $(\boldsymbol{r}_s,z_s)$ in the shallow water waveguide. Using the acoustic vertical modes, one obtains for the pressure field

$$P(\boldsymbol{r},z,t)=2\int_0^\infty S(\omega)\sum_l P_l(\boldsymbol{r},\boldsymbol{r}_s)\psi_l(\boldsymbol{r},z)e^{-i\omega t}\,d\omega, \tag{3}$$

where the ψ_l are the eigenfunctions (modes) and q_l and $\gamma_l/2$ are the real and imaginary parts of the eigenvalues $\xi_l=q_l+i(\gamma_l/2)$ obtained by solving the Sturm-Liouville problem

$$\frac{d^2\psi_l(\boldsymbol{r},z,T)}{dz^2}+\left\{\frac{\omega^2}{[c_0(z)+\delta c(\boldsymbol{r},z,T)]^2}-\xi_l^2(\boldsymbol{r},T)\right\}\psi_l(\boldsymbol{r},z,T)=0 \tag{4}$$

subject to the usual boundary conditions

$$\psi_l|_{z=0}=0,\quad \psi_l|_{z=H^-}=\psi_l|_{z=H^+},$$

$$\frac{1}{\rho}\frac{d\psi_l}{dz}\bigg|_{z=H^-}=\frac{1}{\rho_1}\frac{d\psi_l}{dz}\bigg|_{z=H^+}.$$

In the above, $\boldsymbol{r}_s=(0,0)$ gives the coordinates of the source, ρ,ρ_1 are the densities of the water and bottom, respectively, and the P_l are the mode amplitude coefficients, which are critical to determining horizontal refraction or mode coupling (or both of these phenomena).

Let us examine the P_l more closely, neglecting attenuation for the moment. The P_l can be found via various different approximations. For instance, they are determined by the horizontal rays in the theory of vertical modes and horizontal rays, or in another variant satisfy the PE equation in the theory of vertical modes and PE in the horizontal plane. In a regular waveguide, at a distance far from the source (far field) $P_l(\bar{r},\bar{r}_s)=i\Psi_l,(\boldsymbol{r}_s,z_s)\exp(i\xi_l^0|\bar{r}-\boldsymbol{r}_s|)/\sqrt{8\pi\xi_l^0|\bar{r}-\boldsymbol{r}_s|}$, where superscript 0 denotes values concerned with unperturbed waveguide (without ISW). In the theory of vertical modes and horizontal rays, we take $P_l(\bar{r},\bar{r}_s)$ to be of the useful form

$$P_l(\bar{r},\bar{r}_s)=A_l(\boldsymbol{r})e^{i\theta_l(\boldsymbol{r})}. \tag{5}$$

[If we have several horizontal rays corresponding to one mode coming to one receiver point, we should write a second subindex to denote a sum over horizontal rays in (5), see Ref. 10.]

For $A_l(\boldsymbol{r})$ and $\theta_1(\boldsymbol{r})$, we get the standard eikonal and transport equations of ray theory:

$$(\nabla_r\theta_l)^2=(q_l^0)^2[1+\mu_l(\boldsymbol{r})],$$

$$2\nabla_r A_l\nabla_r\theta_l+A_l\nabla_r^2\theta_l=0. \tag{6}$$

For the correction to the effective refraction index, μ_1, we can employ an expression from perturbation theory, used in Ref. 10.

$$\mu_l=-\nu_l\zeta_s(\boldsymbol{r},T), \tag{7}$$

$$\nu_l=\frac{2Qk^2}{(q_l^0)^2}\int_0^H[\psi_l^0(z)]^2N^2(z)\Phi(z)\,dz. \tag{8}$$

In our statement of the problem, the index of refraction for horizontal rays depends only on y. This y dependence repeats in a regular fashion, as we can see from Eq. (7), via its dependence on the envelope of the ISW train. It is also seen that the coefficient ν_l (and in turn the refractive index for the horizontal rays) corresponding to the lth mode depends on frequency. The details of this frequency dependence were analyzed in Ref. 10, and, as an example, we show the frequency dependence of the horizontal refraction index (HRI) for some modes and for conditions corresponding to the SWARM'95 experiment (see Fig. 3). As a result of the index of refraction frequency dependence, we will also observe frequency dependence of the sound intensity, and in particular a frequency dependence of sound intensity fluctuations. For the SWARM'95 experiment, this frequency dependence

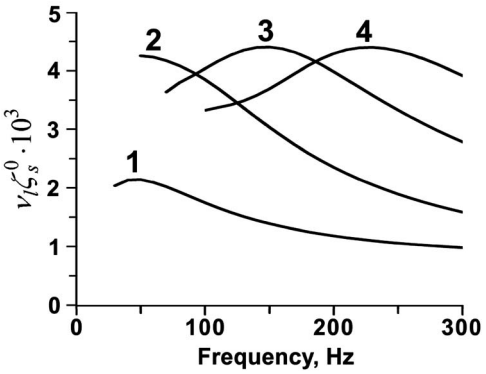


FIG. 3. Frequency dependent index of refraction for horizontal rays at maximum ISW amplitude (~10 m) calculated for the first four lowest acoustic modes.

of the intensity fluctuations was discussed in Ref. 6, where it was remarked that the shape of this dependence repeats the shape of the frequency dependence of the refraction index. Below we explain this interesting fact in more detail.

Let us consider intensity fluctuations using the familiar approximation of horizontal rays/vertical modes. This model corresponds to the one presented in Ref. 6. We assume that the internal solitons have plane wavefronts, parallel to the X axis, and moving in the Y direction at the velocity V, i.e., $\zeta_s(\boldsymbol{r},T)=\zeta_s(\boldsymbol{r}_R,T+(y-y_R)/V)$. Here $\boldsymbol{r}_R=(x_R,y_R)$ is the coordinate of the receiver in the horizontal plane. We will also assume (not strictly correctly) that the shape of the ISW envelope does not change in time, and will denote it as $\zeta_s(y,T)$. In our approach, the parameter μ_l determining the medium of propagation is a function only of y: $\mu_l=\mu_l(y)$. This greatly simplifies our consideration of the ray pattern.

In order to carry out calculations, we need to construct expressions for the horizontal rays. In our formalism, for every mode number l we have a system of horizontal rays (i.e., their trajectories). Thus rays belonging to this system will be characterized both by mode number and by their horizontal launch angle. It is useful to define the tangent vector to this trajectory; in particular, at the source this vector is $\boldsymbol{\tau}_{ls}=(\cos\chi_{ls},\sin\chi_{ls})$, where χ_{ls} is the angle of the horizontal ray with the x axis, corresponding to the lth vertical mode at the point of the source. The trajectories of the horizontal rays (HR), which are outgoing from the source [point (0,0)], can be described by an equation at the current point at the ray (x,y) as $y=y(x)$ or $x=x(y)$. We can also use other parameters and characterize rays by their so-called ray coordinates—the launch angle and the length of the raypath: (χ_{ls},s), where the parameter $ds=\sqrt{dx^2+dy^2}$ is an element of arc along the ray.

We are interested here mainly in the fluctuations of intensity of the sound field. For a broadband signal, we can introduce the spectral-modal intensity

$$\boldsymbol{I}_{l\omega}=\frac{2\pi i}{\rho\omega}|S(\omega)|^2(P_l\nabla P_l^*-P_l^*\nabla P_l). \tag{9}$$

In the approximation of a locally plane wave front $|\nabla P_l|\sim kP_l$, we obtain a simpler expression

$$\boldsymbol{I}_{l\omega}=\frac{4\pi}{\rho c}|S(\omega)|^2|P_l|^2. \tag{10}$$

In ray theory, the amplitude $|P_l|=A_l$ of the separate HRs can be calculated as

$$A_l(s)=\frac{A_l(s_0)}{\sqrt{D_l(s,s_0)}}, \tag{11}$$

where $A_l(s_0)$ is amplitude of ray at a fixed point (usually taken as near the source) and $D_l(s,s_0)$ is the divergence of the horizontal ray tube. This can be presented more elegantly through the Jacobian $J(s)=\partial(x,y)/\partial(s,\chi_{ls})$ of the transformation from Cartesian coordinates to ray coordinates $x=x(s,\chi_{ls}),y=y(s,\chi_{ls})$:

$$D_l(s,s_0)=\frac{J(s)}{J(s_0)}. \tag{12}$$

Thus the spectral modal intensity, propagating along horizontal ray, as a function of distance is

$$I_{l\omega}=\frac{4\pi}{\rho c}|S(\omega)|^2(A_l^0)^2\frac{J(s_0)}{J(s)}=I_{l\omega}^0\frac{J(s_0)}{J(s)}, \tag{13}$$

where $I_{l\omega}^0$ is some reference value of the modal intensity. Its value and the value of the Jacobian at the reference distance $J(s_0)$ do not play an essential role because, in the following considerations, we will focus on relative fluctuations of the intensity.

Due to the changing position of the source relative to the ISW train, we observe different intensity behaviors of the horizontal rays as the ISW wave train passes by. Using Fig. 2, we will next examine intensities at two limiting positions of the internal solitons, corresponding to strong focusing and defocusing of the horizontal rays. Due to the motion of the internal solitons, we observe variation of the ray pattern between these two structures. Let us first consider an area HF where the *focusing* of horizontal rays can take place, and thus fluctuations of the ray trajectories are significant. For this case, ray trajectories can be constructed numerically for any arbitrary shape of ISW. We can estimate the boundary of this area if we know the parameters of the ISW train. Let the ISWs have some amplitude ζ_0, typically $\sim$10 m, and quasiperiod Λ, typically $\sim$700 m. This simple estimate gives a limiting horizontal angle for outgoing rays $\chi_0\sim\sqrt{|\nu_l\zeta_0|}\sim 4°$. Sound fluctuations in this area (HF) can be significant and depend on the mode number, frequency, and position of the receiver—thus they cannot be simply estimated. Numerical calculations for distances $\sim$15 km (similar to the SWARM'95 experiment) give amplitude fluctuations of $\geq$7–10 dB. Concerning *defocusing*, we would remark that if position of the source corresponds to this case, then the horizontal rays go away from the area between two adjacent solitons and so in this area there is a shadow zone.

Let us next consider the HR region ($\chi>\chi_0$, with an upper boundary that we will determine later). Ray trajectories (denoted by dashed lines in Fig. 2) in this area approximately follow the straight lines, which would be seen in the absence of internal waves. We can thus construct the actual trajectory by using this small deviation to justify a perturbation approach.

Of prime importance to us here is the estimation of intensity fluctuations due to horizontal refraction. We will do this estimation using the ray theory transport equation [Eq. (13)] and calculating the divergence of the horizontal rays or, in other words, simply calculating the cross section of the two-dimensional ray tube. Let us now consider rays outgoing from the source, which, as we said, are determined by the connection between the x and y coordinates at the current point of the ray $y=y(x)$ or $x=x(y)$. We will omit the index l, assuming for now that a given HR is for a given vertical mode l. According to Snell's law

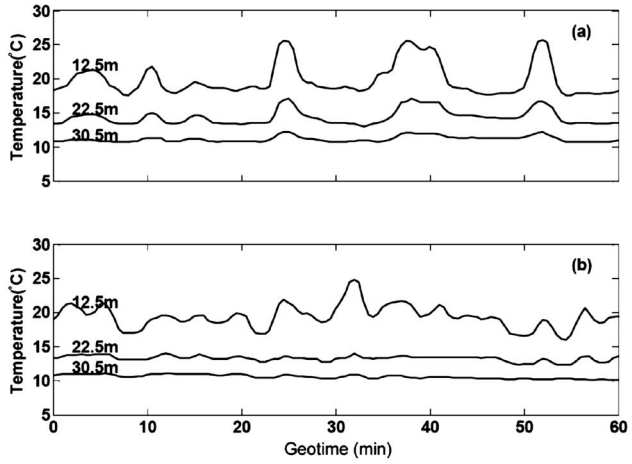


FIG. 4. Recorded temperature data at WHOI vertical array on 4 August 1995. (a) From 19:00:00 to 20:00:00 GMT (case 1). (b) From 20:00:00 to 21:00:00 GMT (case 2). Depths of the corresponding sensors are denoted near the curves.

$$\frac{\cos \chi_{ls}}{\cos \chi_l(y)} = 1 + \mu_l(y)/2 \tag{14}$$

and

$$\frac{dx}{dy} = \cot \chi_l(y), \tag{15}$$

so we have

$$x = \cos \chi_{ls} \int_0^y \frac{1}{\sqrt{\mu_l(y) + \sin^2 \chi_{ls}}}\, dy. \tag{16}$$

This should work, since for internal solitons, $|\mu_l| \approx 10^{-3}$. Thus our approximation is true for $\chi_{ls} \geq 2-3°$; for values less than this, we would need to use another approach. For small perturbations, but not small angles, $|\mu_l(y)| \ll \sin^2 \chi_{ls}$, so we have

$$x = y \cot \chi_{ls} - \frac{\cos \chi_{ls}}{2 \sin^3 \chi_{ls}} \int_0^y \mu_l(y,T)\, dy. \tag{17}$$

Using Eq. (7) we have

$$x = y \cot \chi_{ls} + \frac{\nu_l(\omega) \cos \chi_{ls}}{2 \sin^3 \chi_{ls}} \int_0^y \zeta_s(y,T)\, dy. \tag{18}$$

The second term on the right-hand side of Eq. (18) is the correction to the straight line path, which is the first term, corresponding to absence of IW [dashed lines in Figs. 2(a) and 2(b)]. Remark that in dependence on y coordinate of the source ($y=0$) with respect to the envelope, the integral on the right side of (18) can have a different sign. More exactly, if point $y=0$ corresponds to the minimum of the soliton, then the integral in (18) has positive sign and we have the $\delta x > 0$ "focusing" situation [Fig. 2(a)]. If point $y=0$ falls to the maximum of the soliton, the integral has negative sign—the defocusing situation [Fig. 2(b)].

Let us now carry out the calculations of the ray trajectory and the ray intensity (cross section of the ray tube) as influenced by horizontal refraction. Let us consider rays emitted from the source at angles greater than the critical

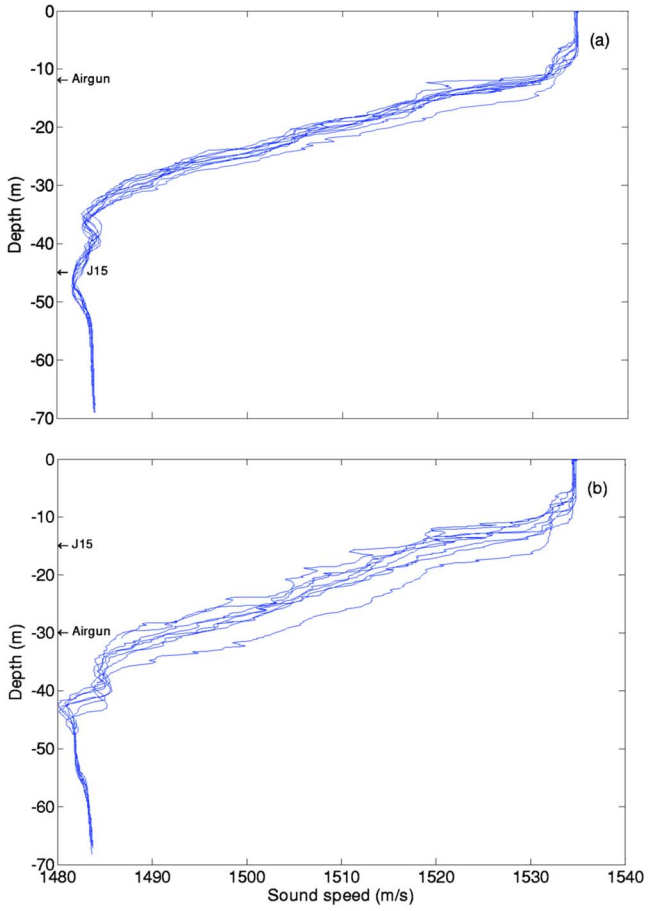


FIG. 5. (Color online) Sound speed profiles, corresponding to different geotimes at the source position, (a) case 1, (b) case 2. Depths of the sources are denoted by arrows.

angle for horizontal focusing $\chi_{ls} > \chi_0$. For this case, the approximate ray path is shown in Fig. 2. The horizontal deviation of the ray is denoted by δx. We can see that this deviation is determined by the integral of the envelope of the solitons, simply speaking the area under the curve. Let us assume the simplest shape for the solitons (a box), so that this area can be estimated as

$$\int_0^y \zeta(y,T)\, dy \sim \frac{y\zeta_0}{2} \tag{19}$$

(here we suppose that "focusing" takes place) and so

$$x = y \cot \chi_{ls} + \frac{\nu_l(\omega)\zeta_0 \cos \chi_{ls}}{4 \sin^3 \chi_{ls}} y = y \cot \chi_{ls} \left(1 + \frac{\nu_l(\omega)\zeta_0}{4 \sin^2 \chi_{ls}}\right). \tag{20}$$

Thus the ray that would go to the point R_1 in the unperturbed case arrives at R_2. Without internal waves, the signal propagates along the corresponding radius vector.

Now let the unperturbed length be determined by the parameter $s = \sqrt{y^2 + (x + \delta x)^2}$. The intensity at the receiver point without internal waves is determined only by the distance to the source and does not depend on the launch angle in the horizontal plane. The equation for the parameter s can be written in the form

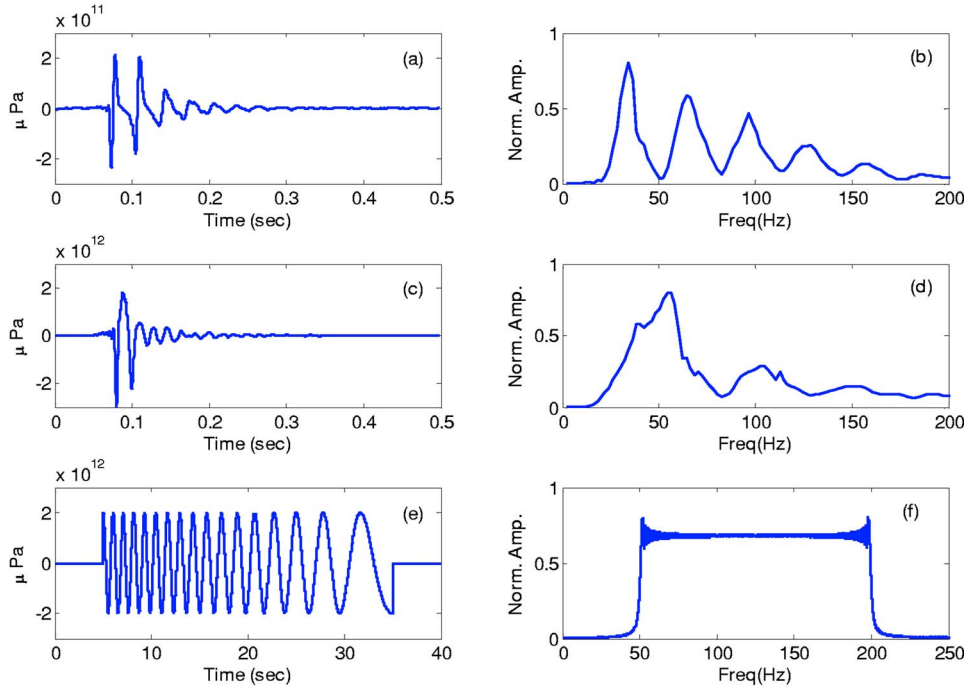


FIG. 6. (Color online) Recorded sound sources signatures measured at 1 m. (a) Airgun at 12 m depth from the sea surface. (b) Airgun spectrum at 12 m depth. (c) Airgun at 30 m depth. (d) Airgun spectrum at 30 m depth. (e) J15 at 15 m depth transmitting LFM sweep. (f) LFM signal spectrum at 15 m depth. Note that while the airgun signature is depth dependent due to the bubble pulse, the J15 signal is independent of the depth.

$$s(y) = \int_0^y \frac{ds}{dy}\, dy = \int_0^y \sqrt{\cot^2 \chi(y) + 1}\, dy$$

$$= \int_0^y \frac{1 + \mu_l(y)}{\sqrt{\mu_l(y) + \sin^2 \chi_s}}\, dy. \tag{21}$$

Using the smallness of the perturbation to the refraction index

$$s \approx \frac{1}{\sin \chi_{ls}} \int_0^y \left(1 + \frac{\mu_l}{2 \sin^2 \chi_{ls}}\right) dy = \frac{1}{\sin \chi_{ls}} y$$

$$+ \frac{\nu_l \zeta_0}{4 \sin^3 \chi_{ls}} y, \tag{22}$$

and so we get

$$y \approx s\left(\sin \chi_{ls} - \frac{\nu_l \zeta_0}{4 \sin \chi_{ls}}\right). \tag{23}$$

We can see connection between x and s. In the same approximation

$$x \approx s \cos \chi_s \tag{24}$$

and the Jacobian at the point R_2 can be expressed as

$$J(s) = \frac{\partial(x,y)}{\partial(s,\chi_{ls})} \approx s\left[1 + \frac{\nu_l \zeta_0 \cos 2\chi_{ls}}{4 \sin^2 \chi_{ls}}\right]. \tag{25}$$

We can now get the intensity in accord with Eq. (13) using the assumption of small grazing angles at the point R_2. For the case of focusing, it is decreasing (in correspondence with experimental data) and at a fixed s

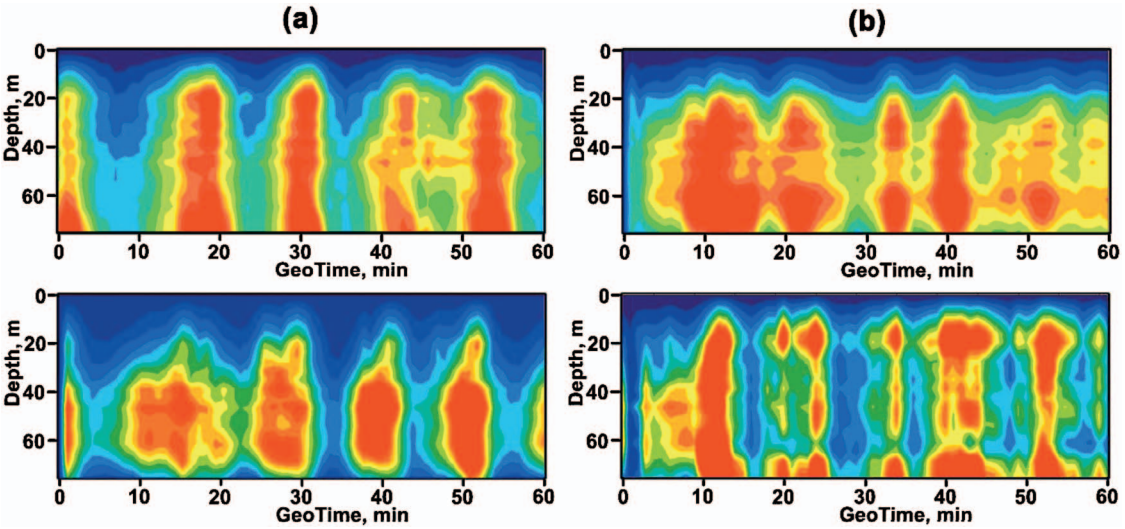


FIG. 7. Fluctuations of sound energy per pulse in normalized units for airgun (top) and J15-LFM (bottom) sources. (a) T=19:00–20:00 GMT, (b) T=20:00–21:00 GMT. Positions of the sources for the cases (a) and (b) are shown in Fig. 5.

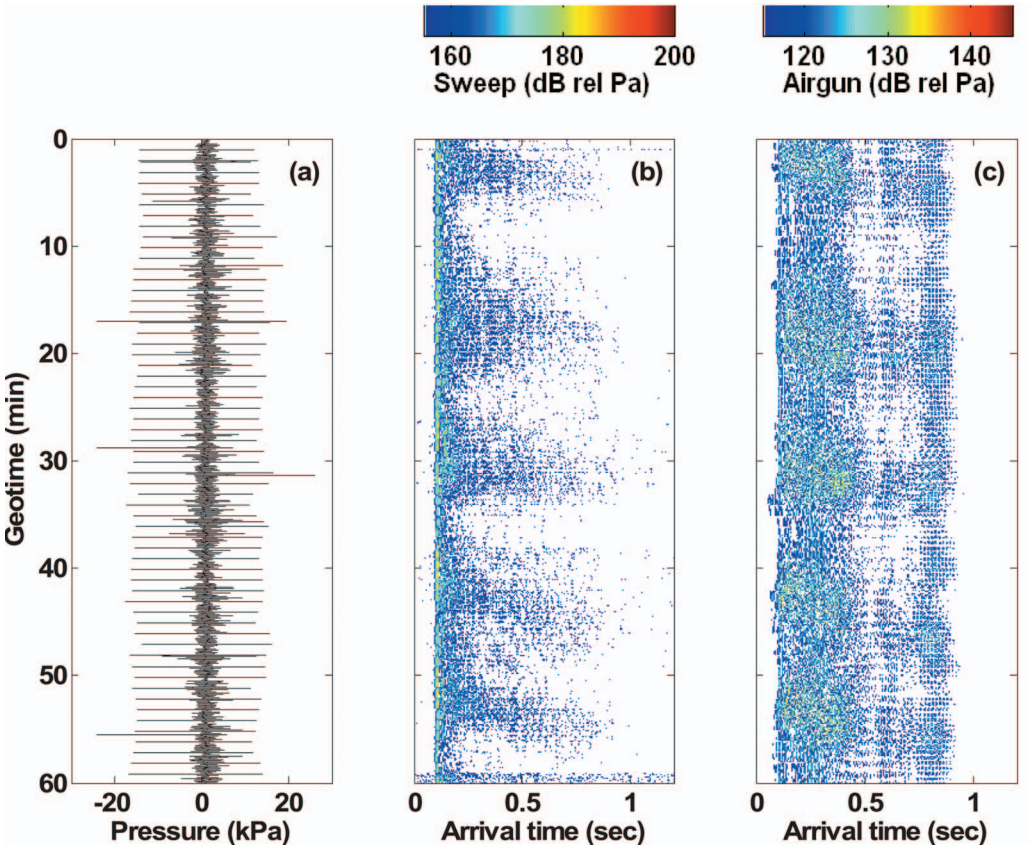


FIG. 8. Acoustic intensity versus geotime, T measured by the WHOI VLA for case 1. (a) Airgun and LFM signatures, (b) intensity of the received airgun source when placed 12 m below the sea surface (above thermocline), and (c) LFM source when placed below thermocline at 45 m below the sea surface.

$$I_{l\omega}^{f} = I_{l\omega}^{0} J(s_0)\left(1 + \frac{\nu_l(\omega)\zeta_0}{4\sin^2\chi_{ls}}\right)^{-1}. \tag{26}$$

For defocusing, the divergence of rays is decreasing, and so the intensity is correspondingly increasing and proportional to the value

$$I_{l\omega}^{d} = I_{l\omega}^{0} J(s_0)\left(1 - \frac{\nu_l(\omega)\zeta_0}{4\sin^2\chi_{ls}}\right)^{-1}. \tag{27}$$

In particular, for solitons with amplitude $\sim$15 m, a value of $\nu_l \sim 5-7\times10^{-4}$ and grazing angle $\sim 0.1(6^0)$, we have $(5-7\times10^{-4}\times15)/4\times0.01\sim0.2-0.3$ and so the ratio of intensities is $I^f/I^d\sim1.8-2$.

Due to the motion of the ISWs, a fixed source and receiver fall into positions first corresponding to the focusing of horizontal rays and then corresponding to defocusing—thus we observe temporal fluctuations of the received signals. We next estimate these fluctuations as a function of the parameters of the problem.

As to the frequency dependence of the fluctuations, this can be seen from the expression for the correction $\delta I_{l\omega}^2 = I_{l\omega}^2 - \langle I_l\omega\rangle^2$. We will consider the value of the scintillation index SI (Ref. 12) describing relative fluctuations of intensity between maximal (defocusing) and minimal (focusing) values to be:

$$\mathrm{SI}_{l\omega}^2 = \left\langle \frac{\delta I_{l\omega}^2}{\langle I_{l\omega}\rangle^2} \right\rangle = \frac{\langle I_{l\omega}\rangle^2 - \langle I_{l\omega}\rangle^2}{\langle I_{l\omega}\rangle^2}. \tag{28}$$

Here $\langle F\rangle = 1/T\int_0^T F\,dT$. We note that in parts of the literature, the SI is sometimes defined without the squared expression. We are following the (squared) convention of Ref. 12 in this work.

Thus, for our case, the scintillation index is

$$\mathrm{SI}_{l\omega}^2 \sim \frac{\nu_l(\omega)\zeta_0}{\sin^2\chi_s}. \tag{29}$$

According to this equation, the frequency dependence of the scintillation index for modal intensity repeats the shape of the frequency dependence of the horizontal refraction index for the same mode. This dependence is shown in Fig. 3.

In the following sections, we will analyze experimental results for SI, obtained in the SWARM'95 experiment, and compare to the theory developed.

III. DATA ANALYSIS AND COMPARISON WITH THEORY

To test the theory presented in previous sections, we show here an analysis of data obtained in the SWARM-95 experiment, to compare with the theory. Since several papers have previously outlined details of these measurements,[1–3] we do not describe the minutiae of the experiment. Here we only describe 2 h of collected data during 4 August 1995.

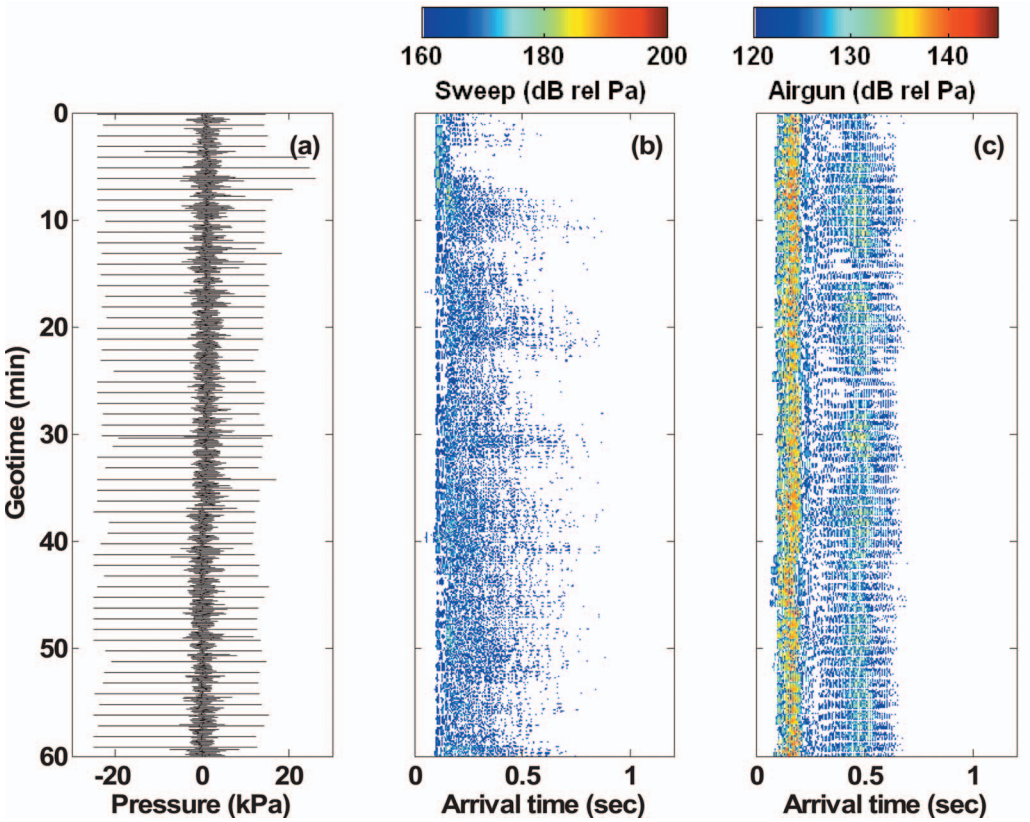


FIG. 9. Acoustic intensity versus geotime, T measured by the WHOI VLA for case 2. (a) Airgun and LFM signatures, (b) intensity of the received airgun source when placed 30 m below the sea surface (below the thermocline), and (c) LFM source when placed above thermocline at 15 m below the sea surface.

During each hour, two acoustic sources having different frequency bands (an airgun at 30–150 Hz, and a J15-3 transducer at 50–200 Hz) were used to simultaneously produce sound in the waveguide. One of the acoustic sources was placed above the thermocline while the second one was placed below. The airgun source fired every 60.03 s while the J15 was transmitting an LFM signal of duration 30 s in between two consecutive airgun shots. The data are presented in two geotime segments—one segment from 19:00 to 20:00 GMT that we call case 1, and the second segment from 20:00 to 21:00 GMT called case 2.

Figure 4 shows thermistor string records for these two time segments at the receiver array. From the six thermistors that were deployed along the vertical line array of the Woods Hole Oceanographic Institution (WHOI-VLA) only the upper water column shows the ISW activity, hence data for depths of 12.5, 22.5, and 30.5 m below the sea surface are only shown here. The temperature records show strong variability in the upper water column, with a decreasing trend at lower depths. In addition to these temperature records, the sound speed profile was measured using a CTD at the source location at 15-min intervals during the propagation. Although not shown here, these data were used in order to get approximate values for the salinity profiles, which were rather constant during this period. Figure 5 shows the sound speed profiles at the source for the 2-h time intervals with the source positions designated at different depths. During the first hour, the airgun was placed at 12 m from the sea surface while the J15 was at 45-m depth. During the second hour, the airgun was at 40 m depth while the J15 was placed at 15-m depth. We note that there is some change in the spectrum of the airgun due to the depth change, but not the J15.

In Fig. 6 we present the source signals measured about 1 m from the source position in both cases. It is noted that the signals are very different from each other. While the airgun signal has a time domain duration of ~0.4 s and energy up to 150 Hz, most of its low frequency energy components peak around 30 Hz. The J15-3 generated a 30-s LFM signal in time domain and its energy is between 50 and 200 Hz.

In order to address the data analysis based on our theory, we first show the temporal dependence of the intensity integrated over the pulse duration. This dependence shows the fluctuation of signals from the acoustic sources for each geotime segment at the WHOI receiver array.[2,6] Figure 7(a) shows the first geotime segment (T=19:00–20:00 GMT) when the airgun was placed at 12 m below the sea surface, and the J15 was placed at 45 m [see Fig. 5(a)]. The LFM signals presented are seen after matched filter processing and the airgun signal being bandpassed filtered between 20 and 300 Hz. In Fig. 7(b) the second geotime segment (T=20:00–21:00 GMT) is shown, when the J15 source is at 15–m and the airgun is at 30-m depth, respectively. It is noticed that the behavior of the waveguide is very similar regardless of the location of the sound source in the water column and for the frequencies of the different sources. Similar results have been reported earlier to relate the intensity fluctuations of the sound signals to the index of refrac-

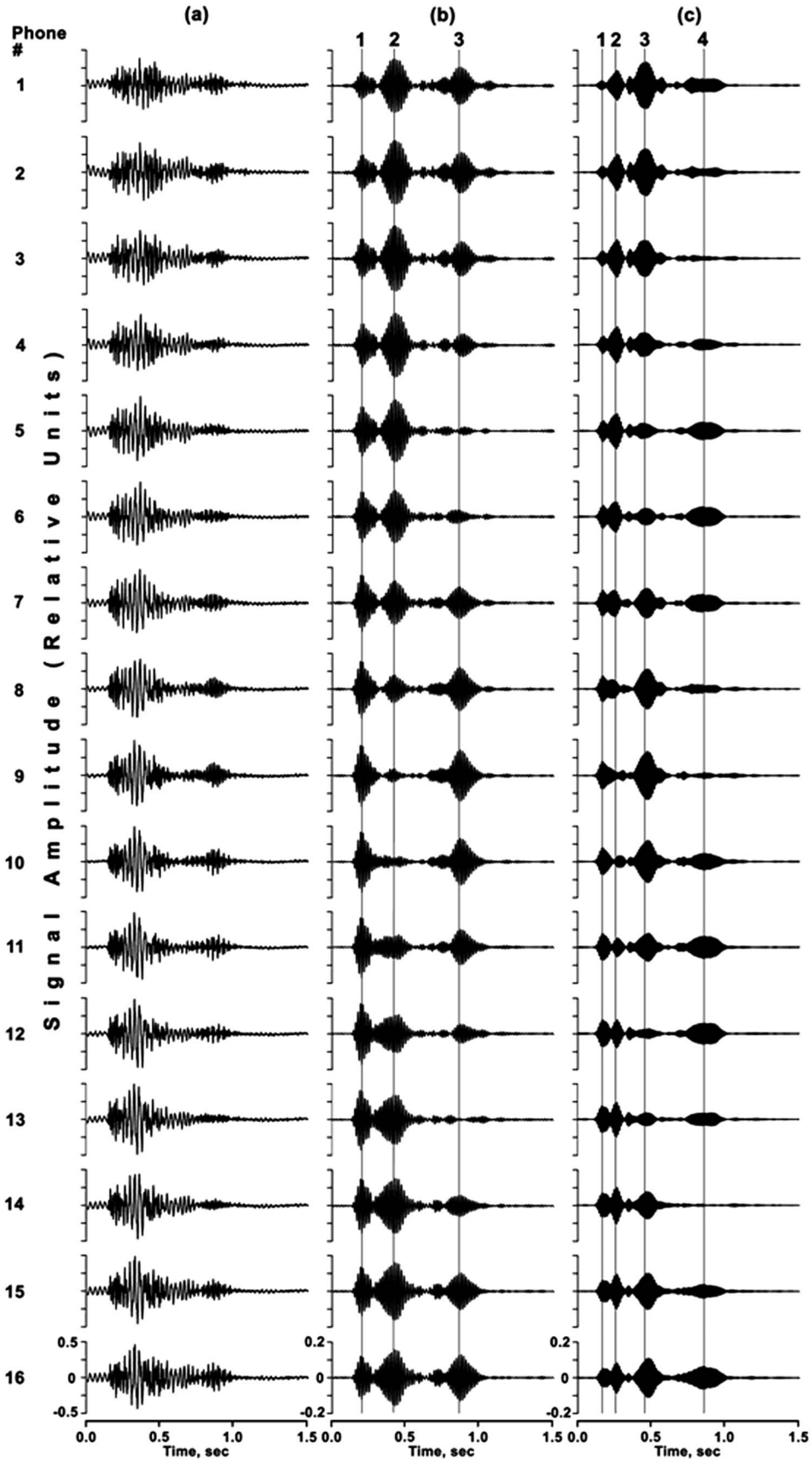


FIG. 10. Received airgun signal for $T=-19:15$ GMT for all 16 WHOI hydrophones. (a) Raw data. (b) Bandpass filter with fc=60 Hz and BW=±10 Hz. (c) Bandpass filter with fc=90 Hz and BW=±10 Hz. Dashed lines indicate arrival times of modes 1–4.

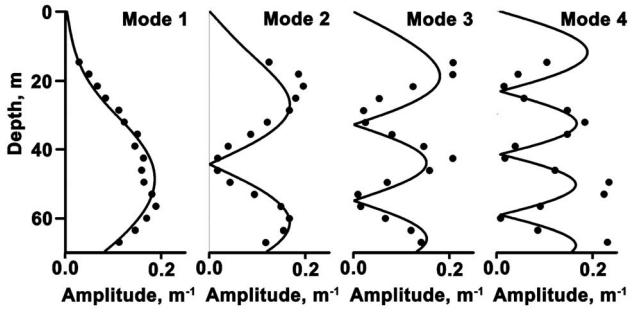


FIG. 11. Depth dependence of the first four mode functions for $T=19:00$–$20:00$ GMT. Dots show averaged value of experimental data and the solid lines show the theoretical calculations based on averaged sound speed profiles for the same time.

tion. Here, we present a more through formalism in the context of experimental verification of the theory.

To further investigate the experimental data, we present the received signals for cases 1 and 2 as a summary of the VLA in the following manner. We compress the received signals for all the elements in the WHOI vertical array into a single point and then stack these points in geotime for each segment of the 2-h period (i.e., cases 1 and 2 shown in Figs. 4 and 5, respectively). As a result, Figs. 8 and 9 are obtained. In each of these figures, three panels are shown. In Fig. 8(a) the source signature measurement near the source is displayed as a function of geotime for case 1. Spikes show the airgun shots and in between; the darker color signals with lower amplitude and longer duration show the LFM sweeps. In Figs. 8(b) and 8(c) the squeezed and stacked array results as a function of geotime are displayed for the source placed above and below the thermocline [Fig. 8(b) shows the received signal at the vertical array for the LFM at 45 m and Fig. 8(c) shows the received signal at VLA for the airgun at 12 m]. Hence, Fig. 8 is a summary of the simultaneous channel response function for the channel activated above and below the thermocline for 1 h (i.e., case 1 from 19:00 to 20:00 GMT). Similarly, Fig. 9 shows the same results for the second hour of our data with the corresponding source positions shown in Fig. 5 (i.e., case 2 from 20:00 to 21:00 GMT).

Several points are noteworthy in these two figures. The most significant feature of these data is their intensity fluctuations as a function of geotime. We see that the number of oscillations of intensity and the typical temporal period between adjacent fluctuations are approximately the same in Figs. 8 and 9 on one side and in Fig. 4 on the other side. It indicates a direct relationship of the sound fluctuations with the internal waves. The duration of the received signal in the second hour [case 2 shown in Figs. 9(b) and 9(c)] is much shorter compared to the first hour [case 1 shown in Figs. 8(b) and 8(c)]. Also as expected, when the source is placed below the thermocline, the received signal depicts a much stronger intensity [Figs. 8(c) and 9(c)].

Next, the temporal dependence of the intensity contained in separate modes per unit frequency interval [called spectral-modal density $I_{l\omega}$ in Eq. (10)] is extracted from the experimental data and is compared with the scintillation index (29). A pulse signal received by vertical array can be shown as $p(\boldsymbol{r},z,t)$, where $\boldsymbol{r}$ is horizontal coordinate of the

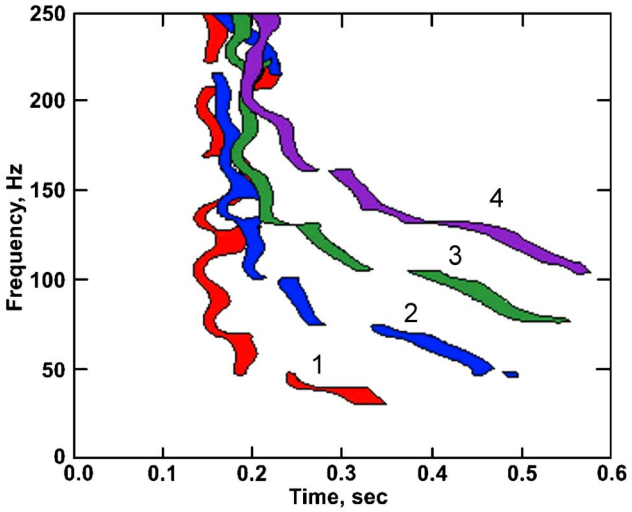


FIG. 12. (Color online) Time-frequency diagram for an airgun pulse at $T=19:15$ GMT showing the first four modes obtained from the experimental data. Modes 1 through 4 are numbered on the curves.

receiver and z is the hydrophone depth. Since only a discrete number of hydrophones exist, the sound pressure is known only at fixed depths z_j and ranges $\boldsymbol{r}_j$. However, for simplicity, in the following analysis we omit the index. The spectrum of the received signal $S(\boldsymbol{r},z,\omega)$ also depends on the hydrophone and analytic (complex) received signal, and has the form

$$P(\boldsymbol{r},z,t)=2\int_0^\infty S(\boldsymbol{r},z,\omega)e^{-i\omega t}\,d\omega,$$

$$S(\boldsymbol{r},z,\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{Re}[P(\boldsymbol{r},z,t)]e^{i\omega t}\,dt. \tag{30}$$

According to our theory, it can be decomposed over vertical modes

$$S(\boldsymbol{r},z,\omega)=\sum_l S_l(\boldsymbol{r},\omega)\psi_l(z), \tag{31}$$

where the value $S_l(\boldsymbol{r},\omega)=S(\omega)P_l(\boldsymbol{r},\omega)$ is called the spectral modal amplitude.

The first step in a modal analysis of the experimental data is to compare the depth dependence of the amplitudes (i.e., the modulus) of the theoretical and experimental modes. To obtain the depth dependence of the experimental modes, we use the difference between the group velocities of different modes $v_l^{gr}(\omega)$. In other words we cut a narrow frequency interval from the total band of the spectrum $S(\omega)$. This can be achieved by using integration with narrow band filtering function $F(\omega,\omega_c)$.

In the following analysis we will use a Gaussian form for $F(\omega,\omega_c)$, where ω_c is the middle of the frequency range of width $\Delta\omega$, $\Delta\omega\ll\omega_c$:

$$F(\omega,\omega_c)=\frac{1}{\Delta\omega\sqrt{2\pi}}\exp\left\{-\frac{(\omega-\omega_c)^2}{2\Delta\omega^2}\right\}. \tag{32}$$

For a narrow frequency band, the time dependence of the received signal can be expressed as

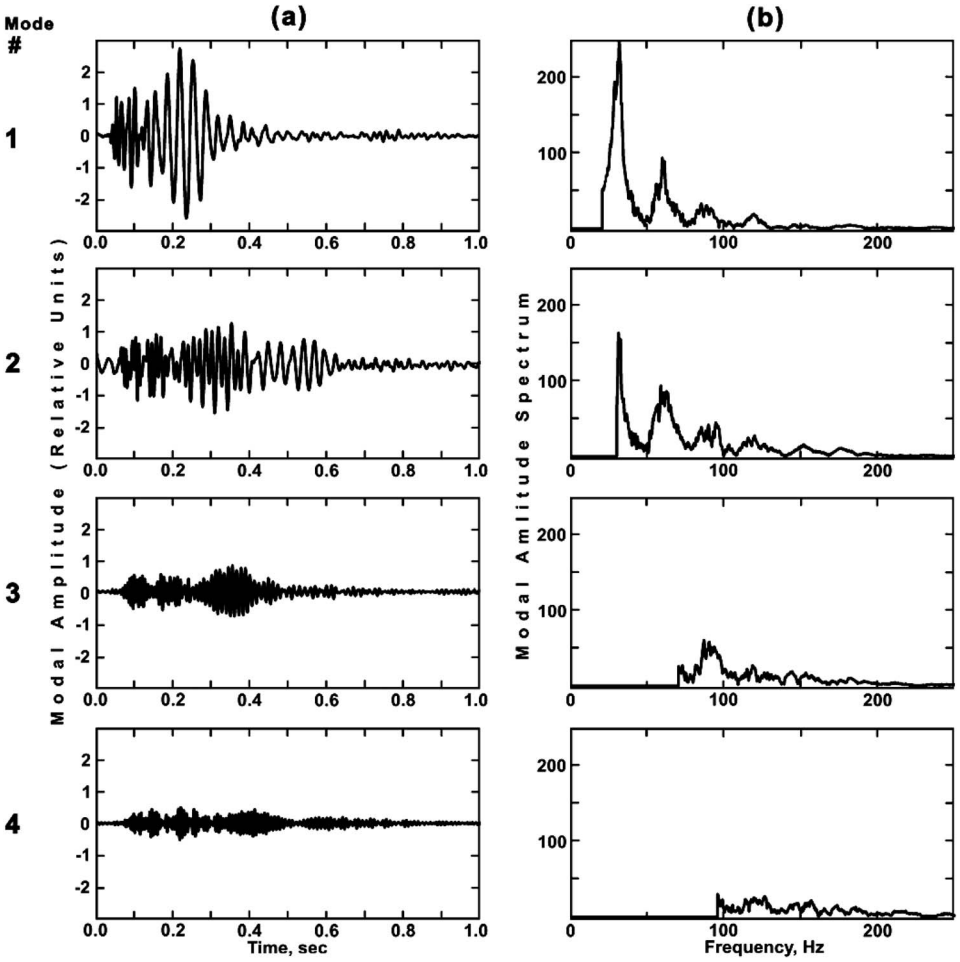


FIG. 13. Filtered modes 1–4 for $T=19:15$ GMT. (a) Relative mode amplitude. (b) Modal amplitude.

$$\tilde{P}(\boldsymbol{r},z,t,\omega_c) = 2\int_0^\infty F(\omega,\omega_c)S(\boldsymbol{r},z,\omega)e^{-i\omega t}\,d\omega \tag{33}$$

where the tilde ($\sim$) denotes the filtered experimental values for fixed center frequency ω_c. The results of the frequency filtering for $\Delta f=10$ Hz (i.e., $\Delta\omega=2\pi\Delta f$) for geotime $T=19:15$ GMT and $\omega_c=60$, 90 Hz, respectively, for different hydrophones are presented in Fig. 10. The maxima in arrival times show the arrivals of different modes due to differences in group velocities, which can be used for mode filtering. We then split arrival times of pulses into intervals $(t_l,t_l+\Delta t)$, where $t_l=L/v_l^{gr}$ is the arrival time of mode l, and Δt is the duration of the pulse. After integration of the pulse envelope (amplitude) we get a function of depth and frequency ω_c, which is proportional to the amplitude of mode l, denoted as $\tilde{\psi}_l(z,\omega_c)$:

$$\tilde{\psi}_l(z,\omega_c) = \int_{t_l}^{t_l+\Delta t} |\tilde{p}(\boldsymbol{r},z,t,\omega_c)|\,dt. \tag{34}$$

Consequently, we receive modulus of the first, second, and third modes, respectively.

Next we should compare the eigenfunctions obtained from the experimental data with the theoretically calculated values. For this we should normalize the selected modes using some relationship. To normalize experimental functions to compare with the analytical ones, we use

$$\boldsymbol{\Psi}_l(z,\omega_c) = \tilde{\psi}_l(z,\omega_c) \Big/ \int_0^H \left[\tilde{\psi}_l(z,\omega_c)\right]^2 dz. \tag{35}$$

The results of mode filtering the data shown in Fig. 10(c) are presented in Fig. 11.

Excellent agreement between experimentally extracted and theoretically calculated modes is obtained; hence, for the next step in our analysis the calculated modes (i.e., the eigenfunctions of Sturm-Liouville theory) are used rather than filtering the experimental data as shown above.

Next the frequency dependence of modal amplitudes is obtained. Modes represented by $\psi_l(z,\omega)$ are orthogonal to each other and the integral over variable z is extended from zero to infinity (including the sea bottom). However, the contribution of the sea bottom to the modal functions is small, particularly for the lower order modes. We can thus approximate

$$\int_0^H \psi_l(z,\omega)\psi_m(z,\omega)\,dz \approx \delta_{nm}, \tag{36}$$

where δ_{mn} is the Kronecker delta function. According to (31) and (36), the spectral amplitude of the lth mode can be ob-

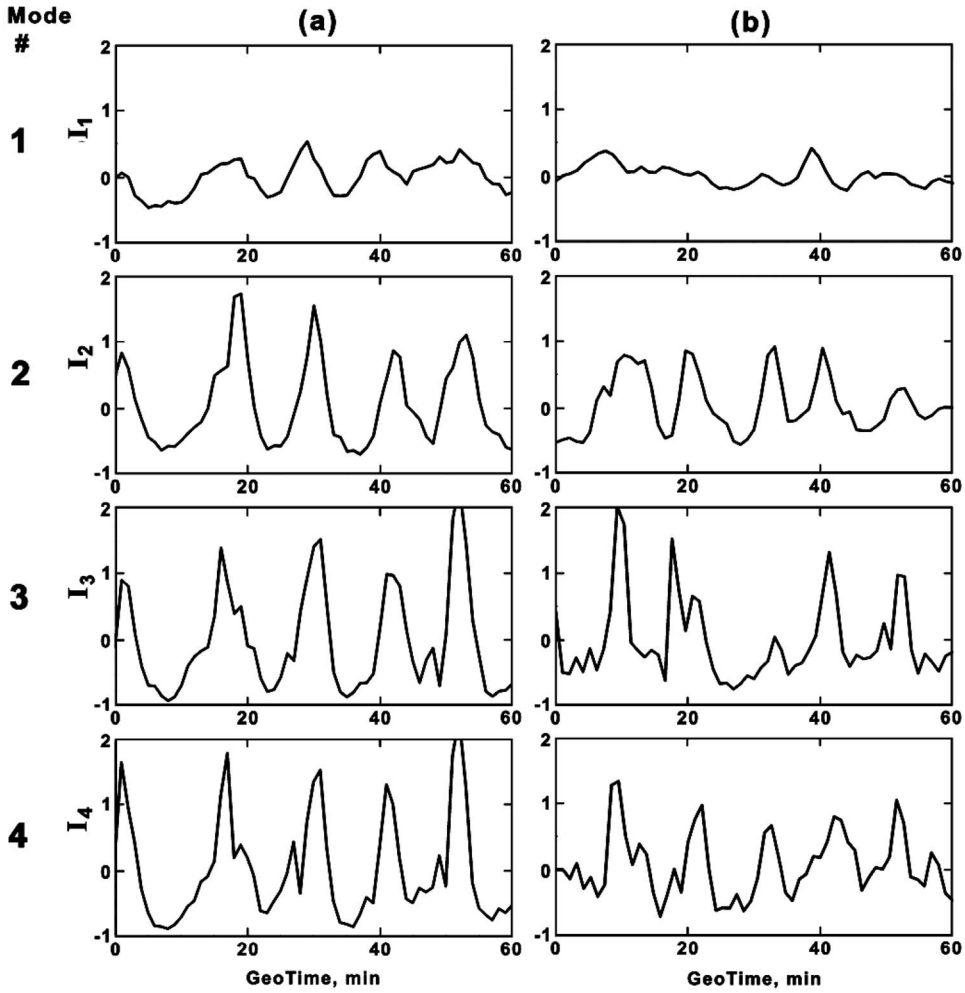


FIG. 14. Relative intensity fluctuations of modes 1–4 obtained from experimental data for (a) $T=19:00$–$20:00$ GMT and (b) $T=20:00$–$21:00$ GMT.

tained by integration of the spectrum of the received signal as a function of depth:

$$S_l(\boldsymbol{r},\omega)=\int_0^H S(\boldsymbol{r},z,\omega)\psi_l(z,\omega)\,dz. \tag{37}$$

The time-frequency diagram for different hydrophone depths of the vertical line array corresponds to the modal decomposition of filtered sound pressure (33) at the receiver:

$$\widetilde{P}(\boldsymbol{r},z,t,\omega_c)=\sum_l \psi_l(z)\widetilde{P}_l(\boldsymbol{r};t,\omega_c),$$

$$\widetilde{P}_l(\boldsymbol{r};t,\omega_c)=2\int_0^\infty F(\omega,\omega_c)S_l(\boldsymbol{r},\omega)e^{-i\omega t}\,d\omega. \tag{38}$$

The value $|\widetilde{P}_l(\boldsymbol{r};t,\omega_c)|$ is shown in Fig. 12. The irregular shape of the time-frequency curves is the result of the source spectrum [see Figs. 6(b) and 6(d)]. These "hyperbolalike" curves coincide with the theoretically obtained values (not shown in this plot) calculated using the aforementioned model. Figure 13 shows the arrival time dependence of modal amplitudes resulting from mode filtering and the corresponding modal spectra for modes 1–4. We note that the modal amplitude reduces as we go from mode 1 to 4 while the cutoff frequency increases. The initial (or radiated) amplitudes of the separate modes are determined by the source position in the water depth. The modal intensity as a function of frequency ω for geotime T is

$$I_{l\omega}(T)=\frac{4\pi}{\rho c}|S_l(\boldsymbol{r},\omega)|^2. \tag{39}$$

Dependence of modal intensity on geotime for fixed frequency (in this case 90 Hz) is shown in Fig. 14; similar curves have been constructed for another frequency in the frequency band of our sources, not shown here. For each curve we can calculate the value of fluctuations for a given geotime interval (e.g., 2 h here) and thus the corresponding scintillation index. The frequency dependence of modal intensity fluctuations through the scintillation index (29) can be obtained, and for this data are shown in Fig. 15. These dependencies, as we can see from Fig. 15, have irregular behavior as a function of frequency. Thus we will consider the smoothed dependence obtained by

$$SI_l^2(\omega)=\frac{\langle \delta I^2(\omega,T_i)\rangle}{\langle I(\omega,T_i)\rangle^2}, \tag{40}$$

where $\langle\rangle$ denotes an averaged value in time. Finally, in Fig. 16 we present these smoothed curves of the scintillation index for two independent geotime periods, for cases 1 and 2.

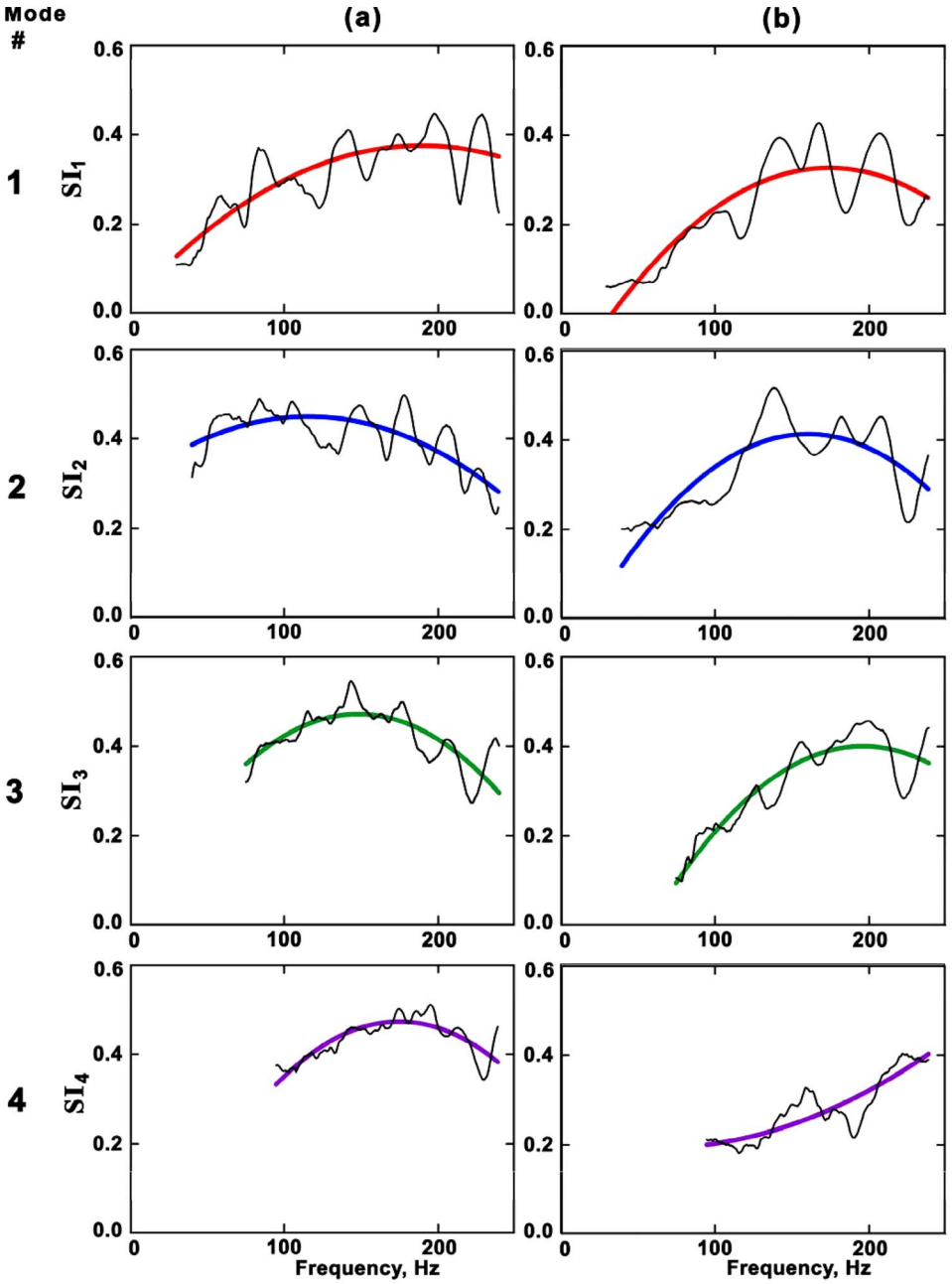


FIG. 15. (Color online) Frequency dependent modal scintillation index, SI (f) of modes 1–4 obtained from experimental data for (a) $T=19:00$–$20:00$ GMT and (b) $T=20:00$–$21:00$ GMT. The data are shown by oscillating curves and averaged values (smooth lines).

IV. SUMMARY AND CONCLUSIONS

A theory of sound propagation through an anisotropic shallow water environment is presented to examine the frequency dependence of the scintillation index in shallow water in the presence of internal waves. The theory of horizontal rays and vertical modes is used to establish the azimuthal behavior of the intensity fluctuations of the broadband acoustic signals propagating through shallow water internal waves.

In an earlier paper[6] we have shown that the temporal variations of depth dependence of intensity of the sound field resulted from horizontal anisotropy (referred to as the violation of circular symmetry) taking place in the presence of solitary internal waves. In addition, we noted the similarity of frequency dependence of modal scintillation index (for temporal fluctuations of intensity) and the index of horizontal refraction of the corresponding modes. We hypothesized that there is a link between the two.

In this paper we have extended our consideration of the nature of fluctuations of the sound field in directional dependence of propagation relative to the wave front of the ISW. Depending on the angle between the direction of the internal wave front and the acoustic track, the mechanism of sound field fluctuations is characterized by either horizontal refraction (HR) or horizontal focusing (HF), is adiabatic (AD), or is a mode coupling mechanism (MC) as shown in Fig. 1.

In this paper we have presented a theory and corresponding experimental results for the sector HR in Fig. 1, where fluctuations are caused by the effect of horizontal re-

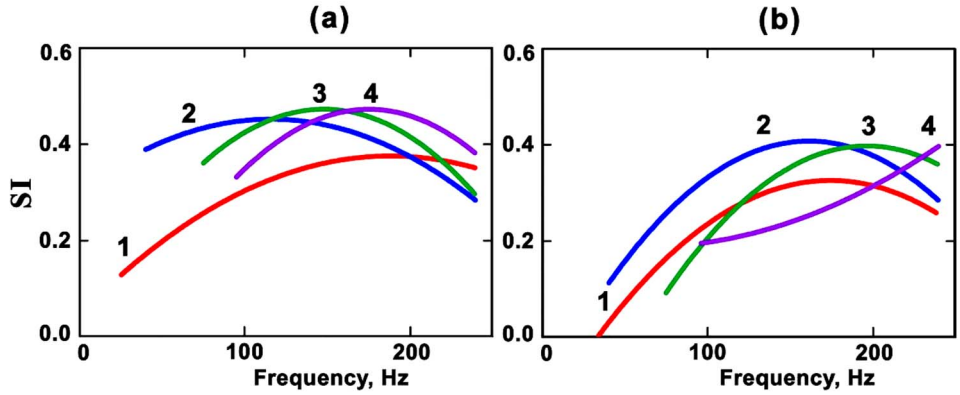


FIG. 16. (Color online) Frequency dependence of modal scintillation index SI (f) for (a) $T=19:00$–$20:00$ GMT and (b) $T=20:00$–$21:00$ GMT.

fraction or (more generally) by redistribution of the sound field in the horizontal plane. Using a ray approximation in the horizontal plane, in this paper we derived an analytical expression for the scintillation index of a sound field due to moving internal waves in shallow water.

The main feature of the frequency dependence of the modal scintillation index (29) is the shape of this dependence shown in Fig. 16. This figure, which is obtained from the experimental data and corresponding to theoretical Fig. 3, shows the relationship between the modal scintillation index, SI, and the parameters of the waveguide. We further see that this does not depend on the amplitude and shape of the internal waves.

Note that factor $\zeta_0/\sin^2\chi_s$ has a constant value for a given sequence of sound pulses. In this connection the value of the horizontal refraction index plays the role of an invariant characteristic of a definite geographical shallow water area, and one that determines the fluctuations of the sound field in the corresponding conditions (i.e., HR sector in horizontal plane).

A good agreement between the experimental and analytical results in this paper, in spite of the idealized modeling of the internal waves, is a confirmation of the fact that the scintillation index (SI) is independent of (or weakly dependent on) the parameters of the internal waves such as amplitude, period, and wave number. We note further that the comparison between the experimental and theoretical values is very good for the first hour [Fig. 16(a)] when there is a distinct internal wave passing through, while it is less good for the second hour [Fig. 16(b)], where the amplitude of the internal waves has considerably reduced. Hence, the anisotropic properties of the shallow water channel are less affected by the more isotropic background internal waves.

ACKNOWLEDGMENTS

This work was supported by the Ocean Acoustics Program (321 OA) of the Office of Naval Research (ONR Grant No. N00014-01-1-0114 to UD and N00014-04-10146 to WHOI) and by the Russian Foundation for Basic Research (RFBR Grant No. 06-05-64853-a). Funding from Delaware Sea Grant program was provided to support graduate students at the University of Delaware.

[1] J. R. Apel, M. Badiey, C.-S. Chiu, S. Finette, R. H. Headrick, J. Kemp, J. F. Lynch, A. E. Newhall, M. H. Orr, B. H. Pasewark, D. Tielburger, A. Turgut, K. von der Heydt, and S. N. Wolf, "An overview of the SWARM 1995 shallow-water internal wave acoustic scattering experiment," IEEE J. Ocean. Eng. **22**, 465–500 (1997).

[2] M. Badiey, Y. Mu, J. F. Lynch, J. R. Apel, and S. N. Wolf, "Temporal and azimuthal dependence of sound propagation in shallow water with internal waves," IEEE J. Ocean. Eng. **27**, 117–129 (2002).

[3] R. H. Headrick, J. F. Lynch, Acoustic normal mode fluctuation statistics in the 1995 SWARM internal wave scattering experiment.," J. Acoust. Soc. Am. **107**, 201–220; and the SWARM group, "Modeling mode arrivals in the 1995 SWARM experiment acoustic transmissions," J. Acoust. Soc. Am. **107**, 220–236 (2000).

[4] D. Rubenstein and M. N. Brill, "Acoustic variability due to internal waves and surface waves in shallow water," in *Ocean Variability and Acoustics Propagation*, edited by J. Potter and A. Warn-Varnas (Kluwer Academic, Dordrecht, 1991), pp. 215–228.

[5] D. Rubenstein, "Observations of cnoidal internal waves and their effect on acoustic propagation in shallow water," IEEE J. Ocean. Eng. **24**, 346–357 (1999).

[6] M. Badiey, B. Katsnelson, J. Lynch, S. Pereselkov, and W. Siegmann, "Measurement and modeling of 3-D sound intensity variations due to shallow water internal waves," J. Acoust. Soc. Am. **117**, 613–625 (2005).

[7] H. Weinberg and R. Burridge, "Horizontal ray theory for ocean acoustics," J. Acoust. Soc. Am. **55**, 63–79 (1974).

[8] S. D. Frank, M. Badiey, J. F. Lynch, and W. L. Siegmann, "Analysis and modeling of broadband airgun data influenced by nonlinear internal waves," J. Acoust. Soc. Am. **116**(6), 3404–3422 (2004).

[9] S. D. Frank, M. Badiey, and W. L. Siegmann, "Experimental evidence of three-diemensional acoustic propagation caused by nonlinear internal waves," J. Acoust. Soc. Am. **118**(2), 723–734 (2005).

[10] B. G. Katsnelson and S. A. Pereselkov, "Low-frequency horizontal acoustic refraction caused by internal wave solitons in a shallow sea," Acoust. Phys. **46**, 684–691 (2000).

[11] B. G. Katsnelson and S. A. Pereselkov, "Space-frequency dependence of the horizontal structure of the sound field in the presence of intense internal waves," Acoust. Phys. **50**, 169–176 (2004).

[12] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978), Vol. **2**.

Experimental detection and focusing in shallow water by decomposition of the time reversal operator

Claire Prada,[a] Julien de Rosny, Dominique Clorennec, Jean-Gabriel Minonzio, Alexandre Aubry, and Mathias Fink
Laboratoire Ondes et Acoustique, ESPCI, 75005 Paris, France

Lothar Berniere, Philippe Billand, Sidonie Hibral, and Thomas Folegot[b]
Atlantide, Technopole Brest Iroise, 29200 Brest, France

(Received 27 October 2006; revised 14 May 2007; accepted 18 May 2007)

A rigid 24-element source-receiver array in the 10–15 kHz frequency band, connected to a programmable electronic system, was deployed in the Bay of Brest during spring 2005. In this 10- to 18-m-deep environment, backscattered data from submerged targets were recorded. Successful detection and focusing experiments in very shallow water using the decomposition of the time reversal operator (DORT method) are shown. The ability of the DORT method to separate the echo of a target from reverberation as well as the echo from two different targets at 250 m is shown. An example of active focusing within the waveguide using the first invariant of the time reversal operator is presented, showing the enhanced focusing capability. Furthermore, the localization of the scatterers in the water column is obtained using a range-dependent acoustic model.
 [DOI: 10.1121/1.2749442]

PACS number(s): 43.30.Gv, 43.30.Vh, 43.60.Tj [DRD] Pages: 761–768

I. INTRODUCTION

Time reversal focusing in a waveguide using a source receiver array (SRA) was demonstrated in an ultrasound experiment (Fink, 1997). Taking advantage of the multiple reflections at the waveguide interfaces, time reversal allows high resolution focusing. This property was proven by several underwater time reversal experiments starting with the one by Kuperman and his team (Kuperman *et al.*, 1998). Then, the strong potential of time reversal techniques for underwater communication in shallow water was demonstrated (Edelmann *et al.*, 2002). High resolution offered by time reversal has also been exploited for detection and separation of scatterers in an ultrasonic waveguide using the decomposition of the time reversal operator (DORT method). This method is a scattering analysis technique derived from the study of the iterative time reversal process (Prada and Fink, 1994). It was applied in an ultrasonic water waveguide with a flat rigid bottom, demonstrating multitarget detection and selective focusing with high resolution. The resolution of this method was used to separate the signal reflected by two close scatterers and then focus selectively at any of them (Mordant *et al.*, 1999). It was then the object of several studies for ocean applications (Lingevitch *et al.*, 2002; Yokoyama *et al.*, 2001). Its ability to separate the echo of a target from bottom reverberation was shown in a laboratory experiment (Folegot *et al.*, 2003).

Recently, the DORT method has been tested at sea with a vertical SRA and using an echo repeater to simulate the target response (Gaumond *et al.*, 2006). The signal transmitted back to the SRA provided by an echo repeater offers much higher signal to noise ratio than a passive target but, unfortunately was free of bottom reverberation. In addition, the SRA was mounted to a small vessel that heaved significantly due to wave motion, which induced serious loss of coherence that affects the technique. Another recent paper (Sabra *et al.*, 2006) proposes to use the time reversal technique to enhance focusing on a target located on the bottom in presence of reverberation. A reflectivity map from the measured array response matrix is built, using passive (i.e., numerical) iterative time reversal. This method allowed successful detection of an ensemble of 12 spheres of diameter 50 cm on the bottom using a 96 element billboard array of central frequency 3.5 kHz, in 50 m water depth at range 200 m.

The present paper focuses on the application of the DORT method to detection in very shallow water in the presence of strong reverberation. A rigid vertical SRA with fully programmable parallel processed generators has been developed in the 10–15 kHz frequency band. After initial tests in a pool basin (Clorennec *et al.*, 2005; Folegot *et al.*, 2005), the system was deployed in the Bay of Brest for sea trials during spring 2005. Backscattered data from small submerged targets have been recorded for distances up to 600 m in a water depth varying from 10 to 18 m and the DORT method was applied for detection and localization of these targets. Furthermore, focusing was achieved either by time reversal or by transmission of the first eigenvector of the time reversal operator.

The system and the experimental setup are described in Sec. II. The signal measurement and analysis technique are developed in Sec. III. Several results from two different experiments are presented in Secs. IV and V. The first one is a detection experiment with two targets at the same range but different depth, showing the ability to separate the echo of

[a]Electronic mail: claire.prada-julia@espci.fr
[b]Currently at NURC, La Spezia, Italy.

FIG. 1. The pier of Sainte Anne du Portzic in the vicinity of Brest (France) where the system was deployed. The white lines show the insonified zone.

each target. The second one shows the ability to separate the echo of a target from bottom reverberation and compares the focusing obtained by active time reversal or by transmission of the first eigenvector of the time reversal operator.

II. EXPERIMENTAL SETUP

The experiments took place during spring 2005 in the Bay of Saint Anne du Portzic in France, a shallow water inlet (Fig. 1). This site is characterized by a significant spatial variability of the bottom interface, a rocky coastline generating strong reverberation and strong tidal currents (up to 4 knots, i.e. 2 m/s). The bathymetry has been recorded using multibeam sonar (Fig. 2). The sound speed c was measured with a profiling sound velocimeter at several distances in the whole water column and was constant at 1495 m/s. The SRA is deployed from a pier in a water depth varying from 7 to 12 m depending on tide (Fig. 3). It consists of a rigid support frame carrying 24 source/receiver transducers operating in the 10–15 kHz frequency band with a maximum source level of 203 dB re 1 μPa at 1 m each. The vertical transducer positions are adjustable, with a maximum array aperture of 12 m. In this experiment, they were equally spaced with a total aperture $D=9.4$ m. Each of the 24 channels is individually controlled and amplified during transmission and reception. In addition, a 16 element flexible vertical receiver array (VRA) can be deployed from a small vessel in order to sample the acoustic field produced by the SRA. As described in Folegot *et al.* (2005), communication and synchronization between the two arrays is established via a wireless local area network. In the experiments shown here, the targets are trihedral corner acoustic retroreflectors (TCAR) consisting of three mutually perpendicular intersecting plates made of air filled honeycomb composite (Fig. 4). These targets are small, easy to handle from a small ship and moor with an anchor. The biggest target has maximum dimension 60 cm$\approx 5\lambda$ and the smallest 40 cm.

III. DATA ACQUISITION AND ANALYSIS

The principle of the DORT method is described in several papers (Mordant *et al.*, 1999; Prada and Fink, 1994). It requires the measurement of the array response function $\mathbf{K}(t)$ made of the $N\times N$ interelement impulse responses of the array ($N=24$ in the experiment). Then, the invariants of the time reversal operator are calculated using the singular value decomposition (SVD) of the frequency response matrix $\mathbf{K}(\omega)$.

A. Measurement of the array response matrix

The array response matrix $\mathbf{K}(t)$ is measured using linear frequency modulated (LFM) sweeps transmitted from the SRA. In order to get a sufficient signal to noise ratio (SNR), the array response matrix is acquired using the Hadamard basis as proposed in Lingevitch *et al.* (2002). Such a basis is

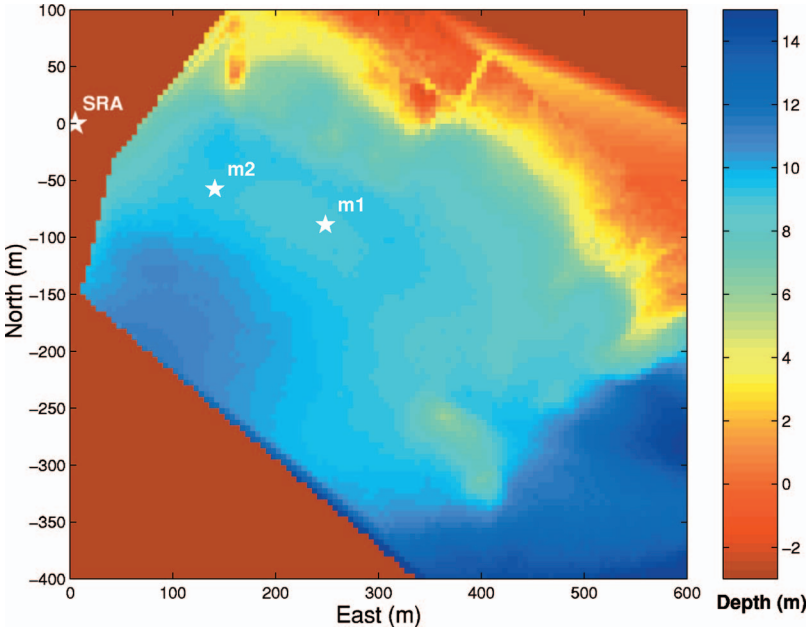


FIG. 2. Bathymetry of the bay of Sainte Anne du Portzic. The array was deployed at coordinates (0,0). The targets were successively deployed at points $m1$ and $m2$.

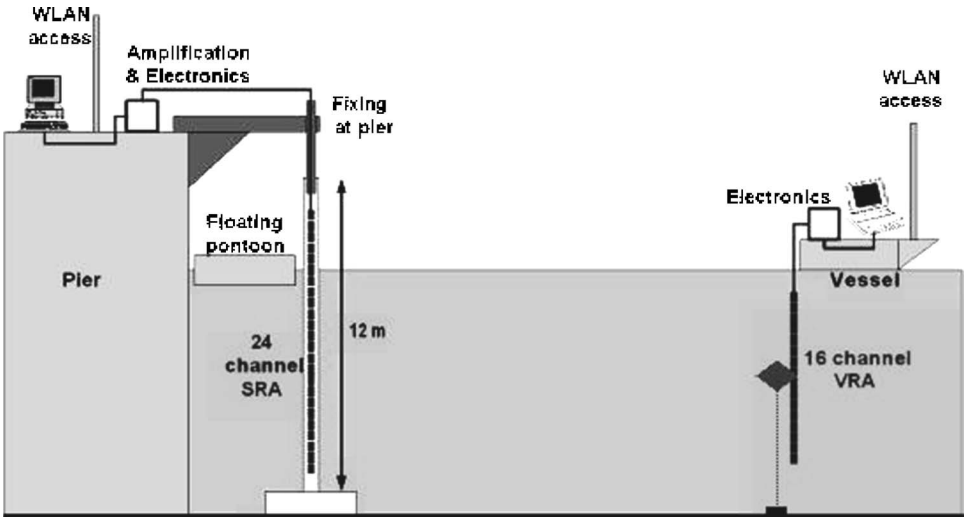


FIG. 3. Experimental setup. The SRA is deployed from a pier. The VRA is deployed from a small pleasure boat.

defined for $N=24$ and the SNR is improved by a factor $\sqrt{N} \approx 5$. Then, the N^2 received signals are correlated by the transmitted LFM, ensuring optimal SNR.

B. Analysis of the array response matrix

As the reverberation leads to significant backscattered signals, the array response matrix is built from short time windows. More precisely, the matrix $\mathbf{K}(r_0,\omega)$ is the Fourier transform of $\mathbf{K}(t)$ between time t_0 and $t_0+\Delta t$, where t_0 is related to the detection range r_0 through the equation $t_0=2r_0/c$ (c being the sound velocity) and Δt the window length. Due to the directivity of the transducer that is 40° in azimuth, and due to the unevenness of the bottom, the reverberation is very large, thus short time windows were used. A $\Delta t=3$ ms window was found to be a good compromise to include the complete response of one target and minimize the effects of reverberation.

The singular value decomposition of $\mathbf{K}(r_0,\omega)$ is done for regularly shifted time windows and the singular values are presented as a function of range r_0. An increase of the singular values at a given range corresponds to an increase in the backscattered energy, which can be associated with either a discontinuity of the bottom (a rock or a slope change), or the presence of targets. This can be elucidated by exploiting the corresponding singular vectors.

C. Backpropagation of singular vectors in free space

Singular vectors are numerically backpropagated to determine whether they correspond to a given target or the bottom reverberation. As the medium is complex and the wavelength ($\approx$12 cm) is rather short, the simplest way to backpropagate the data is to assume free space propagation. For the current point (r_0,h), the propagation vector in free space is defined as

$$\mathbf{H}(r_0,h,\omega)=\left(\frac{e^{-ikr_1}}{r_1},\frac{e^{-ikr_2}}{r_2},\cdots,\frac{e^{-ikr_N}}{r_N}\right)$$

where k is the wave number and $r_j=\sqrt{r_0^2+(h-h_j)^2}$ is the distance between the jth transducer of height h_j and the point (r_0,h). Then, for a given singular vector $\mathbf{V}(r_0,\omega)$ of the matrix $\mathbf{K}(r_0,\omega)$, the backpropagated field at point (r_0,h) is the scalar product of the singular vector by the propagation vector. The image $C(r_0,h)$ is then obtained by averaging the absolute values of the field over frequencies between 11 and 14 kHz:

$$C(r_0,h)=\sum_{\omega}|\langle\mathbf{V}(r_0,\omega),\mathbf{H}(r_0,h,\omega)\rangle|.$$

The resulting image is very interesting if the free space diffraction spot size ($\lambda r_0/D$) is smaller than half the water depth. For instance, with an array aperture $D=9$ m, a wavelength $\lambda=0.12$ and a 15 m water depth, the method can be used for distances r_0 up to 600 m. We shall see from several examples that a singular vector associated with a given target

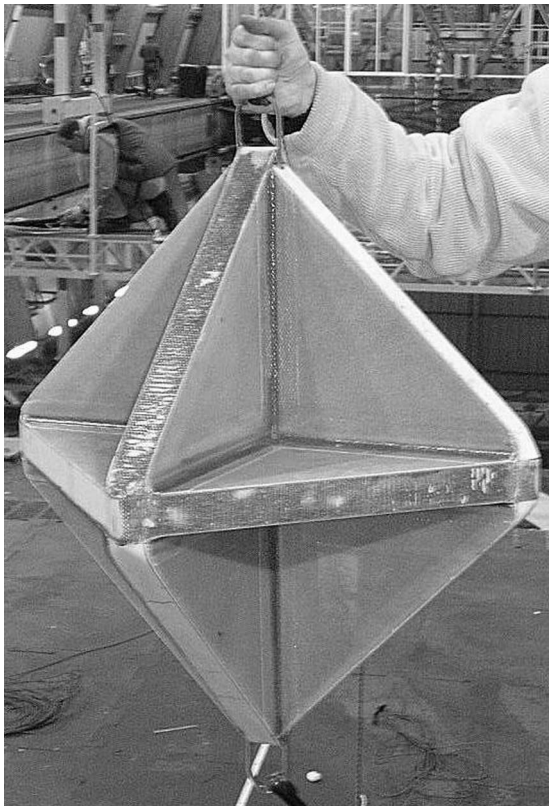


FIG. 4. The trihedral corner acoustic retroreflector.

will produce focusing on the target, plus focusing on each of the repeated images of the target with respect to the interfaces of the waveguide.

This type of backpropagation is simple as it only requires the knowledge of the array geometry and the sound velocity in water. The resulting image is unusual, but it provides a lot of information in a rapid manner. The water-air interface being flat compared to the bottom, the image with respect to this interface is generally very well defined, while images with respect to the bottom and higher order images are often more spread out. Besides, this free space backpropagation does not benefit from the guided propagation to increase resolution, that is why the backpropagation was achieved taking into account the water channel geometry as described in the next paragraph.

D. Backpropagation of singular vectors using a model of the waveguide

To account for guided propagation and achieve high resolution focusing, a second type of backpropagation is performed using the range-dependent acoustic model (RAM) (Collins and Westwood, 1991). RAM is based on the parabolic approximation and assumes an axisymmetric medium. As the bathymetry depends on azimuth, the bottom profile in the direction of the target is taken from the bathymetry map (Fig. 2) using an estimate of the target's position that was measured with a GPS. Besides, as the bathymetry map only starts 60 m away from the array, a linear bottom profile is assumed for the first 60 m. At last, as the transducers are mounted on a heavy rigid structure, several floats are regularly distributed along the array to reduce the weight. At low tide, part of the array is above the surface producing greater constraints. It results in a variation of the array tilt from 2° to 3° depending on tide. This angle is taken into account in the model.

In order to observe the quality of the focusing both in depth and range, the experimental singular vector is numerically backpropagated from the array. The field is displayed in the whole water column from the array to the target range. The operation differs from the free space backpropagation where the field at a given range r_0 corresponds to the singular vector of $\mathbf{K}_{r_0}$.

The aforementioned processing does not exploit the fully programmable generators. Active focusing was also achieved using either simple time reversal or transmission of particular singular vectors. In the following, two experiments are presented showing detection, localization and then active focusing.

IV. FIRST EXPERIMENT: DETECTION AND SEPARATION OF TWO TARGETS

In this experiment, two TCAR are used. Both TCAR are placed at approximately 250 m from the SRA at 5.5 and 8.5 m from the bottom (point $m1$ in Fig. 2). Thus, the targets are at about 2500 λ from the array and in a water depth about 100 λ. At the target distance, the free space diffraction spot is about 3.4 m, so that they are not very well resolved. In order to measure the array response matrix $\mathbf{K}(t)$, 200-ms-long LFM sweeps are transmitted from the SRA using the Hadamard basis.

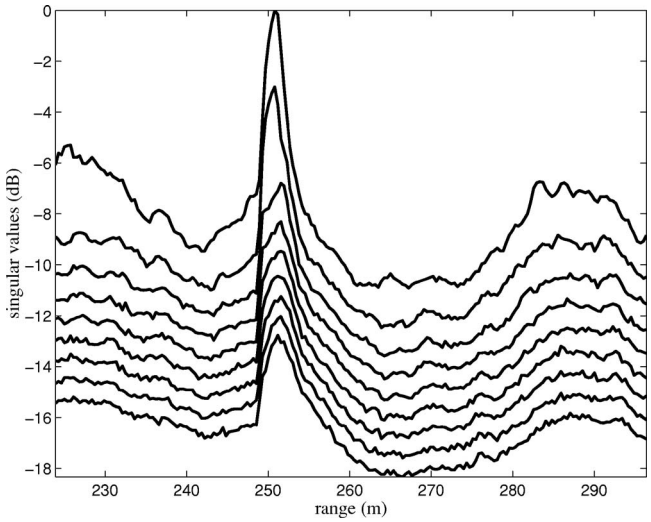


FIG. 5. Two targets at 253 m: Experimental singular values of $\mathbf{K}_{r_0}$ calculated for a sliding 3 ms time window.

A. Decomposition of the array response matrix

For the analysis, a $\Delta T=3$ ms time window is shifted by 0.5 ms steps. For each distance r_0, the matrices $\mathbf{K}_{r_0}$ are calculated at equally spaced frequencies from 11 to 14 kHz. After SVD, the singular values are averaged over frequencies and normalized. They are presented as a function of distance from 220 to 300 m (Fig. 5). It appears that two singular values emerge at $r_0=253$ m, the first is about 6 dB (the second about 4 dB) above reverberation singular values. For this distance, the two dominant singular values are clearly separated from the others in the whole frequency band from 10 to 15 kHz (Fig. 6). To determine the information contained in the dominant singular vectors, they are backpropagated in free space at the distance of the targets (Fig. 7). The first vector focuses at about 9 m from the bottom and the second at 6 m from the bottom, which are close to the targets

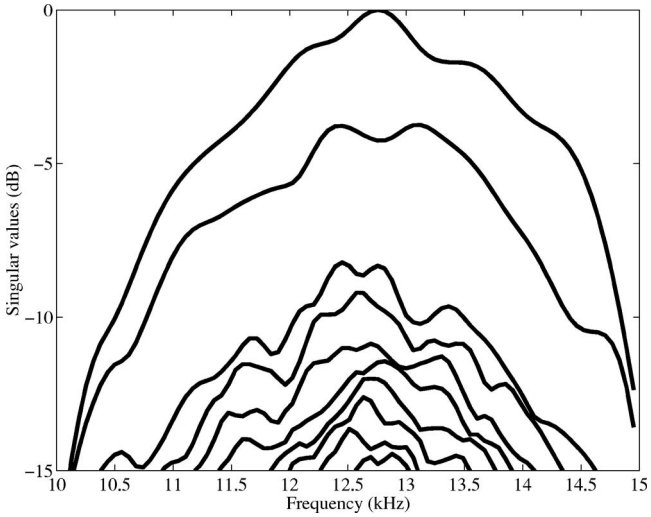


FIG. 6. Two targets at 253 m: Experimental singular values at $r_0=253$ m, as a function of frequency. Two singular values are clearly above the others.

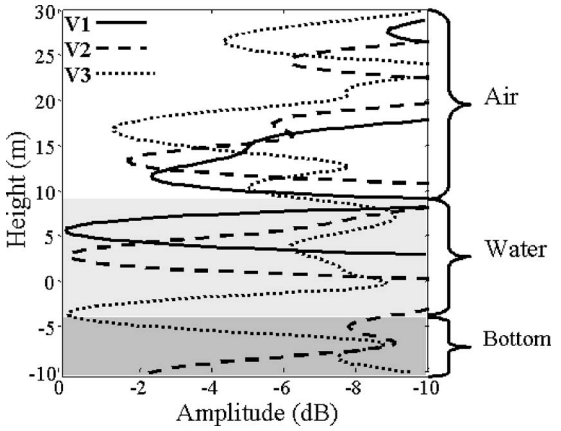


FIG. 7. Two targets at 253 m: Numerical backpropagation of the three first experimental singular vectors calculated for r_0=253 m. The two first vectors focus on the two targets. The third singular vector focuses at the bottom. Height 0 m corresponds to the sea floor at the position of the SRA.

positions. The third singular vector is also backpropagated and focuses at the bottom, which is about 4 m below the antenna foot at this range. This is in agreement with the water tank experiment shown in Folegot *et al.* (2003).

In the absence of a target, information on the bottom can be found on the backpropagation in free space of the first and second singular vectors (Fig. 8). These images provide the location of the target, along with an average bottom profile. For time windows before or after the target echo, the singular vectors correspond to reverberation and focus at the bottom and at its images with respect to the interfaces. In particular, the increase in the singular values around 230 m and between 280 and 290 m can be definitely attributed to bottom reverberation. The range resolution of these images is limited by the time window length Δt=3 ms and the LFM bandwidth, leading to about 3.5 m axial resolution. This resolution limitation is also observed on the singular values (Fig. 5).

B. Backpropagation using RAM

The backpropagation of the first and second singular vectors corresponding to the targets are calculated for several frequencies using RAM and then averaged. The SRA tilt is taken equal to 3.2° and the water depth at the array equal to 8.6 m. The focusing is clearly achieved on each target with a very good resolution in depth (Fig. 9). In order to appraise the quality of the focusing, the comparison between the backpropagation in free space and with RAM code at the distance of the targets is displayed in Fig. 10. For each singular vector, the two types of propagation focus at the same depth, the improvement in resolution at −6 dB is about 3.5, which means that at least one bottom and one surface reflection contribute to the reconstruction with RAM. A better knowledge of the water channel and the use of a three-dimensional code would certainly result in even better resolution.

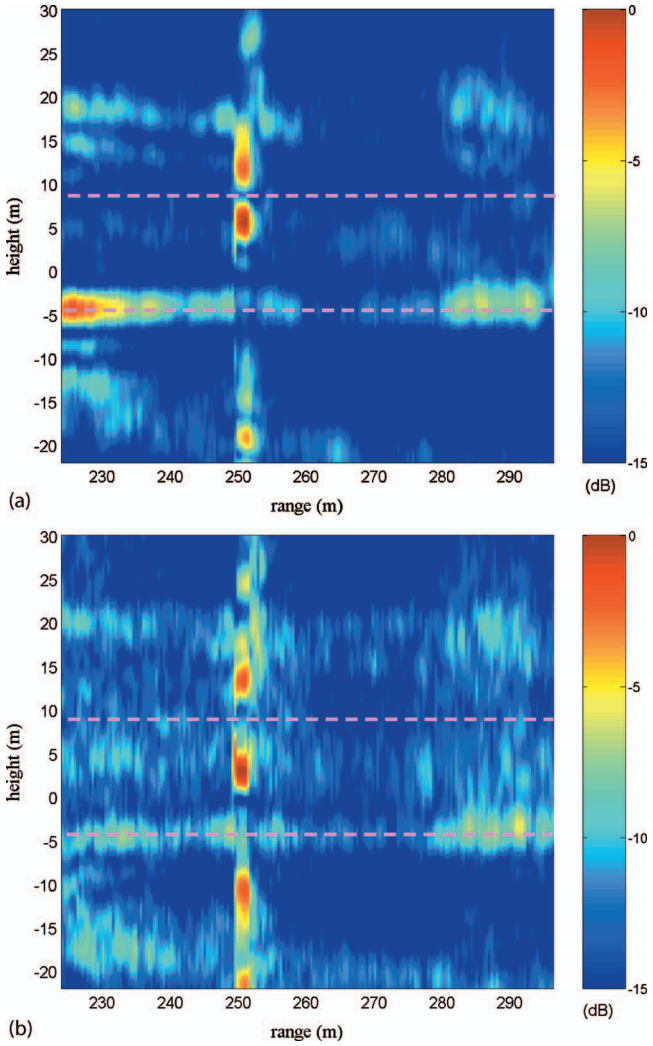


FIG. 8. Two targets at 253 m: Numerical backpropagation in free space of the first (a) and second (b) experimental singular vectors obtained for shifted time windows. The pink dashed lines indicate the bottom and the surface. The color scale is in decibels. Height 0 corresponds to the antenna foot.

V. SECOND EXPERIMENT: ACTIVE FOCUSING ON A SINGLE TARGET

In the second experiment, the TCAR is placed approximately at 140 m from the array and at 7 m from the bottom (point *m*2 in Fig. 2). The purpose is to compare simple active time reversal of the target echo with the active transmission of a singular vector. The vertical receiver array is deployed at the position of the target to sample the acoustic field. This relatively short distance was chosen to avoid the strong tide current that renders measurement on the VRA difficult for longer distances. The water depth at the array position is 10 m so that all the elements are immersed. As in the first experiment, the array response matrix is acquired using the Hadamard basis and 150 ms frequency sweeps.

A. Time reversal and DORT analysis

The array response matrices are calculated for sliding 3 ms time windows. The singular value decomposition of each matrix is then achieved. For comparison with the singular values, the energy of the echo after broadside transmission is also calculated for the same set of time windows.

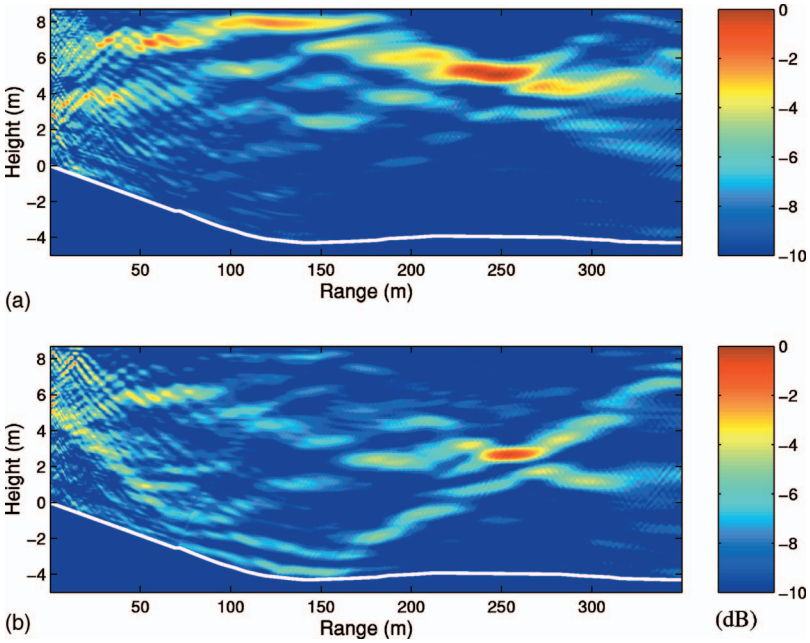


FIG. 9. Two targets at 253 m: Average field obtained by RAM backpropagation of the first (a) and second (b) experimental singular vectors calculated on the same time window as in Fig. 7, i.e., r_0=253 m.

Broadside transmission corresponds to the transmission of the first Hadamard vector (all elements transmit the same LFM simultaneously). The echo received on the N transducers are noted $r_1(t),\ldots,r_N(t)$ and the energy is calculated as

$$E(r_0)=\sum_{j=1}^{N}\sum_{t=t_0}^{t_0+\Delta t} r_j(t)^2.$$

The singular values and the normalized energy are represented as a function of distance in decibel (Fig. 11). A clear enhancement at the target distance can be observed on those curves. The fact that several singular values increase might be explained by the presence of the boat, the anchor, the VRA, and probably by the flucuations of the medium during acquisition. The backscattered energy after broadside transmission is higher after the target than before, this is explained by the slope change in the bottom profile that occurs near the target (see Figs. 2 and 13). Then the numerical backpropagation in free space of the echo after broadside transmission is calculated [Fig. 12(a)]. A local maximum occurs at the height and range of the target but the greatest lobe occurs on the bottom (−4 m) probably corresponding to the anchor maintaining the target. The energy spreading below, above, and just behind the target might be due to the presence of the boat and the VRA. Beyond the target, the focusing occurs on the bottom and the image of the bottom with respect to the surface, meaning that the broadside transmission produces strong reverberation after 140 m.

On the contrary, the backpropagation of the first singular vector focuses with good signal to noise ratio at the target and at its images with respect to the interfaces [Fig. 12(b)]. This confirms that it solely contains signal from the target. Around 130 m the singular vector mostly focuses on the image of the bottom with respect to the surface, meaning that the ray path with one reflection at the surface is the dominant one.

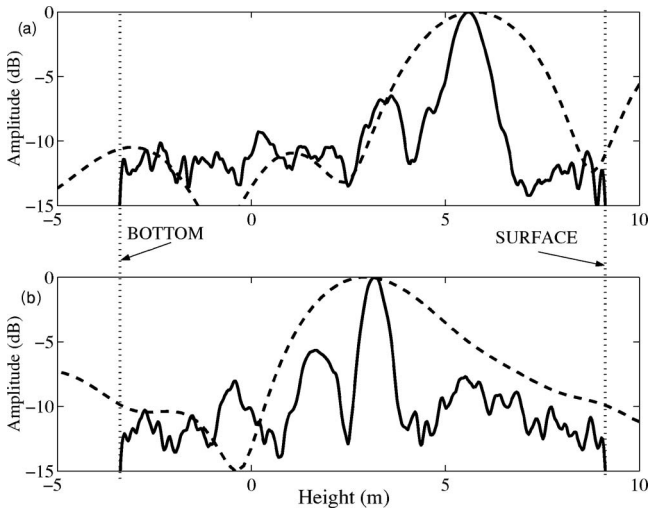


FIG. 10. Two targets at 253 m: Backpropagation of the first (a) and second (b) singular vectors calculated for r_0=253 m in free space (dashed line) and using RAM (solid line).

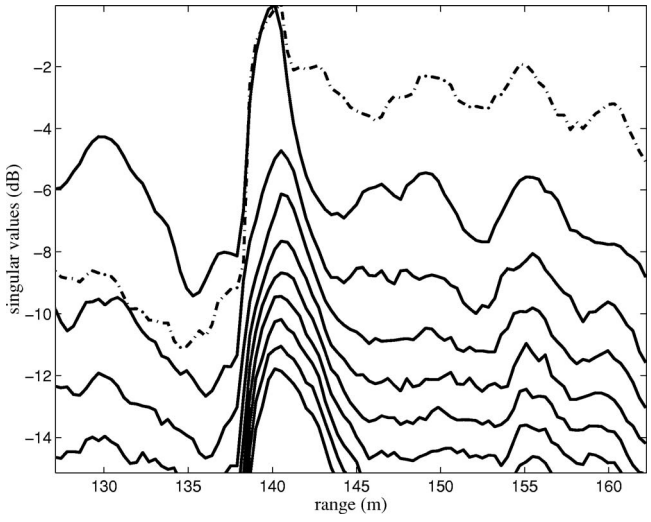


FIG. 11. One target at 143 m: Singular values as a function of distance (solid line) and normalized energy of the echo after broadside transmission (dash-dot) calculated for 3 ms shifted time windows.

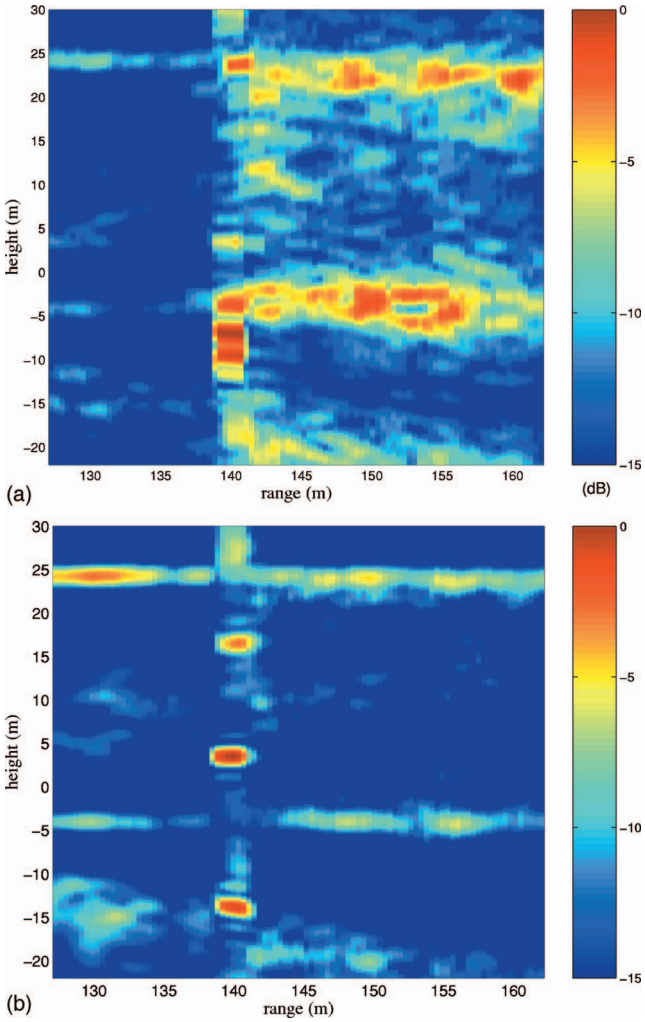


FIG. 12. One target at 143 m: Numerical backpropagation in free space of the echo after broadside transmission (a) and of the first singular vector (b) calculated for 3 ms shifted time windows.

B. Backpropagation using RAM

To calculate the backpropagation with RAM, the array tilt is taken equal to 2.9° and the water depth at the array to 10.35 m. The first and second singular vectors are backpropagated over 250 m for several frequencies and the amplitude field is averaged (Fig. 13). The focusing clearly occurs at the target depth for the first singular vector, and the focal spot is about three times thinner than in free space as in the first experiment. The second singular vector focuses on the bottom, while some energy is also focused on the target (however 5 dB below). This might be explained by the fact that the target is not point-like (5 λ) and probably moving around an average position.

C. Active focusing with time reversal and DORT

The programmable parallel processed generators are used to time reverse the echo of the target. The VRA is deployed from the boat at the target distance and only eight elements spanning 3.5 m around the target are used to control the field. In the first stage, a time window is selected on the echo measured after broadside transmission, and the selected echo is time reversed. This transmission is repeated five times and each time, the transmitted signal is measured at the SRA (Fig. 14, left-hand side). The focusing occurs at height 3.8 m with significant secondary lobes. In the second stage, the first singular vector is calculated on the same time window at the central frequency. Then, this vector is transmitted to the SRA and the signal measured at the VRA (Fig. 14, right-hand side). As for time reversal, the transmission is repeated 5 times. Again, the maximum occurs at height 3.8 m, but with low secondary lobes.

In both cases, the focusing is achieved at the target. At this distance, the free space point spread function is about 1.7 m wide. In both cases, the main lobe is less than 1 m large, probably of the order of 50 cm, which means that more than two reflections (at bottom and surface) contribute to the focusing. For transmission of the singular vector, the secondary lobes are below 10 dB, which confirms the fact that the decomposition separates the echo of the target from the other contributions that can be attributed to the anchor, the VRA, the boat, or bottom reverberation. The strong tidal

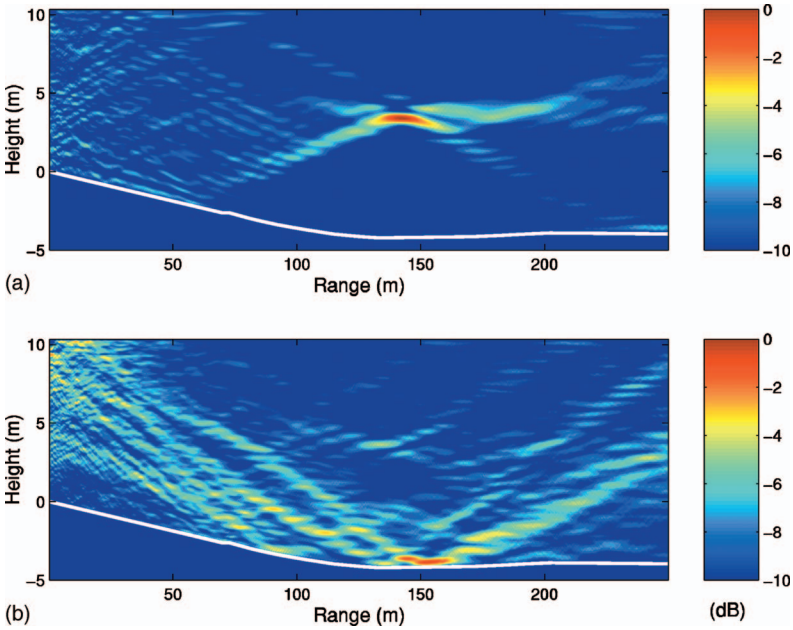


FIG. 13. One target at 143 m: Numerical backpropagation using RAM code of the first (a) and second (b) experimental singular vectors. Field are normalized and represented in decibel scale.

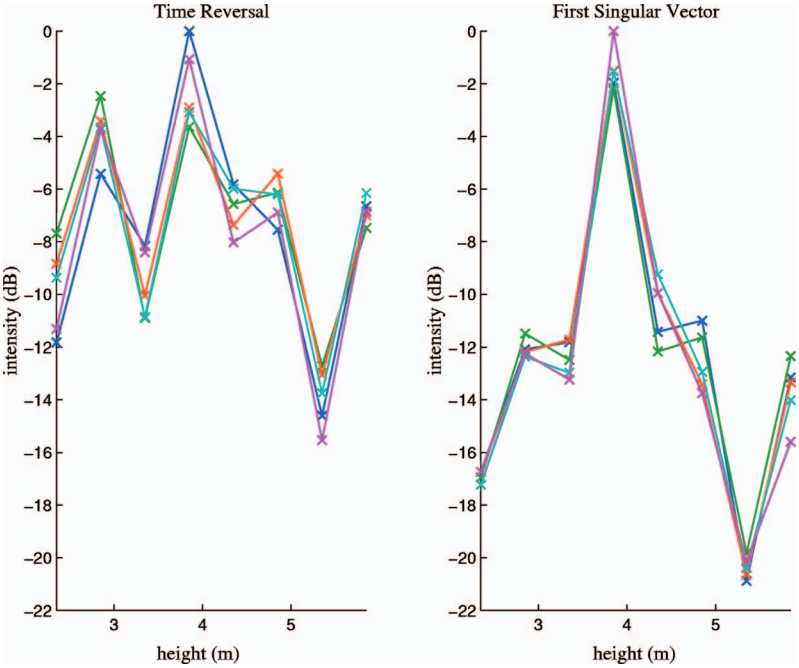


FIG. 14. One target at 143 m: Active focusing at the target using time reversal (left) and the first singular vector (right).

currents in the Bay of Brest made it difficult to deploy the vertical receiver array at longer distances. This is why the focusing experiment was done at a short range compared to the expected range of the system. Further experiments are required to determine the real limits of this system.

VI. DISCUSSION

Detection experiments using the DORT method applied on data measured with a 24 elements vertical source-receiver array were presented. The experiments have been carried out in a reverberation limited shallow water environment using corner retroreflectors.

The results confirm those obtained in laboratory experiments. Two targets have been individually detected and correctly located within the water depth. The method appears to be robust and provides much better information on the target localization than conventional beam-forming on flat broadside transmissions.

In an experiment with a single target, active selective focusing by transmission of a singular vector has been shown to produce a strongly localized focal spot at the position of the scatterer. The obtained resolution was more than three times better than in free space. To the authors' knowledge, this is the first time that the DORT method was applied successfully to backscattered data in a real shallow water ocean environment with targets at range 2500 λ, in depth 100 λ. Beyond detection applications, this selective focusing ability also opens perspectives in underwater communication.

ACKNOWLEDGMENTS

The authors are grateful to Dr Yann Stephan, from the French Hydrographic and Oceanographic Office, for providing the bathymetric model in the area. This work was funded by the French Armament Procurement Agency (DGA/SPN) under Contract No. 02 77 154 470 75 53.

Clorennec, D., de Rosny, J., Minonzio, J.-G., Prada, C., Fink, M., Folegot, T., Billand, P., Tauvry, S., Hibral, S., and Berniere, L. (**2005**). "First tests of the DORT method at 12 kHz in a shallow water waveguide," in Proceedings IEEE Oceans'05 Europe, Brest, France, Vol. **2**, 1205–1209.

Collins, M. D., and Westwood, E. K. (**1991**). "A higher-order energy-conserving parabolic equation for range-dependent ocean depth, sound speed, and density," J. Acoust. Soc. Am. **89**, 1068–1075.

Edelmann, G., Akal, T., Hodgkiss, W., Kim, S., Kuperman, W., and Song, H. (**2002**). "An initial demonstration of underwater acoustic communication using time reversal," IEEE J. Ocean. Eng. **27**, 602–609.

Fink, M. (**1997**). "Time reversed acoustics," Phys. Today **50**, 34–40.

Folegot, T., Billand, P., Tauvry, S., Hibral, S., Berniere, L., de Rosny, J., Clorennec, D., Minonzio, J.-G., Prada, C., and Fink, M. (**2005**). "Design of a time reversal mirror for medium scale experiments," in Proceedings IEEE Oceans'05 Europe, Brest, France, Vol. **2**, pp. 1210–1213.

Folegot, T., Prada, C., and Fink, M. (**2003**). "Resolution enhancement and separation of reverberation from target echo with the time reversal operator decomposition," J. Acoust. Soc. Am. **113**, 3155–3160.

Gaumond, C. F., Fromm, D. M., Lingevitch, J. F., Menis, R., Edelmann, G. F., Calvo, D. C., and Kim, E. (**2006**). "Demonstration at sea of the decomposition-of-the-time-reversal-operator technique," J. Acoust. Soc. Am. **119**, 976–990.

Kuperman, W. A., Hodgkiss, W. S., Song, H. C., Akal, T., Ferla, C., and Jackson, D. R. (**1998**). "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **103**, 25–40.

Lingevitch, J. F., Song, H. C., and Kuperman, W. A. (**2002**). "Time reversed reverberation focusing in a waveguide," J. Acoust. Soc. Am. **111**, 2609–2614.

Mordant, N., Prada, C., and Fink, M. (**1999**). "Highly resolved detection and selective focusing in a waveguide using the DORT method," J. Acoust. Soc. Am. **105**, 2634–2642.

Prada, C., and Fink, M. (**1994**). "Eigenmodes of the time reversal operator: A solution to selective focussing in multiple target media," Wave Motion **20**, 151–163.

Sabra, K. G., Roux, P., Song, H.-C., Hodgkiss, W. S., Kuperman, W. A., Akal, T., and Stevenson, J. M. (**2006**). "Experimental demonstration of iterative time-reversed reverberation focusing in a rough waveguide application to target detection," J. Acoust. Soc. Am. **120**, 1305–1314.

Yokoyama, T., Kikuchi, T., Tsuchiya, T., and Hasegawa, A. (**2001**). "Detection and selective focusing on scatterers using the decomposition of time reversal operator method in pekeris waveguide model," Jpn. J. Appl. Phys., Part 1 **40**, 3822–3828.

Acoustic detection of North Atlantic right whale contact calls using spectrogram-based statistics

Ildar R. Urazghildiiev[a] and Christopher W. Clark
Bioacoustics Research Program, Cornell Laboratory of Ornithology, Ithaca, New York 14850-1999

(Received 20 November 2006; revised 7 May 2007; accepted 14 May 2007)

This paper considers the problem of detection of contact calls produced by the critically endangered North Atlantic right whale, *Eubalaena glacialis*. To reduce computational time, the class of acceptable detectors is constrained by the detectors implemented as a bank of two-dimensional linear FIR filters and using the data spectrogram as the input. The closed form representations for the detectors are derived and the detection performance is compared with that of the generalized likelihood ratio test (GLRT) detector. The test results demonstrate that in the presence of impulsive noise, the spectrogram-based detector using the French hat wavelet as the filter kernel outperforms the GLRT detector and decreases computational time by a factor of 6. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2747201]

PACS number(s): 43.30.Wi, 43.60.Bf [WWA] Pages: 769–776

I. INTRODUCTION

The North Atlantic right whale (NARW), *Eubalaena glacialis*, is critically endangered.[1] The population size is estimated to be about 350 individuals due to low population growth and increasing numbers of vessel collisions and entanglements with fishing gear.[2] Monitoring for the occurrence of NARW has been accomplished by detecting NARW contact calls on data recordings obtained from distributed autonomous hydrophone systems.[3–9] More recent monitoring efforts sample for calling right whales over large areas using large numbers of broadly distributed, autonomous systems (>20) for many months at a time. Such large-scale efforts yield enormous data sets totaling many years, and the analyses of these present an analytical challenge. The method of data analysis involving human operators to visually and aurally evaluate data spectrograms is impractical in projects which collect huge amounts of data, so the design of effective automated detection techniques is of critical importance.

The fundamental theoretical difficulty of the automated detection technique design is that both the whale signals and ambient noise are random processes with unknown statistical and spectral properties. As a result, the problem of NARW contact call detection has no optimal solution. In Urazghildiiev and Clark (2006),[10] the generalized likelihood ratio test (GLRT) detector was obtained under the assumption that ambient noise belongs to the class of Gaussian processes. Such an assumption is violated for ambient noise contaminated by impulsive noise. In fact, the class of distributions to which ambient noise belongs is larger than the class of Gaussian processes. Therefore, the GLRT detector, which is optimal within the class of Gaussian processes, does not guarantee optimality within the full class of possible ambient noise distributions. It means that there exist detectors different from the GLRT that provide higher detection performance in the presence of impulsive noise.

Another weakness of the GLRT detector proposed in Ref. 10 is high computational costs. The reported run-time needed to compute the GLRT-based statistic using a filter bank of 12 filters and a 3 GHz processor is approximately 1 h per 1 day of data recording. Correspondingly, with the GLRT-based statistic, 10 years of data recording would require about 150 days of continuous computing. Other factors that can increase the analysis time include a higher sampling rate and more filters. Involving more processors may mitigate the analysis time problem, but in general the high computational cost essentially limits the effectiveness of the GLRT detector in cases where very large data sets must be analyzed.

Thus, the fact that the GLRT detector is optimal under certain conditions does not necessarily mean that it is the right detector to use or even that is a satisfactory detector. Although the optimality of the GLRT provides a good starting point, the design of NARW detectors providing the desirable run-time and robustness against the full class of ambient noise distributions remains an important practical problem.

In this paper, the problem of an automatic detector design is solved with application to the class of passive acoustic NARW monitoring systems performing analysis of long-term data recordings. In such systems, the main requirement is reduction of run-time without negatively affecting the detection performance. To find a solution, we use some specific restrictions applied to such a class of systems. We assume that the automatic detector provides information on the time of occurrence of signals, but the human operator makes the final decision by visual inspection of the corresponding areas on the data spectrogram. As was shown in Ref. 11, in the case of signals with low signal-to-noise ratio (SNR) the GLRT-based detector provides higher detection performance than the human operator. However, the operator will likely reject weak signals detected by the GLRT detector if the signals are nonvisible on the spectrogram. Therefore, in the applications involving human operators in the decision-making process, only signals with sufficiently high SNR so

[a]Electronic mail: iru2@cornell.edu

as to be visible in the spectrogram should be detected by the automatic detector. In the strict sense, applying the short-time Fourier transform (STFT) to compute the spectrogram has no statistical motivation. As a result, the heuristic spectrogram cross correlation[12–14] and the "edge"[15] detection algorithms provided lower detection probability under a given probability of false alarm. However, calculating the detection statistic from the data spectrogram does result in a decrease in run-time. Therefore, we find the optimal detector structure within the class of detectors using the data spectrogram to compute the detection statistic. As an optimality criterion, the Neyman-Pearson criterion is used. The resultant detector is compared with the GLRT in terms of detection performance and run-time using data sets collected off Massachusetts in Cape Cod Bay and the Great South Channel, and in the Southeast Atlantic off Savannah, GA, during periods when right whales were present.

II. DATA MODEL AND PROBLEM FORMULATION

Let $x(t)$, $t=1,2,\ldots$ denote the discrete-time-varying data recorded from the hydrophone system. For any data segment $\mathbf{x}(t)=(x(t),x(t+1),\ldots,x(t+N-1))^T\in E^N$ consisting of N samples, the following hypotheses can be introduced:

$$H_0:\mathbf{x}(t)=\mathbf{w}(t),\quad H:\mathbf{x}(t)=A\mathbf{s}(\boldsymbol{\lambda})+\mathbf{w}(t) \tag{1}$$

where $\mathbf{w}(t)=[w(t),w(t+1),\ldots,w(t+N-1)]^T\in E^N$ is the noise vector; $\mathbf{s}(\boldsymbol{\lambda})=[s(1,\boldsymbol{\lambda}),s(2,\boldsymbol{\lambda}),\ldots,s(N,\boldsymbol{\lambda})]^T\in E^N$ is the signal vector; A is a positive scalar representing the signal amplitude; $\boldsymbol{\lambda}$ is the vector of signal parameters; E^N is Euclidean N-dimensional space, and the symbol "T" denotes the transpose. The null hypothesis H_0 represents the case of signal absence, and the alternative hypothesis H corresponds to the case of signal presence. We assume that N is taken to ensure that the duration of the segment is close to 1 s, the typical duration of contact call signals. The *a priori* probabilities of the hypotheses, $p(H_0)$ and $p(H)$, are unknown, but we assume that[10]

$$p(H_0)\gg p(H). \tag{2}$$

Based of the analysis of the vector $\mathbf{x}(t)$, one of the hypotheses, H_0 or H, will be accepted. Subject to Eq. (2), we introduce the Neyman-Pearson optimality criterion implying maximization of the detection probability under the given probability of false alarm.[16–18]

Let us introduce the following notations: U_X is the set of distributions of the data vector $\mathbf{x}(t)$; U_D is the set of acceptable detectors; $d\in U_D$ is the detector from U_D; and $\alpha(\beta|d)$ is the probability of detection corresponding to the probability of false alarm, β, and provided by the detector d.

The problem considered in this paper is to find the detector $\hat{d}\in U_D$ that for a given U_X satisfies the condition

$$a(\beta|\hat{d})=\max_d \alpha(\beta|d),\quad \beta\in[0,1]. \tag{3}$$

To derive a detector structure in a closed form that allows practical implementation, the sets U_X and U_D should be specified.

The set U_D is determined based on practical requirements and restrictions applied to the physical system. In this paper, we consider a class of passive acoustic NARW detections systems that involve a human operator to run the analysis and to make a final detection decision. Examples of such systems using autonomous seafloor recorders and autonomous real-time buoys for collecting the data are considered in Refs. 3 and 4. After running the automatic detector, the human operator confirms the presence or absence of NARW contact calls by visual inspection of the detections annotated in the data spectrogram. By this process, only signals visible on the spectrogram are of interest and should be detected by the automatic detector.[11] It is assumed that the STFT using a sliding short-time window of K samples ($K\ll N$) and overlapped by K_{ov} samples is applied to the data samples $x(t)$. Let the matrix $\mathbf{G}(i)=\{G(k,n),k=1\ldots K,n=i\ldots i+N_S-1\}\in E^{K\times N_S}$, $i=0,1,\ldots$ denote a spectrogram calculated for the data vector $\mathbf{x}(iK_0)$ where $K_0=K-K_{\text{ov}}$. Then the hypotheses (1) can be represented as

$$H_0:\mathbf{G}(i)=\mathbf{G}_W(i),\quad H:\mathbf{G}(i)=\mathbf{G}_S(\boldsymbol{\lambda})+\mathbf{G}_W(i), \tag{4}$$

where the matrices $\mathbf{G}_S(\boldsymbol{\lambda})$ and $\mathbf{G}_W(i)$ represent the spectrograms of the signal and noise, respectively. The automatic detector calculates a scalar random variable $z(i)$ as a particular transformation of the data vector $\mathbf{x}(iK_0)$. The variable $z(i)$ is referred to as a statistic.[16] The statistic is compared with a threshold C, and the signal is considered detected if $z(i)\geqslant C$. We assume that the human operator verifies any event $z(i)\geqslant C$ as a NARW contact call by visually inspecting the spectrogram $\mathbf{G}(i)$. To mitigate the run-time problem, we require the statistic $z(i)$ to be calculated from the data spectrogram, $z(i)=z(\mathbf{G}(i))$. We also make the following assumptions: (a) the complete data recording $x(t)$ is available for computing $z(i)$ and for making a decision; and (b) the human operator determines the value of the threshold, C, in the data analysis process based on current noise conditions. For example, the operator can set the threshold so as to obtain a certain number of detections, $z(i)\geqslant C$, per 24 h of observation, or set the threshold based on the local background noise level. The problem of choosing the threshold is beyond the scope for this paper. Instead, we focus our attention on the problem of optimizing the detector structure. Taking into account the above-noted assumptions, we restrict U_D to the class of detectors that calculate the statistic $z(i)$ from the data spectrogram. A more detailed description of U_D is given in Sec. III.

Observations show that NARW contact calls are transient, locally narrowband, frequency-modulated signals. The energy in about 99% of NARW contact calls is distributed within the 40–250 Hz frequency band. Therefore, we model NARW contact calls as polynomial-phase signals so that[10]

$$s(t,\boldsymbol{\lambda})=\cos[\theta_{\lambda}(t)+\varphi_0], \tag{5}$$

$$\theta_{\lambda}(t)=2\pi\sum_{m=1}^{M}(m)^{-1}f_{m-1}t^m, \tag{6}$$

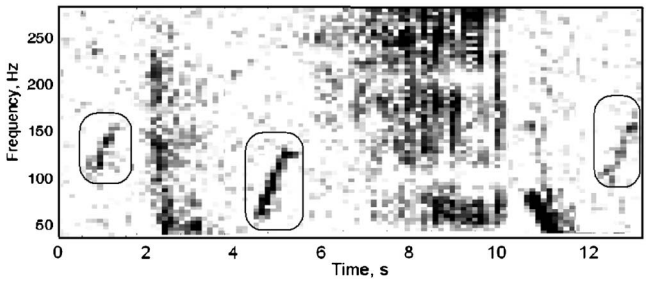


FIG. 1. A spectrogram of three NARW contact calls observed in the presence of ambient noise, including a burst of impulsive noise between the first and second calls.

$$\boldsymbol{\lambda} = (f_0, \ldots, f_{M-1})^T \in U_\lambda, \tag{7}$$

where U_λ is the set of polynomial coefficients, and φ_0 is the initial phase. The admissible set of $\boldsymbol{\lambda}$ can be obtained from empirical distributions of the model parameters.[10] The polynomial coefficients in Eq. (7) unambiguously specify the instantaneous frequency (IF) of the signal,

$$f(t,\boldsymbol{\lambda}) = \frac{1}{2\pi}\frac{d\theta_\lambda(t)}{dt} = \sum_{m=0}^{M-1} f_m t^m. \tag{8}$$

Note that the IF approximates the positions of the peak absolute values of the spectrogram $\mathbf{G}_S(\boldsymbol{\lambda})$. The frequency corresponding to the maximum of the nth column of the matrix $\mathbf{G}_S(\boldsymbol{\lambda})$ can be represented as

$$\tilde{f}(n,\boldsymbol{\lambda}) \approx f(nK_0,\boldsymbol{\lambda}). \tag{9}$$

This property is important in practice since the visual analysis of the spectrogram is one of the primary methods of signal detection in bioacoustics. For a quadratic-phase signal, the first three polynomial coefficients can be interpreted as follows: f_0 is the start frequency, and f_1 and f_2 represent the slope and the curvature, respectively. The initial phase is assumed to be a random value uniformly distributed over the interval $0-2\pi$. The model parameters $\{A,\boldsymbol{\lambda}\}$ are assumed to be unknown and nonrandom.

A spectrogram illustrating three NARW contact calls is shown in Fig. 1. The spectrogram was computed using the Fast Fourier Transform (FFT) with rectangular window, $K=256$ samples and $K_{\text{ov}}=128$ overlapping samples. The sampling frequency, F_S, was 2 kHz so that the duration of each short-time segment was $K/F_S=0.128$ s. Note that a dark area on the NARW contact call spectrogram represents the IF that can be approximated by Eq. (9).

Observations show that within the frequency band occupied by signals the statistical and spectral properties of ambient noise change dramatically. Ambient noise contains a continuous random process whose power spectrum density (PSD) does not change essentially over tens of seconds and more. We refer to this process as background noise. There is also a nonzero probability of occurrence of impulsive noise arising due to human activity (manmade noise), the presence of acoustically active animals (biological noise), the interaction of the sensor with the medium (mechanical noise), and other factors. The design of a general noise model feasible for development of practical detection schemes is a difficult problem. As was shown in Ref. 10, the background component of noise in Eq. (1) can be modeled as a locally stationary Gaussian random process. Impulsive noise is assumed to be an arbitrary process with finite energy and duration. The example of ambient noise spectrogram is shown in Fig. 1. Thus, we assume that ambient noise can be represented as a Gaussian process contaminated by unknown impulsive processes. We define the set U_X to which the distributions of $\mathbf{x}$ belong as[19,20]

$$U_{X|H} = \{W(\mathbf{x} - As(\boldsymbol{\lambda}))\}, \tag{10}$$

$$U_{X|H_0} = \left\{ \varepsilon_0 W(\mathbf{x}) + \sum_{i=1}^{N_\varepsilon} \varepsilon_i H_i(\mathbf{x}) \right\},$$
$$\sum_{i=1}^{N_\varepsilon} \varepsilon_i = 1 - \varepsilon_0, \quad \int_x H_i(\mathbf{x}) d\mathbf{x} = 1, \tag{11}$$

where $W(\mathbf{x})$ is the distribution of the zero-mean Gaussian process; ε_0 is a scalar representing the percentage of the Gaussian process, $H_i(\mathbf{x})$ is the distribution of the ith contaminating process, ε_i is a scalar representing the percentage of contamination by the ith process ($\varepsilon_0 \gg \varepsilon_i$), and N_ε is the number of contaminating processes. As was shown in Ref. 10, the Gaussian model is acceptable for about 90% of the observed ambient noise conditions and during any time interval 8–16 s in length, variations in the covariance matrix of Gaussian noise are negligibly small. Therefore, we suppose that $\varepsilon_0 \approx 0.9$. The values $H_i(\mathbf{x})$, ε_i, and N_ε are assumed to be unknown.

Equations (10) and (11) represent the statistical data model for the optimization problem, Eq. (3), considered in this paper. The detailed structure of the set of acceptable detectors, U_D, as well as a technique used to find the optimal detector $\hat{d} \in U_D$ are considered in the next section.

III. SPECTROGRAM-BASED STATISTICS

A realizable spectrogram-based detection scheme can be implemented using a bank of two-dimensional (2D) linear FIR filters. The output of a 2D filter can be represented as a bilateral convolution of the spectrogram:

$$u(i) = \sum_{k=1}^{K}\sum_{n=1}^{N_S} G(k,i-n)Q(k,n) = \mathbf{G}(i) * \mathbf{Q}, \tag{12}$$

where $Q(k,n)$ is the kernel, $\mathbf{Q}=\{Q(k,n)\} \in E^{K\times N_S}$ is the matrix representing the kernel, and the asterisk (*) denotes the bilateral convolution.

Since background noise is modeled as a locally stationary Gaussian process, we implement a normalization of the spectrogram in the first stage. The normalized spectrogram can be calculated as

$$\tilde{G}(k,n) = G(k,n)/\hat{B}(k), \tag{13}$$

where $\hat{B}(k)$ is the noise PSD estimate calculated over the time interval for which background noise is stationary. Note that normalization, Eq. (13), is also referred to as prewhitening.[17] A robust technique of calculating $\hat{B}(k)$ based

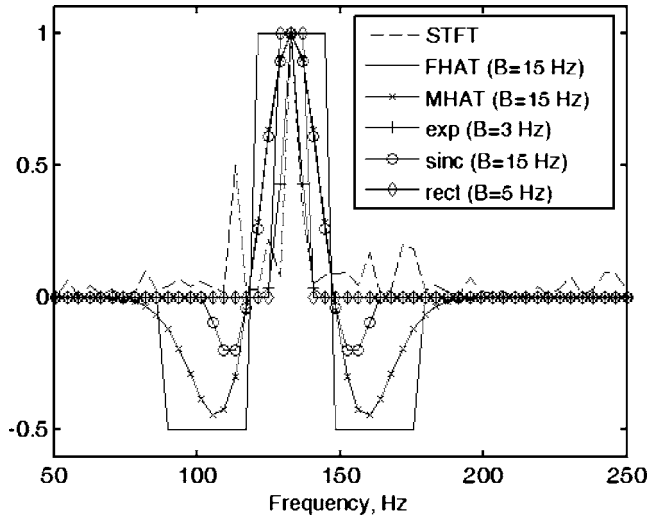


FIG. 2. A normalized STFT of the NARW contact call as well as FHAT, MHAT, exponential, rectangular, and truncated sinc windows in frequency domain with nearly optimal bandwidths.

on median filtering of $G(k,n)$ was considered in Urazghildiiev and Clark (2006).[10] Substituting the normalized spectrogram in Eq. (12), we obtain

$$u(i) = \tilde{\mathbf{G}}(i) * \mathbf{Q}, \tag{14}$$

where $\tilde{\mathbf{G}} = \{\tilde{G}(k,n)\} \in E^{KN_S}$ is the matrix of a normalized data spectrogram.

As follows from Eq. (14), the problem of optimization of the detector structure can be reduced to finding the optimal kernel, $\hat{\mathbf{Q}}$. Since the set of acceptable kernel is infinite, certain restrictions regarding kernel types must be introduced to implement a practical solution.

Subject to Eq. (10), the acceptable kernels should be chosen as a certain approximation of the signal spectrogram:

$$\mathbf{Q} = c\mathbf{G}_S(\boldsymbol{\lambda}), \tag{15}$$

where c is a positive constant. A column of the matrix $\tilde{\mathbf{G}}(i)|_H \approx \mathbf{G}_S(\boldsymbol{\lambda})$ representing a short-time FFT spectrum of NARW contact calls is shown in Fig. 2. As Figs. 1 and 2 illustrate, most of the signal energy in any short-time segment of the signal is concentrated in a limited bandwidth, B, centered on the IF. The PSD of whitened background noise is almost uniform, but the bandwidth of impulsive noise can vary considerably. Therefore, we include bandwidth as a kernel parameter and find the optimal kernel as a function $\mathbf{Q} = \mathbf{Q}(\boldsymbol{\lambda}, B)$, $\boldsymbol{\lambda} \in U_\lambda$, $B > 0$.

As was shown in Ref. 10, the natural variability of a signal's IF is high so no single filter can provide acceptable detection performance. Therefore, we apply a bank of 2D linear FIR filters and calculate the statistic from the filter bank output given by

$$\tilde{u}(i,\mathbf{Q}) = \max_p u(i,\mathbf{Q}(\boldsymbol{\lambda}_p,B)), \quad p = 1 \ldots P, \quad \boldsymbol{\lambda}_p \in U_\lambda, \tag{16}$$

where P is the number of filters used in the filter bank, and $u(i,\mathbf{Q}(\boldsymbol{\lambda}_p,B)) = \tilde{\mathbf{G}}(i)^* \mathbf{Q}(\boldsymbol{\lambda}_p,B)$ is the output of the pth filter.

Observe that the sequence (16) can immediately be used for comparing the statistic with the threshold and making a detection decision. However, any two values of $\tilde{u}(i)$ and $\tilde{u}(i+n)$, $0 < n \leq N_S$ are correlated since overlapping short-time data segments are used for calculating $\tilde{u}(i)$. As a result, a vocalization may yield a sequence of statistics that exceed the threshold. In almost all cases there are no practical reasons to consider more than one value of the statistic per signal. Therefore, the rate at which the statistic is calculated on the filter bank output, Eq. (16), should be reduced. We assume that each nonoverlapped data segment $\mathbf{x}(jK_0)$ may contain only one signal. Hence, only one value of the statistic per data segment should be used for the subsequent analysis. To avoid the loss of signal peaks on the filter bank output, the statistics corresponding to the data segment $\mathbf{x}(jK_0)$ can be calculated as

$$z(j) = \max \Psi(j), \quad j = 0,1, \ldots , \tag{17}$$

where the set $\Psi(j) = \{\tilde{u}(jN_S + k_j), \tilde{u}(jN_S + k_j + 1), \ldots, \tilde{u}([j+1]N_S)\}$ represents a sequence of filter bank outputs, and the index $1 \leq k_j \leq N_0$, $k_0 = 1$ is taken to ensure the minimal time interval N_0 between two adjacent values of the statistics, $z(j-1)$ and $z(j)$. This time interval can be chosen as half of the signal duration, $N_0 = 0.5N$. If actual signal duration is close to N and any two signals are separated by more than N_0 samples, the algorithm (17) elicits only one peak value of $\tilde{u}(i)$ per vocalization.

As follows from Eqs. (16) and (17), the detector calculating the statistic $z(j)$ can be specified by a pair $d = \{P, \mathbf{Q}(\boldsymbol{\lambda}_p, B)\}$ representing the number of filters P and the types of the kernels $\mathbf{Q}(\boldsymbol{\lambda}_p, B)$ used in a filter bank. For highly variable NARW contact calls, the optimal value of the parameter P does not exist. In practice, the detection performances and computational costs increase as the number of filters increase. Therefore, the parameter P should be taken based on an acceptable trade-off between detection performances and computational costs. Correspondingly, the detector optimization problem is reduced to obtaining the clear form representation for the function $\mathbf{Q}(\boldsymbol{\lambda}, B)$ and finding the optimal value of B such that the statistic $z(j)$ satisfies the condition (3).

Under an unknown class of noise distributions, Eq. (11), the analytical solution to this problem does not exist. Therefore, we propose the following numerical technique.

In a first stage, we introduce various kernel types known from the literature. Subject to Eq. (15), the kernel approximating the signal spectrogram should be considered, because the spectrographic representations of real NARW contact calls may differ from the polynomial-phase model. Therefore, we consider such kernels as rectangular, exponential, "truncated sinc" and "truncated cos." If impulsive noise is present, maximization of the criterion (3) can be achieved by suppressing noise impulses. Our observations show that a certain percentage of impulsive noises have an almost uniform PSD within the frequency band occupied by signals (see Fig. 1). To reject such noises, we apply the FHAT ("French hat") and MHAT ("Mexican hat") wavelets.[21]

TABLE I. Data sets used when testing the statistics.

Data name	Place of recording	Time of recording	Hydrophone location	Hydrophone depth (m)	Number of NARW calls detected, M_S
CCB4	Cape Cod Bay, MA	18 December 2002—18 January 2003	Latitude: N41.934° Longitude: W70.181°	31	1284
GSC02	Great South Channel	1 May 2002—28 May 2002	Latitude: N41.708° Longitude: W69.609°	150	1502
SESAC04	15 km off Savannah, GA	29 November 2004—18 February 2005	Latitude: N31.789° Longitude: W80.826°	15	1041

Let us introduce the variable $r_k(n)=kF_S/K-\tilde{f}(n,\boldsymbol{\lambda})$. Then the kernels being investigated here can be defined as follows:

$$Q_1(k,n)=\begin{cases}1 & \text{if } |r_k(n)|\leqslant B\\ 0 & \text{otherwise}\end{cases} \tag{18}$$

is the rectangular kernel;

$$Q_2(k,n)=\exp\left\{-\frac{r_k(n)^2}{2B^2}\right\} \tag{19}$$

is the exponential kernel;

$$Q_3(k,n)=\begin{cases}\operatorname{sinc}(r_k(n)/B) & \text{if } |r_k(n)|\leqslant 2B\\ 0 & \text{otherwise}\end{cases} \tag{20}$$

is the truncated sinc kernel;

$$Q_4(k,n)=\begin{cases}\cos(\pi r_k(n)/2B) & \text{if } |r_k(n)|\leqslant B\\ 0 & \text{otherwise}\end{cases} \tag{21}$$

is the truncated cos kernel;

$$Q_5(k,n)=\begin{cases}1 & \text{if } |r_k(n)|\leqslant B\\ -0.5 & \text{if } B<|r_k(n)|\leqslant 2B\\ 0 & \text{otherwise}\end{cases} \tag{22}$$

is the FHAT wavelet; and

$$Q_6(k,n)=\left(1-\left(\frac{r_k(n)}{B}\right)^2\right)\exp\left\{-\frac{r_k(n)^2}{2B^2}\right\} \tag{23}$$

is the MHAT wavelet. Here

$$\operatorname{sinc}(x)=\begin{cases}1 & \text{if } x=0\\ \sin(\pi x)/(\pi x) & \text{otherwise.}\end{cases} \tag{24}$$

Note that using a single MHAT wavelet for calculating the detection statistic was also proposed in Mellinger and Clark (2000).[12] For each kernel $\mathbf{Q}_m(\boldsymbol{\lambda},B)$, $m=1,\ldots,6$, the optimal value of the bandwidth is found in the second stage. Subject to Eq. (3), the optimal kernel bandwidth of the detector $d(B)=\{P,\mathbf{Q}_m(\boldsymbol{\lambda}_p,B)\}$ is the value $\hat{B}$ that satisfies the condition

$$\alpha(\beta|d(\hat{B}))=\max_B \alpha(\beta|d(B)). \tag{25}$$

In the final stage, the detectors $d_m(\hat{B}_m)$ with optimal kernels are tested using the criterion (3), and the optimal detector $\hat{d}=\{P,\hat{\mathbf{Q}}(\boldsymbol{\lambda}_p,\hat{B})\}$ and corresponding kernel $\hat{\mathbf{Q}}(\boldsymbol{\lambda}_p,\hat{B})\in\{\mathbf{Q}_m(\boldsymbol{\lambda}_p,B),m=1,\ldots,6\}$ are determined. The detectors $d_m(\hat{B}_m)$ comprise a finite discrete set $U_D=\{d_1,d_2,\ldots,d_6\}$. For such a set, the condition (3) can be rewritten as

$$\alpha(\beta|\hat{d})=\max_m \alpha(\beta|d_m),\quad \beta\in[0,1]. \tag{26}$$

The result of applying this technique to different empirical data sets is considered in the next section.

IV. TESTS

Three data sets described in Table I were used in our tests. The data were collected at a $F_S=2$ kHz sample rate using bottom-mounted hydrophone recorders.[4] The STFT of the digitized data was computed using a rectangular window, $K=256$ samples, and $K_{\text{ov}}=128$ overlapping samples. The signal duration was 1.024 s ($N=2048$), resulting in Eq. (14) matrixes, $\tilde{\mathbf{G}}(i)$ and $\mathbf{Q}(\boldsymbol{\lambda}_p,B)$, with $N_S=16$ columns. The robust prewhitening algorithm proposed in Ref. 10 was used to compute the normalized spectrogram $\tilde{\mathbf{G}}(i)$. The filter bank output, Eq. (16), was calculated using $P=14$ vectors $\boldsymbol{\lambda}_p\in U_\lambda$. The discrete set U_λ consisting of 271 vectors, $\boldsymbol{\lambda}$, was used, and was obtained from training data consisting of 721 NARW contact calls (see Ref. 10 for details). For each detector, the sequence of statistics was computed using Eq. (17).

The first goal of the tests was to obtain the detector satisfying the condition (26). The data sets GSC02 and SESAC04 were used as the training data sets for this purpose. The optimal detector was determined independently for each data set. An experienced human operator analyzed the data and determined the time of occurrence, t_i, of each NARW signal by visual inspection of the spectrogram. Using the data sets GSC02 and SESAC04, two signal sets were constructed from the data recordings as $U_S=\{\mathbf{x}(t_i)|_H,i=1,\ldots,M_S\}$ where M_S is the number of NARW contact calls in the data set as detected by the human operator (see Table I). For a given threshold, the empirical probability of detection was calculated as $\alpha(C)=n(C)/M_S$ where $n(C)$ is the number of signal segments for which the statistic exceeds the threshold, (i.e., the number of events for which $z(\tilde{\mathbf{G}}(i)|_H)\geqslant C$). To calculate the probability of false alarm, a number of data chunks with no detected NARW contact calls and different impulsive noise rates were used. All such chunks were 24 h long so that each kth noise set was constructed as $U_W=\{\mathbf{x}(jK_0)|_{H_0},j=j_k,j_k+1,\ldots,j_k+M_W-1\}$, where M_W

$=84\,375$ and the index j_k specifies the start time of the kth noise segment used in the tests. The false alarm probability was calculated as $\beta(C)=m(C)/M_W$, where $m(C)$ is the number of noise segments for which the statistic exceeded the threshold (i.e., the number of the events for which $z(\tilde{\mathbf{G}}(i)|_{H_0}) \geq C$). The plots of the pairs $\alpha(C)$ and $\beta(C)$ over the range of thresholds $-\infty < C < \infty$ produce the function $\alpha(\beta|d)$. This function specifies the detection performance of the detector, d, and is referred to as the receiver operating characteristic (ROC).[17,18]

In the first stage, the optimal bandwidth for each kernel type was found. For this purpose, the ROC curves were computed for the values of $B_k=\{1,3,5,10,15,20,25,30\ \text{Hz}\}$. Some weighting function windows specified by the columns of the corresponding matrix $\mathbf{Q}_m(\boldsymbol{\lambda}_p,B)$ with nearly optimal values of bandwidth are shown in Fig. 2. The results of this test revealed the following. For the conventional kernels, Eqs. (18)–(21), the optimal bandwidth, $\hat{B}_m$, depends on the total number of impulsive noise events in the sample. When the number of impulsive noise events is low, the main cause of decreased detection performance is the variability of signal parameters. In such cases, the optimal value of the kernel bandwidth is between 5 and 10 Hz. As the number of impulsive noise events increases, the optimal bandwidth of the conventional kernels [Eqs. (18)–(21)] tends to zero. This is because increased rejection of impulsive noise, and therefore increased detection performance, is achieved by decreasing kernel bandwidth. The optimal bandwidth of the FHAT and MHAT wavelets was invariant to the number of impulsive noise events. When all training data sets were considered, the optimal bandwidth of the wavelets was found to be $\hat{B}_5 \approx \hat{B}_6 \in [15,20]$ Hz. For all training data used in our tests, the kernels $\mathbf{Q}_5(\boldsymbol{\lambda}_p,\hat{B}_5)$ and $\mathbf{Q}_6(\boldsymbol{\lambda}_p,\hat{B}_6)$ based on the FHAT and MHAT wavelets provided the highest $\alpha(\beta|d)$ under any given $\beta \in [0,1]$.

In the second stage, the spectrogram-based detectors were compared with the GLRT, and the data set CCB4 was used as the test data set. The GLRT detector[10] was implemented using the same vectors $\boldsymbol{\lambda}_\mathbf{p} \in U_\lambda$, $p=1,\ldots,14$. To compute the probability of detection, the signal set, U_S, of $M_S=1284$ NARW contact calls detected by the human operator was used. The probability of false alarm was calculated from six 24 h data samples, U_W, each taken from the CCB4 data set when no calls were detected by the human operator. Because of space limitations, only two days of data recordings when the highest and the lowest number of impulsive noise events were observed (29 December and 31 December 2002, respectively) are represented here. The corresponding ROC curves are shown in Figs. 3 and 4.

Figures 3 and 4 illustrate the basic results obtained using different training and testing data. Within the set U_D, the detectors using the FHAT and MHAT wavelets provided the highest probability of detection for a given false alarm probability. In addition, these detectors outperformed the known GLRT-based detector. This fact can be explained by the following. First, only signals with relatively high SNR and visible on the spectrogram were used in our tests. For such signals, the losses in the SNR due to the STFT did not essentially affect the detection performances. Second, ambient noise included wide-band, short-duration noise transients (0.1–1 s) and having nearly uniform PSD within the frequency range of 50–150 Hz. This kind of impulsive noise produces high values of the GLRT-based statistic which results in an increase in false alarm probability for the GLRT detector. At the same time, a certain number of wide-band noise impulses were rejected by the FHAT and MHAT wavelets having the property

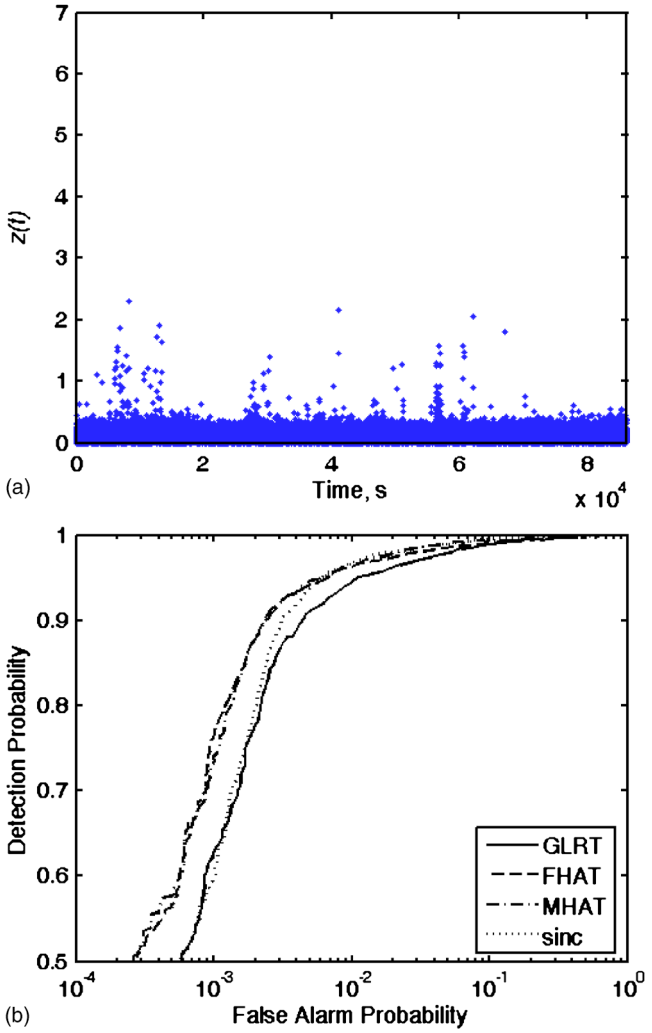


FIG. 3. (Color online) The values of FHAT-based statistic (top frame) and the ROC provided by the GLRT and the spectrogram-based detectors using FHAT, MHAT, and truncated sinc kernels (bottom frame). Data collected at Cape Cod Bay on 29 December, 2004, when the lowest impulsive noise rate was observed.

$$\sum_k Q_5(k,n) = \sum_k Q_6(k,n) = 0. \tag{27}$$

In fact, the FHAT and MHAT wavelets performed some kind of preprocessing of the data, thereby providing detector robustness in the presence of some types of ambient noise. As a result, applying FHAT and MHAT wavelets ensured better detection performance as compared with the GLRT-based detector and the spectrogram-based detector with the kernels specified by Eqs. (18)–(21).

It is of interest to note that computing the filter bank output using the FHAT wavelet is computationally more ef-

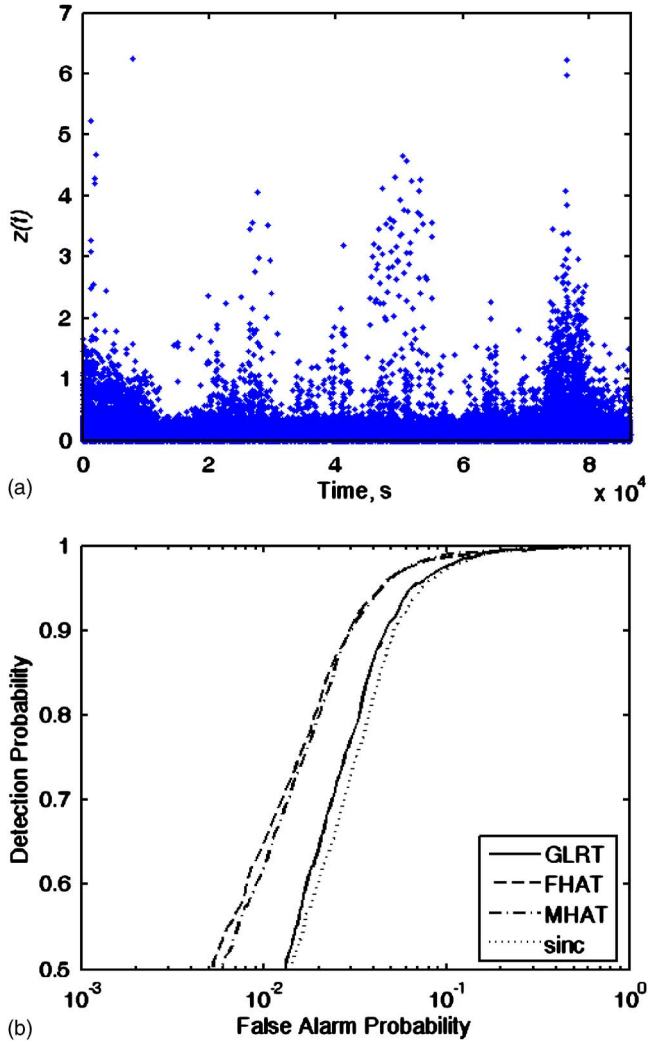


FIG. 4. (Color online) The values of FHAT-based statistic (top frame) and the ROC provided by the GLRT and the spectrogram-based detectors using FHAT, MHAT, and truncated sinc kernels (bottom frame). Data collected at Cape Cod Bay on 31 December, 2004, when the highest impulsive noise rate was observed.

ficient than the MHAT wavelet, because the FHAT wavelet requires many fewer multiplications. Indeed, if we introduce the integer variables $k(n)=[\tilde{f}(n,\boldsymbol{\lambda})K/F_S]$ and $k_B=[BK/F_S]$, then the FHAT wavelet (22) can be rewritten as

$$Q_5(k,n)=\begin{cases}1 & \text{if } k(n)-k_B\leqslant k\leqslant k(n)+k_B\\ -0.5 & \text{if } k(n)+k_B<k\leqslant k(n)+2k_B\\ -0.5 & \text{if } k(n)-2k_B\leqslant k<k(n)-k_B\\ 0 & \text{otherwise.}\end{cases} \tag{28}$$

Using Eq. (28), the pth filter output, Eq. (14), can be represented as

$$u(i,\boldsymbol{\lambda}_p)=\Sigma_1(i,\boldsymbol{\lambda}_p)-0.5\Sigma_2(i,\boldsymbol{\lambda}_p)-0.5\Sigma_3(i,\boldsymbol{\lambda}_p), \tag{29}$$

where

$$\Sigma_1(i,\boldsymbol{\lambda}_p)=\sum_{n=0}^{N_S-1}\sum_{k=\hat{k}(n)-k_B}^{\hat{k}(n)+k_B}\tilde{G}(\omega_k,i-n), \tag{30}$$

$$\Sigma_2(i,\boldsymbol{\lambda}_p)=\sum_{n=0}^{N_S-1}\sum_{k=\hat{k}(n)+k_B+1}^{\hat{k}(n)+2k_B}\tilde{G}(\omega_k,i-n), \tag{31}$$

$$\Sigma_3(i,\boldsymbol{\lambda}_p)=\sum_{n=0}^{N_S-1}\sum_{k=\hat{k}(n)-2k_B}^{\hat{k}(n)-k_B-1}\tilde{G}(\omega_k,i-n). \tag{32}$$

It follows from Eqs. (29)–(32) that calculation of the FHAT-based statistic is reduced to a summation of the normalized spectrogram elements specified by the indexes $k(n)$ and k_B.

Computational times required to calculate the FHAT-based detector were tallied using a standard PC with Pentium-IV 3 GHz processor and MATLAB 7.0 R14. For the detector with 14 filters specified by Eq. (29), the run-time was 490 s. This is about six times faster than that required by the sample-based GLRT detector[10] and about two times slower than that required by other spectrogram-based detectors.[12–15] Relative to the task of analyzing a large data set with this particular processor, detecting NARW contact calls in 10 years of data would require about 20 days of computation time.

Thus, the detector specified by Eqs. (16) and (17) and using the FHAT kernel $\hat{\mathbf{Q}}=\mathbf{Q}_5(\boldsymbol{\lambda}_p,\hat{B}_5)$ with $\hat{B}_5=15$ Hz has higher detection performance and provides a significant decrease in the run-time as compared with the GLRT-based solutions. It is important to note that this result was obtained using given sets U_S, U_W, and U_D. In general, the statistical properties of noise depend on time of day, season, habitat, type of the sensor used, and other factors, and may differ from those observed in our tests. Additionally, the admissible set of kernels is larger than that used here. As a result, kernels may exist that are different from $\mathbf{Q}_5(\boldsymbol{\lambda}_p,\hat{B}_5)$ and that have better detection performance. From this perspective, the proposed solution can be considered as locally optimal over the given sets U_S, U_W, and U_D. However, the large amount of data used in this test, as well as the similarity of the optimal detector structures obtained for the different data sets, supports the conclusion that the proposed solution is highly likely to yield robust results with high probabilities of detections and low probabilities of false alarms.

V. CONCLUSION

The problem of detecting NARW contact calls in the presence of background noise and impulsive noise was considered. To mitigate the run-time problem, the class of acceptable detectors can be constrained by the detectors implemented as a bank of 2D linear FIR filters and using the data spectrogram as the input. Test results demonstrate that the detector using the FHAT wavelet with a bandwidth of 15 Hz as a filter kernel maximizes the detection probability under a given probability of false alarm. This result can be explained by the ability of the FHAT wavelet to suppress wideband noise transients having nearly uniform PSD. Another important property of the FHAT wavelet is that its implementation in a filter bank significantly reduces computational costs.

The run-time needed to calculate the proposed detection statistic is about six times less than that required by the

GLRT detector. These properties make the detector developed in this paper an attractive solution for cases requiring detection analysis of very large data sets.

ACKNOWLEDGMENTS

The authors wish to thank M. Fowler, D. Ponirakis, A. Warde, and E. Rowland for their assistance in marking right whale calls. Thanks also to M. Chu for editing the draft version of the manuscript. The research was funded by NOAA Grant No. NA03NMF4720493.

[1]S. Kraus, M. W. Brown, H. Caswell, C. W. Clark, M. Fujiwara, P. K. Hamilton, R. D. Kenney, A. R. Knowlton, S. Landry, C. A. Mayo, W. A. McLellan, M. J. Moore, D. P. Nowacek, D. A. Pabst, A. J. Read, and R. M. Rolland, "North Atlantic right whales in crisis," Science **309**, 561–562 (2005).

[2]Department of Commerce, NOAA, "Endangered and threatened species; proposed endangered status for North Atlantic right whales," Federal Register **71** No. 248, 77704–77716 (2006).

[3]Stellwagen Bank National Marine Sanctuary, "Passive acoustic monitoring," http://stellwagen.noaa.gov/science/passive_acoustics.html. Last viewed online 7 May 2007.

[4]Cornell Lab of Ornithology, Bioacoustics Research Program, "Undersea recording: Pop-Ups," http://www.birds.cornell.edu/brp/hardware/pop-ups. Last viewed online 7 May 2007.

[5]J. N. Matthews, S. Brown, D. Gillespie, M. Johnson, R. McLanaghan, A. Moscrop, D. Nowacek, R. Leaper, T. Lewis, and P. Tyack, "Vocalization rates of the North Atlantic right whale (*Eubalaena glacialis*)," J. Cetacean Res. Manage. **3**, 271–282 (2001).

[6]K. M. Stafford, S. L. Nieukirk, and C. G. Fox, "Low-frequency whale sounds recorded on hydrophones moored in the eastern tropical Pacific," J. Acoust. Soc. Am. **106**, 3687–3698 (1999).

[7]C. W. Clark and W. T. Ellison, "Potential use of low-frequency sounds by baleen whales for probing the environment: Evidence from models and empirical measurements," in *Echolocation in Bats and Dolphins*, edited by J. Thomas, C. Moss, and M. Vater (The University of Chicago Press, Chicago, 2000) pp. 564–582.

[8]C. Clark, J. Borsani, and G. Notarbartolo-di-Sciara, "Vocal activity of fin whales, *Balaenoptera physalus*, in the Ligurian Sea," Marine Mammal Sci. **18**, 281–285 (2002).

[9]D. K. Mellinger, S. L. Nieukirk, H. Matsumoto, S. L. Heimlich, R. P. Dziak, J. Haxel, M. Fowler, C. Meinig, and H. V. Miller, "Seasonal occurrence of North Atlantic right whales (*Eubalaena glacialis*) at two sites on the Scotian Shelf," Marine Mammal Sci. **23**(4), 2007.

[10]I. Urazghildiiev and C. Clark, "Acoustic detection of North Atlantic right whale contact calls using the generalized likelihood ratio test," J. Acoust. Soc. Am. **120**, 1956–1963 (2006).

[11]I. Urazghildiiev and C. Clark, "Detection performances of experienced human operators compared to a likelihood ratio based detector," J. Acoust. Soc. Am. **122** (2007).

[12]D. Mellinger and C. Clark, "Recognizing transient low-frequency whale sounds by spectrogram correlation," J. Acoust. Soc. Am. **107**, 3518–3529 (2000).

[13]D. Mellinger, "A comparison of methods for detecting right whale calls," Can. Acoust. **32**, 55–65 (2004).

[14]L. Munger, D. Mellinger, S. Wiggins, S. Moore, and J. Hilderbrand, "Performance of spectrogram cross-correlation in detecting right whale calls in long-term recordings from the Bering Sea," Can. Acoust. **33**, 25–34 (2005).

[15]D. Gillespie, "Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram," Can. Acoust. **32**, 39–47 (2004).

[16]E. L. Lehman, *Testing Statistical Hypotheses* (Wiley, New York, 1986).

[17]H. L. Van Trees, *Detection, Estimation and Modulation Theory* (Wiley, New York, 2001), Part I.

[18]A. Hero, "Signal detection and classification," in *Digital Signal Processing Handbook*, edited by E. Madisetti and D. Williams (CRC Press, New York, 1999).

[19]Y. Harin, *Robustness in Statistical Pattern Recognition* (Kluwer Academic, Dordrecht, 1996).

[20]G. R. Arce, *Nonlinear Signal Processing: A Statistical Approach* (Wiley-Interscience, Hoboken, NJ, 2001).

[21]*Handbook of Formulas and Tables for Signal Processing*, edited by A. Poularicas (CRC Press, New York, 1999).

Underwater tunable organ-pipe sound source

Andrey K. Morozov[a] and Douglas C. Webb
Webb Research Corporation, 82 Technology Park Drive, East Falmouth, Massachusetts 02536

(Received 24 July 2006; revised 18 January 2007; accepted 31 May 2007)

A highly efficient frequency-controlled sound source based on a tunable high-Q underwater acoustic resonator is described. The required spectrum width was achieved by transmitting a linear frequency-modulated signal and simultaneously tuning the resonance frequency, keeping the sound source in resonance at the instantaneous frequency of the signal transmitted. Such sound sources have applications in ocean-acoustic tomography and deep-penetration seismic tomography. Mathematical analysis and numerical simulation show the Helmholtz resonator's ability for instant resonant frequency switching and quick adjustment of its resonant frequency to the instantaneous frequency signal. The concept of a quick frequency adjustment filter is considered. The discussion includes the simplest lumped resonant source as well as the complicated distributed system of a tunable organ pipe. A numerical model of the tunable organ pipe is shown to have a form similar to a transmission line segment. This provides a general form for the principal results, which can be applied to tunable resonators of a different physical nature. The numerical simulation shows that the "state-switched" concept also works in the high-Q tunable organ pipe, and the speed of frequency sweeping in a high-Q tunable organ pipe is analyzed. The simulation results were applied to a projector design for ocean-acoustic tomography.
[DOI: 10.1121/1.2751268]

PACS number(s): 43.30.Yj, 43.38.Ew, 43.30.Jx [JAC] Pages: 777–785

I. INTRODUCTION

Low-frequency broadband underwater sound sources are used in sonar systems, ocean-acoustic tomography, and seismic deep-penetration profiling systems.[1–3] In these systems, various physical mechanisms are applied to low-frequency signal generation, such as pneumatic, electromagnetic, magnetostrictive, electrostatic, hydroacoustic, and parametric. A variety of underwater sound sources (including Tonpilz, Helmholtz resonator, flexural, and bubble transducer.) are described in the scientific literature.[4–9] Sound-source design requires accounting for certain fundamental physical principles and problems that are inherent in underwater transducers that use arbitrary mechanisms of energy transformation. Emitted acoustic energy is proportional to a square volume velocity and frequency; to produce high-amplitude and low-frequency signals, an underwater source must generate large volume displacement. As a result, a sound source with a small radiation area has a large imaginary part of impedance, bigger than its real part.

There are two potential approaches to building a projector with high efficiency and large radiated acoustic power. The first approach is to increase the amplitude of the displacement. This is accomplished in transducers by using flexible membranes, bars, or plates of large area and dimension. Such transducers require complicated and expensive techniques and in most cases require pressure compensation.

Another option is to use high-Q resonators near the resonance frequency. The reactive part of the resonator in resonance with the radiated signal eliminates the reactance of the radiation impedance. A high ratio of radiation reactance to the real part of the radiation impedance suggests use of a high Q of the resonator. Practical experience and common sense tell us that high-Q resonant sources should have high efficiency, relatively smaller dimensions, and uncomplicated design. A narrow-frequency bandwidth is the only disadvantage of such resonant sound sources. Application of the tunable or switchable high-Q resonant sound source is one way to generate signals, which occupy a very large frequency bandwidth. The projector need not produce a linear, time-invariant broadband transformation of input signals to generate a broadband acoustic wave field. The sound source does not have to emit all components of a broadband signal simultaneously, as is necessary for linear time invariant systems. A high-Q narrow-band resonant system can generate a broadband signal if it is tuned synchronically with the instantaneous signal frequency and always maintained in a resonance state. Moreover, it is possible to use both a quick (instant) switch of a resonance frequency simultaneously with an instantaneous signal frequency shift and slow tracking of the signal frequency. As noted by Larson *et al.*,[10] Munk proposed this method in 1980 as a "state-switched" sound source concept when he was analyzing different approaches for design of high-efficiency sound sources for an ocean-acoustic tomography experiment. The "state-switched acoustic source" can switch among several different resonant states instantly and synchronically with the discrete frequency manipulated signal. If the source at any moment in time has only one fundamental resonant frequency and always maintains resonance with the signal, it radiates a highly efficient acoustic wave. Reference 10 describes this concept in detail, with the example of a simple mass-spring harmonic oscillator and a description of a prototype underwater state-switched sound source with two states, 810 and 1022 Hz.

[a]Electronic mail: moro@webbresearch.com

Although a variety of designs for high-frequency state-switched sources is described in the literature,[11–18] there is no good design for a state-switched transducer with a frequency smaller than 800 Hz. The state-switched transducer adequate for frequency coded signal transmission is used in communications and underwater sonar. Many applications require neither complex, coded-signal transmission nor simultaneous transmission of all broadband acoustic-energy spectrum components. Different frequencies transmitted in sequence, or swept signals, are appropriate when the medium under investigation does not change appreciably over the duration of the transmission. In that case, one can use relatively slow resonance frequency tuning of a high-Q system corresponding to the instantaneous frequency of the linear frequency-modulated signal. The swept frequency modulation does not degrade the time resolution. The Cramer-Rao bound for time resolution for a system operating above threshold and using a matched filter is $\sigma_{\Delta T}=1/(\sqrt{\mathrm{SNR}}W)$, where SNR is the ratio of total received energy to a spectral level of noise, and W is the total system bandwidth, not an instantaneous one, and the only implications of a FM signal is the range-Doppler coupling. In our analysis, we will distinguish total system bandwidth from the instantaneous bandwidth of narrow-band tunable resonator. The swept frequency transducer meets all the requirements of stationary systems for ocean monitoring, including ocean-acoustic tomography and seafloor monitoring. Three papers (Refs. 19–21) describe a frequency swept sound source based on a tunable organ pipe. It is a reliable, broadband, depth-independent, highly efficient sound source capable of long-term operation. This makes it superior to the marine vibroseis,[22] which is less efficient, needs pressure gas compensation, and cannot be used at large depths.

A simple and practical model of a tunable organ pipe[20] will be used further for simulation of the frequency and time responses of a tunable organ-pipe source.

This paper's objective is to consider general aspects of tunable resonant sound-source theory and to provide detailed analysis of time-response processes in the practical design of a real tunable organ pipe. The organization of the paper is as follows.

Section II provides theoretical analysis of the problem, beginning with consideration of a general concept for a "quick frequency adjustment" (QFA) filter. Suppose that a broadband signal with an arbitrarily changing phase narrow or wideband phase modulation is required on the output of a high-Q filter with the bandwidth, which can be much smaller than signal spectrum bandwidth. A QFA filter is a resonant filter, whose resonant frequency coincides at any moment with the instantaneous frequency of the input signal; an example is a high-Q tunable projector that maintains resonance with the transmitted signal at any given moment. Such a projector has high efficiency and all other advanced characteristics of high-Q resonant systems along with a broadband radiated signal spectrum. The theory of QFA filters as considered here is based on the tunable Helmholtz resonator. Our research shows that a simple tunable Helmholtz resonator can work as a QFA filter: We show analytically and by numerical simulation that the simplest lumped element tunable resonator can be successfully used as a state-switched sound source or as a frequency swept source.

In Secs. III and IV the analysis moves from the simplest lumped sound source to a complex distributed system, such as a tunable organ pipe. A numerical model of a real tunable organ pipe is developed. The electrical circuit model of the organ pipe has a form of a transmission line segment and is equivalent to high-frequency electromagnetic resonator models. This gives a general form to the results of the analysis. Computer simulation of an organ pipe conducted in Sec. IV shows that it can also be used as a highly efficient, state-switched transducer or a frequency swept projector. The analysis is then applied to design of a real sweeping transducer for acoustic tomography and global ocean monitoring. Section V presents result of tunable organ pipe experimental testing. The projector described has been used in NPAL 2004 experiments in the Pacific Ocean.

Section VI offers the summary and conclusions.

II. SIMPLEST SECOND-ORDER TUNABLE RESONANT CIRCUIT

The ability to change resonance frequency quickly is a property of any lumped element resonant system of the second-order. Indeed, the solution for simplest resonant system described by the second-order differential equation is completely determined for any time from half-space $t \geqslant t_0$ by $x(t_0)$ and its derivative $x'(t_0)=(dx/dt)(t_0)$. The system has no memory for the parameters or coefficients, and thus no memory for the resonance frequency. If the resonance frequency switches from ω_1 to ω_2, the oscillation instantly changes its wave form from $x(t)=A_1 \sin(\omega_1 t+\varphi_1)$ to $x(t)=A_2 \sin(\omega_2 t+\varphi_2)$, where $A_2=\sqrt{x^2(t_0)+x'^2(t_0)/\omega_2^2}$ and $\varphi_2=\arctan(x(t_0)\omega_2/x'(t_0))-\omega_2 t_0$ are amplitude and phase of a final oscillation. Note that if the state switches when $x(t_0)=0$ or $x'(t_0)=0$, then the phase does not change value, and amplitude jumps to the value $A_2=A_1\omega_1/\omega_2$; in all other cases, phase changes as well. To avoid additional phase manipulation during the frequency change, the original state-switched concept expected instant resonance frequency switch only when $x(t_0)=0$ or $x'(t_0)=0$. The amplitude signal hop still accompanies any frequency switching in accordance with $A_2=\sqrt{x^2(t_0)+x'^2(t_0)/\omega_2^2}$. The frequency sweeping or chirp signal can be performed by the resonance frequency switching among several frequencies.[10,11] All above-described formulas were obtained from simple requirements for continuity of signal, and its derivative. It should be noted that different variants of that principle can have specific properties, but, in any case, the state-switched process conserves energy, and any energy changing in a high-Q resonance system requires a slow transient process.

To analyze the general concept of a quick frequency adjustment filter, the system from Fig. 1 is considered. Suppose that a broadband signal with an arbitrarily changing phase is required on the output of a high-Q tunable filter. The resonant frequency of the filter is controlled by computer to match the instantaneous frequency of the input signal at all times. The QFA filtering can be implemented using a phase-locked loop (PLL) to follow the changing frequency of the

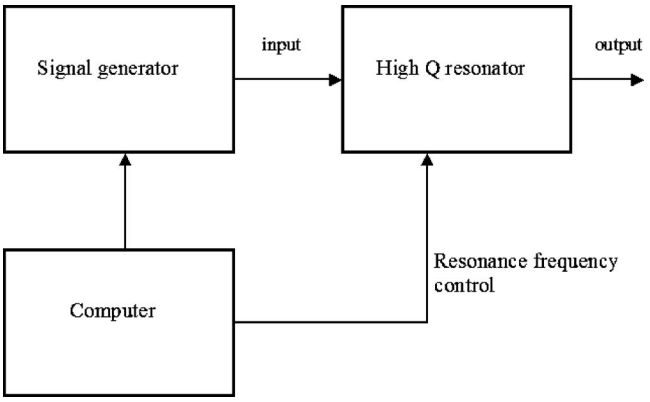


FIG. 1. The control system for QFA filter.

signal. The details of such a PLL are provided in Refs. 19–21 and are not included in this analysis. The objective of this paper is to demonstrate the ability of the QFA filter to reproduce a general broadband signal with an arbitrarily changing phase without constraints on the first- or second-order derivatives, that is, the rates of change, of that phase.

The differential equation for a simplest resonant circuit in a canonical form is

$$\frac{d^2V}{dt^2} + 2\delta\omega_r\frac{dV}{dt} + \omega_r^2 V = \omega_r^2 V_0, \tag{1}$$

where $\delta = 1/2Q$ is the loss factor; Q is the quality, or resonance, factor; and ω_r is the resonance frequency.

The Helmholtz resonator is the simplest example of a lumped element resonant system. Several papers describe tuning methods,[11–18] and some examples of tunable Helmholtz resonators can be found in others (see Refs. 23–26). In the case of the Helmholtz resonator, the wave forms V_0 and V are the volume velocities and other parameters of Eq. (1):

$$\omega_r = c\sqrt{\frac{S}{l\Omega}}, \tag{2}$$

$$Q = \frac{2\lambda l}{S}. \tag{3}$$

The Helmholtz resonator parameters are: Ω is the resonator volume, S is the throat area, l is the throat length, $\beta = 1/K$ is the compressibility of water, and K is the bulk modulus.

Our analysis does not include detailed consideration of the tunable mechanism of the Helmholtz resonator. For example, the resonator can be tuned by opening a path to an additional container with a compressible liquid. As a result, the resonance frequency $\omega_r = \omega_r(t)$ will be variable. The objective is to show the potential ability of the Helmholtz resonator to quickly adjust to the signal with a variable instantaneous frequency. This ability allows it to radiate a broadband signal $V(t) = A\sin(\varphi(t))$ with the continuous instantaneous frequency $\omega(t) = d\varphi(t)/dt$ by a quick adjustment of the resonance frequency of the high-Q, narrow-band Helmholtz resonator.

The quick adjustment condition means that any time resonance frequency of the tunable resonator $\omega_r(t)$ is equal to the instantaneous signal frequency $\omega(t)$,

$$\omega_r(t) = \omega(t) = \frac{d\varphi(t)}{dt}. \tag{4}$$

The radiated signal derivative forms are

$$\frac{dV(t)}{dt} = \frac{d\varphi(t)}{dt}A\cos(\varphi(t)),$$

$$\frac{d^2V(t)}{dt^2} = \frac{d^2\varphi(t)}{dt^2}A\cos(\varphi(t)) - \left(\frac{d\varphi(t)}{dt}\right)^2 A\sin(\varphi(t)). \tag{5}$$

Substituting these formulas into the initial equation (1) and taking into account the adjustment frequency condition (4), the form of a drive signal $V_0(t)$ becomes

$$V_0(t) = \left(\frac{d\varphi(t)}{dt}\right)^{-2}\left(2\delta\omega_r\frac{d\varphi(t)}{dt} + \frac{d^2\varphi(t)}{dt^2}\right)A\cos(\varphi(t))$$

$$= A\left(2\delta + \frac{1}{\omega^2(t)}\frac{d\omega(t)}{dt}\right)\cos(\varphi(t)). \tag{6}$$

We just proved that Eq. (1) with the right part in the form

$$\omega_r^2(t)V_0(t) = A\left(2\delta\omega_r^2(t) + \frac{d\omega(t)}{dt}\right)\cos(\varphi(t))$$

has the solution $V(t) = A\sin(\varphi(t))$. The output of the QFA filter has the form $V(t) = A\sin(\varphi(t))$ with the arbitrarily changing phase $\varphi(t)$, if the input signal has the form of Eq. (6),

$$V_0(t) = A\left(2\delta + \frac{1}{\omega_r^2(t)}\frac{d\omega(t)}{dt}\right)\cos(\varphi(t)).$$

There is no limit on the derivatives of the first or second order, and no limitation on the phase spectrum bandwidth.

In the simplest CW case, when $V(t) = A\sin(\omega_r t)$, the wave form for $V_0(t)$ is

$$V_0(t) = 2\delta A\cos(\omega_r t). \tag{7}$$

In the case of a linear frequency modulated signal (LFM), when $V(t) = A\sin(\omega_0 t + 0.5at^2)$, the wave form for $V_0(t)$ is

$$V_0(t) = (2\delta + a/(\omega_0 + at)^2)A\cos(\varphi(t)). \tag{8}$$

The output of the QFA filter in a form $V(t) = A\sin(\omega_0 t + 0.5at^2)$ can be achieved by exciting it with the wave form in Eq. (8) with no limitation of the rate a. As result, the system bandwidth can be much larger than the instantaneous bandwidth of a high-Q tunable system.

The amplitude of signal $V_0(t)$ in Eq. (6) has two components. The first component, $A\delta\omega^{-1}(t)$, is quasistationary and has the same form as in CW case. The second,

$$A\frac{1}{\omega^{-2}(t)}\frac{d\omega(t)}{dt},$$

is a changing of amplitude due to frequency changes during a time comparable with the period. That component can be positive (sweeping from low frequency to high frequency) or negative (sweeping from high frequency to low frequency). This component is essential only for very fast frequency

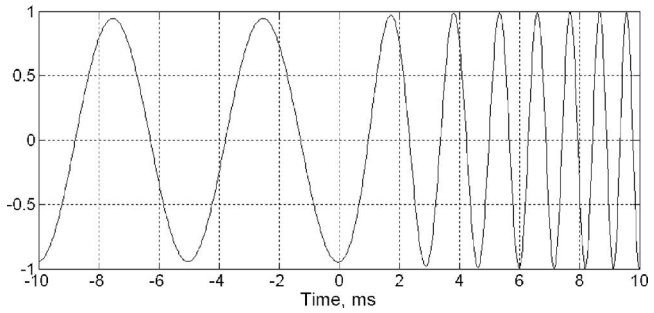


FIG. 2. The response of tunable Helmholtz resonator to fast frequency sweep from 200 to 1200 Hz for 10 ms.

changes, when from period to period it is changed 20% or more.

The numerical simulation demonstrates the ability of the simplest lumped element resonator to quickly adjust its resonance frequency to an instantaneous frequency of a broadband signal. The response of a tunable Helmholtz resonator to a fast frequency sweep of 200–1200 Hz for 10 ms is shown in Fig. 2. Figure 3 presents the response to a frequency hop of 200–1200 Hz for the same resonator. The QFA concept allows the generation of very short broadband chirp signals and generalizes the above-mentioned switched-state concept. The concept is an inherent property of second-order lumped-element resonant circuits, but it also applies to more complicated circuits, such as the organ pipe discussed in the following as an example of a distributed acoustic system.

III. TUNABLE RESONATOR TUBE

References 18–20 describe a design for a tunable, resonant organ-pipe sound source. It is a very reliable projector with the ability to radiate swept-frequency signals with high efficiency, high power, and unlimited operating depth. The projector is a freely flooded, mechanically tunable organ pipe with a Tonpilz acoustical driver. A computer-controlled electrical actuator keeps the projector in resonance with the swept-frequency signal by means of phase-lock-loop feedback. This projector combines the efficiency and simplicity of resonant tube projectors with the possibility of using wide frequency ranges.

The organ-pipe design is a configuration of two slotted resonator tubes driven by a coaxially mounted, symmetrical Tonpilz transducer. To change the resonant frequency of the

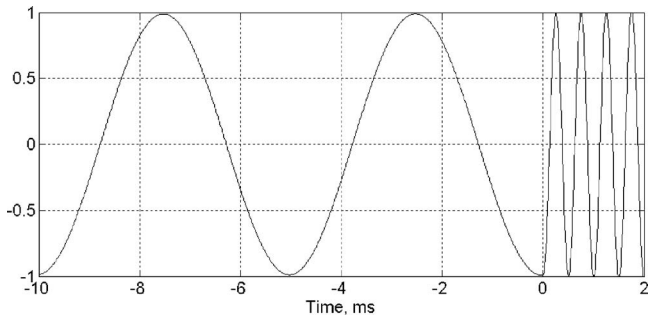


FIG. 3. The response of tunable Helmholtz resonator to instant frequency hop from 200 to 2000 Hz.

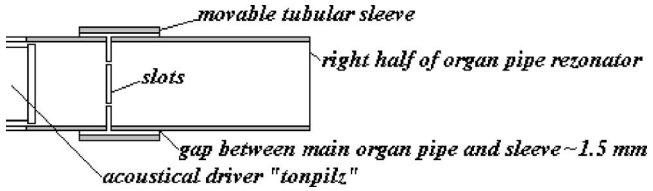


FIG. 4. Draft drawing of the tuning mechanism of organ pipe sound projector, the linear actuator is not shown.

projector, the resonator tubes were fitted with slots (or vents) at a distance of one-third the resonator tube length measured from the acoustical driver.[27] Two stiff coaxial tubular sleeves of larger diameter move axially along the resonator tubes, changing the exposure of the slots (Fig. 4). The inertia of the water layer in the gap between the two coaxial tubes depends on the position of the sleeves relative to the tube slots. The position of the sleeves causes a change in the equivalent acoustic impedance of the slots, thus changing the resonant frequency. As a result, the resonant frequency varies with the position of the sleeves relative to the slots. A computer-controlled actuator moves the sleeves and keeps the projector in resonance with a swept-frequency signal.

A simplified equivalent electrical-circuit model can be successfully applied similar to the tunable organ-pipe resonant sound source. The model is based on the similarity of mechanical differential equations and equations for ordinary electrical elements, such as capacitors, inductors, resistors, and transformers.[20] The model is simpler than the finite element analysis,[20] and it does not need special software or powerful computers to facilitate the prediction of precise projector parameters. This model was continuously compared with experimental data from the actual projector test. The comparison showed that it truly reflects organ-pipe sound physics. In this model, we did not take into account the inertia of the aluminum pipe walls, the losses in the walls, the deformation of the Tonpilz transducer shell, the radiation from the orifice, or other small details of the actual projector performance.

Let us assume that a Tonpilz acoustical driver includes m ceramic stacks in cylindrical form composed of n piezoelectric ceramic longitudinally polarized cylinders. The entire area of the ceramic stacks is A_c. The length of one ceramic cylinder is t_c and the length of all the stacks is l_c. The piezoelectric ceramic polarization direction (three by convention) is in the axial direction. The reduced constitutive relations[19] for a piezoelectric ceramic are shown in the simple equations

$$S_3 = s_{33}^E T_3 + d_{33} E_3,$$

$$D_3 = \varepsilon_3^T E_3 + d_{33} T_3, \tag{9}$$

where S_3 is the three-strain component, T_3 is the three-stress component, E_3 is the electric field in the three directions, and D_3 is the electric displacement in the three directions. The piezoelectric material properties are given by compliance s_{33}^E, piezoelectric strain coefficient d_{33}, electric permittivity ε_3^T, and density ρ_c. The ceramic stack emits sound pressure with

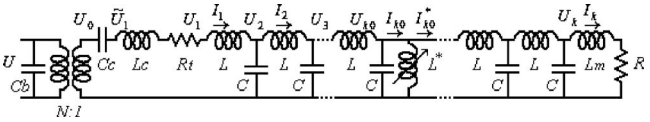


FIG. 5. Resonant organ pipe equivalent circuit.

a frequency f into the water through a piston with area A_p, length l_p, and material density ρ_p.

The acoustic Tonpilz driver is placed between a pair of open-ended, cylindrical tubes with individual length L, cross-sectional area A, and thickness h. Each section has orifices with the area A_o formed in the walls at a distance L_o from the acoustic driver. The movable coaxial sleeve has length l_s and radius g, which is larger than the radius of the main tubular section. The projector tubes are freely flooded with water (density ρ_o and sound velocity c_o).

Figure 5 shows the simplified electrical equivalent circuit for one-half of the symmetric resonator tubes, where $C_b = nA_c\varepsilon_{33}^T(1-k_{33}^2)/t_c$ is the capacity of the piezoelectric ceramic for a clamped circuit, $C_b = nA_c\varepsilon_{33}^T/t_c$ is the same capacity for an open circuit, $k_{33} = d_{33}/\sqrt{\varepsilon_{33}^E s_{33}^E}$ is the coupling coefficient (a property of the piezoelectric ceramic material), $C_c = l_c s_{33}^E A_p^2/A_c$ is the ceramics stiffness equivalent capacitor,

$$L_c = \frac{4\rho_c A_c l_c}{A_p^2\pi^2} + \frac{\rho_p l_p}{A_p}$$

is the lumped equivalent inductance of the combined ceramics and piston inertia, $R_t = (1-A_c/A_p)^2\pi\rho_o f^2/c_o$ is the radiation resistance from the Tonpilz center, $N = d_{33}A_c/(A_p s_{33}^E t_c)$ is the transformation coefficient, A_p is the piston area, $L = \rho_o d/A$ is the inertia of the water mass in the tubular section with length d, $C = C_{\text{water}} + C_{\text{wall}} = A\beta d + 2AR_a d/(Eh)$ is the combined capacitance of the wall stiffness and water compressibility, $\beta = 1/(c_0\rho_0)$ is the bulk modulus of water, $d = L/k$ is the length of one section (where k is the number of sections in the numerical model and L is the length of the tube), L^* is the variable inductance equivalent for the water inertia in the gap between the resonator tube and the movable sleeve, $L_m = 0.25\rho_o\sqrt{\pi/A}$ is the inductance of the added mass of the open resonator tube end, and $R = \pi\rho_o f^2/c_o$ is the radiation resistance.[20]

The key element of the projector design is the variable inductance L^*. The inductance is derived from the position of the movable sleeve relative to the position of the slot in the resonator tube. When the slot is completely uncovered, this inductance can be calculated from Eq. (10) for the added mass m_a of the open orifice, which is

$$m_a = 0.5\rho\sqrt{\pi}A_o^{3/2}. \tag{10}$$

When the moving sleeves close the slots, the inertia of the water in the gap between the resonator tube and the sleeve increases. The resulting dependence of the variable inductance L^* on the displacement for a circular orifice can be approximately represented by Eq. (11), where x is the displacement of the movable sleeve from the center position,

$$L^* = \frac{\rho_o}{2}\sqrt{\frac{\pi}{A_o}} + \frac{\rho_o(0.5l_s - x)(0.5l_s + x)}{2\pi r g l_s}. \tag{11}$$

A short segment of a resonator tube (Fig. 5) with length dX is simulated by an LC resonant circuit, which can be described by a simple matrix equation

$$\boldsymbol{V}_{n+1} = \boldsymbol{A}\boldsymbol{V}_n, \tag{12}$$

where

$$\boldsymbol{V}_n = \begin{bmatrix} U_n \\ I_n \end{bmatrix}, \quad \boldsymbol{V}_{n+1} = \begin{bmatrix} U_{n+1} \\ I_{n+1} \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} 1 & -i\omega L \\ -i\omega C & 1-\omega^2 LC \end{bmatrix}.$$

The admittance of the resonator tube can be calculated from the continued fraction.

$$y = i\omega C_b + N^2/(1/(i\omega C_c) + i\omega L_c + R_t + z), \tag{13}$$

where

$$z = i\omega L + 1/\Bigg(i\omega C + 1/\Bigg(i\omega L + \cdots 1/\Bigg(i\omega L + 1/\Bigg(i\omega C + \frac{1}{i\omega L^*} + 1/\Bigg(i\omega L + \cdots 1/\Bigg(i\omega C + \frac{1}{i\omega L_m + R}\Bigg)\cdots\Bigg)\Bigg)\cdots\Bigg)\Bigg).$$

The radiated volume velocity and pressure sound level P_{spl} can be calculated by matrix equation

$$U_0 = UN,$$

$$I_1 = U_0/(1/(i\omega C_c) + i\omega L_c + R_t + z),$$

$$U_1 = I_1 z,$$

$$\mathbf{V}_1 = \begin{bmatrix} U_1 \\ I_1 \end{bmatrix}, \quad \mathbf{V}_k = \begin{bmatrix} U_k \\ I_k \end{bmatrix},$$

$$\mathbf{V}_k = \mathbf{A}^{k-k_o}\begin{bmatrix} 1 & 0 \\ \dfrac{-1}{i\omega L^*} & 1 \end{bmatrix}\mathbf{A}^{k_o-1}\mathbf{V}_1,$$

$$P_{\text{spl}} = 20\,\text{Log}_{10}(\rho_o f|I_k|). \tag{14}$$

The simulation example of a tunable resonator tube was accomplished with the parameters listed in Table I.

The input projector voltage was $U = 1500$ V. The dependence of the sound pressure level re 1 μpa at the distance of 1 m versus the position of the movable sleeve is shown in Fig. 6. The displacement of the sleeve was changed in 2 mm steps. The resonant frequency depends on a fluently changing sleeve position, and can be tuned to any frequency in the 200–300 Hz range. The Q factor of the organ pipe was approximately 100. The model gives a good qualitative assessment of the bandwidth and other resonator-tube projector properties.

TABLE I. Piezoelectric-driver and resonator-tube parameters.

Piezoelectric Tonpilz driver parameters		Resonator tube parameters	
Length of one ceramic cylinder t_c (m)	0.0127	Tube length L (m)	1.397
Length of all the stacks l_c (m)	0.0762	Tube radius R_a (m)	0.1683
Entire area of the ceramic stacks A_c (m^2)	0.0081	Tube cross-sectional area A (m^2)	0.08899
Compliance s_{33}^E (m^2/N)	18.5×10^{-12}	Tube thickness h (m)	0.0095
Electric permittivity ε_{33}^T (m^2/N)	11510.2×10^{-12}	Orifice distance L_0 (m)	0.4572
Piezoelectric strain coefficient d_{33}(C/N)	253×10^{-12}	Orifice area A_0 (m^2)	0.0768
Ceramic density ρ_c (kg/m^3)	7500	Gap under sleeve g (m)	0.002
Pipe length l_p (m)	0.06985	Coaxial sleeve length l_s (m)	0.02
Material (aluminum) density ρ_p (kg/m^3)	2700	Water density ρ_0 (kg/m^3)	1005
Piston area A_p (m^2)	0.061311	Sound velocity c_0 (m/s)	1490
Number of piezoelectric ceramic cylinders n	6	Compressibility of water β (m^2/N)	4.9416e−01
Number of ceramic stacks m	4	Young modulus E (N/m^2)	6.895e+09

IV. TUNING PROCESS SIMULATION

The electrical circuit model can be used to simulate the resonant-tube tuning process. In a time domain analysis, the model (Fig. 5), can be rewritten as a differential equation system for capacitor voltages and inductance currents. Note that voltages $\tilde{U}_1(t)$, $U_1(t)$ are measured in different places on the circuit, and resistor R_t is not taken into account,

$$\frac{d\tilde{U}_1(t)}{dt} = \frac{dU_0(t)}{dt} - \frac{1}{C_c}I_1(t),$$

$$\frac{dI_1(t)}{dt} = \frac{1}{L_c + L}(\tilde{U}_1 - U_0),$$

$$\frac{dU_2(t)}{dt} = \frac{1}{C}(I_1(t) - I_2(t)),$$

$$\frac{dI_2(t)}{dt} = \frac{1}{L}(U_2 - U_3),$$

$$\frac{dU_3(t)}{dt} = \frac{1}{C}(I_2(t) - I_3(t)),$$

$$\vdots,$$

$$\frac{dI_{k0}(t)}{dt} = \frac{1}{L}(U_{k0} - U_{k0+1}) + \frac{1}{L^*}U_{k0},$$

$$\frac{dI^*_{k0}(t)}{dt} = \frac{1}{L}(U_{k0} - U_{k0+1}),$$

$$\frac{dU_{k0+1}(t)}{dt} = \frac{1}{C}(I^*_{k0}(t) - I_{k0+1}(t)),$$

$$\vdots,$$

$$\frac{dI_k(t)}{dt} = \frac{1}{L_r}U_k(t) - \frac{R}{L_r}I_k(t) \qquad (15)$$

The radiated volume velocity and sound pressure level are represented by the output current $I_k(t)$.

The implicit Crank-Nicholson method was used for numerical simulation of the ordinary differential equation system. Figure 7 presents the solutions.

The instantaneous frequency hop is presented in Figs. 7(a) and 7(b). The exciting signal is shown in gray. This signal is shifted 90° from the output because of the capacitor C_c. The signal maintains the same phase relative to the output signal during the entire process. The frequency shift in the reference signal is 100 Hz. The orifice area was instantaneously opened to provide a corresponding change in the organ-pipe resonant frequency. Engineering such a fast acoustical vent presents a special technological problem that is not considered in this paper. This type of acoustical modulation can be accomplished, for example, by blocking movable piston vibrations using magneto-rheological liquid or other technologies.[11–18] The purpose of this research is to demonstrate the possibility of such an approach. The simulation shows that simultaneously changing the excitation sig-

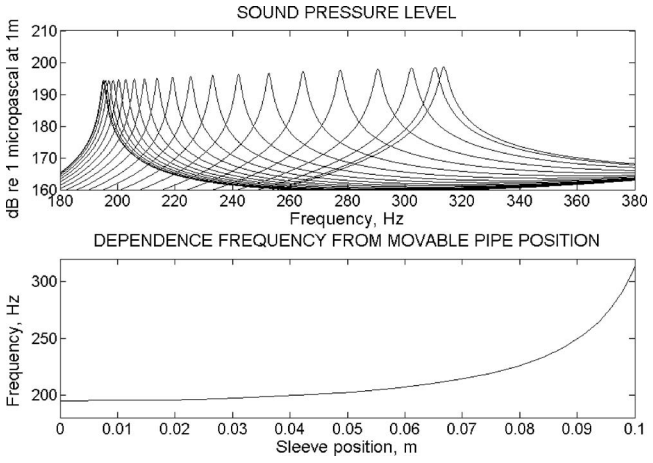


FIG. 6. The sound pressure level of a tunable resonator tube for different positions of a movable sleeve; the difference between any two displacements is 2 mm.

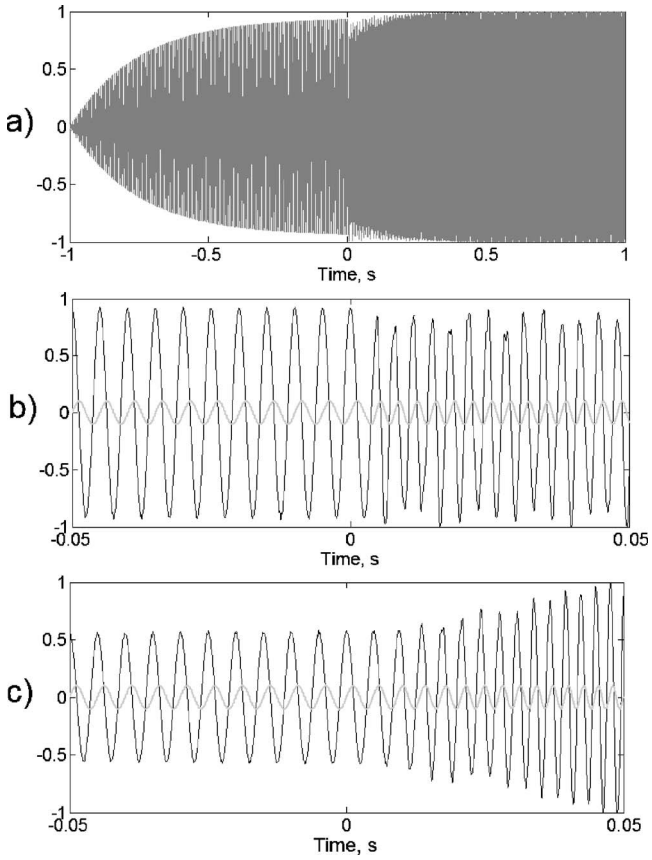


FIG. 7. (a) The transient response for a frequency hop from 200 to 300 Hz at time moment $t=0$. (b) The transient response of a tunable organ pipe for a frequency hop from 200 to 300 Hz at time moment $t=0$, large time scale. (c) The tunable organ pipe frequency sweeping from 200 to 300 Hz for 50 ms.

FIG. 8. Low frequency deepwater sound source in the SCRIPPS testing pool.

nal and the resonance frequency of a tunable organ pipe leads to a quick change in the radiated wave form. The difference with a Helmholtz resonator is an irrelevant transient process in high-frequency resonance harmonics. The high-frequency response process is approximately 100 ms, with amplitude fluctuations of about 10%. However, the high-Q organ pipe retains the ability for quick frequency hopping.

The same conclusion applies to fast frequency sweeping. A sweep from 200 to 300 Hz was simulated with the same equation system (15) and presented in Fig. 7(c). There is a small transient response to the sweeping of the sound source frequency from 200 to 300 Hz for 50 ms. These transient fluctuations can be easily corrected using phase locked loop.

V. EXPERIMENTAL TESTING OF TUNABLE ORGAN PIPES

Three identical, tunable, organ-pipe projectors with frequency swept signals were built for ocean-acoustic tomography and long-range sound propagation experiments in spring 2004. During manufacture, the wall thickness was changed from 0.0095 m, as was used during simulation, to 0.0125 m. As a result, the resonance frequency increased to 25 Hz and sources were sweeping from 225 to 325 Hz. The prototype source manufactured with a 0.0095 m wall[20] swept in the expected 200–300-Hz frequency band. Unfortunately, the first prototype used an actuator with a very low rate, which did not allow testing of high-rate frequency sweeping. The system used a tunable organ-pipe resonator with an electromechanical actuator and interior hydrophone, which were combined into a PLL controlled circuit. The linear mechanical actuator moved the coaxial sleeve to maintain resonance with the instantaneous frequency of the signal transmitted. The system was built for long-term ocean monitoring in deep water (~5000 m), and it was equipped with a battery set, a low-consumption controller, a hybrid rubidium clock, and an acoustic navigation system. Figure 8 is a recent photo of the sound-source system in the Scripps Institution of Oceanography test pool. Before deployment in 2004, the system was tested in the Seneca Lake Sonar Test Facility. The results of that test are presented in the following.

Although three sound sources were built with approximately equal parameters, only one of them will be considered in this paper. The complete sound-source system, connected to a standard measurement system, was submerged in a vertical position to 95.4 m below the surface platform. The hydrophone was located 11.41 m from the source and at the same depth.

The main parameter of the long-term system is its efficiency. The test shows that the tunable organ pipe has the expected high efficiency of an ordinary organ pipe and approximately the same directivity with 3 dB gain in a vertical plane. In a horizontal plane, the source, with attached rigid electronics housing, is omnidirectional. Figure 9 shows the efficiency of the sound source. Note that this efficiency was

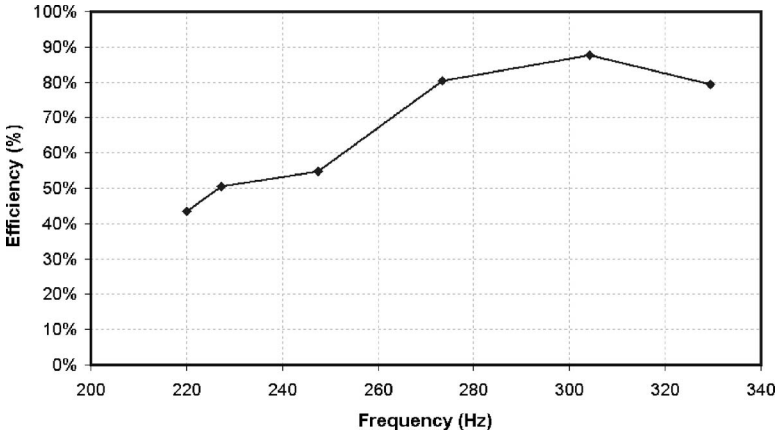


FIG. 9. The efficiency of sound source vs frequency in an isotropic approximation.

measured by an automatic test system, which plots the efficiency for omnidirectional or isotropic projectors. The actual organ pipe was closer to the dipole than to the omnidirectional source and, hence, its real efficiency was smaller. Nevertheless, even after correction for directivity, the efficiency of the source is very high.

The sound-source system incorporated a powerful Tecnadyne Inc. linear actuator capable of a 5 cm/s maximum stroke rate. To sweep the frequency over 100 Hz, the actuator needed to shift the sleeve 10 cm. To save power the actuator was used at one-third of its rated voltage; nevertheless, it could move 8 cm in approximately 3 s, allowing us to test for high-rate sweeping and to check the above-described theory.

We conducted sound-source testing for three different durations of 100 Hz frequency sweeping: 135, 10, and 5 s. All tests were done with the PLL system active. Transmission of each broadband 225 to 325 Hz swept frequency signal begins with a section of CW signal with a 225 Hz carrier frequency. This signal is used to adjust the resonant frequency of the system to the start position before frequency sweeping begins. The PLL feedback continues to function over the period of the swept signal transmission, keeping the resonant frequency in compliance with the instantaneous signal frequency. Figure 10 shows the spectrogram of the three signals with the different frequency-sweeping rates. The signals were transmitted in a series with a small interval between them. The flat part of the spectrogram plots the start position's adjustment period.

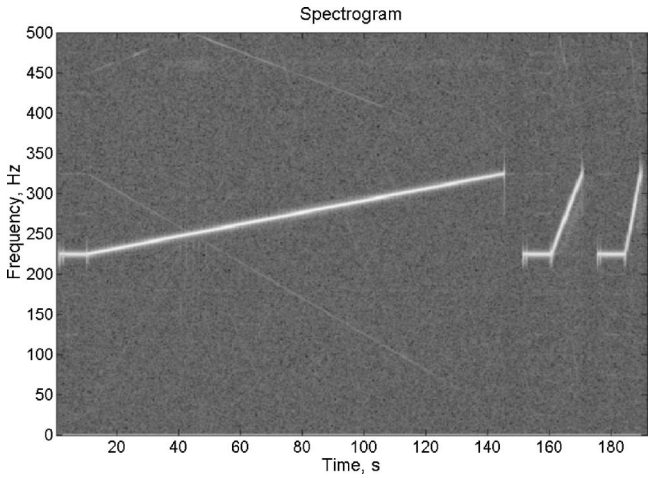


FIG. 10. Spectrogram for 135, 10, and 5 s sweep over 100 Hz bandwidth.

The correlation functions between the reference signal and the actual transmitted signal are shown in Fig. 11. A solid line plots theoretical dependence in Fig. 11, which demonstrates very good agreement between theoretical and experimental correlations. The internal digital controller analyzed and recorded PLL error, which was only a few degrees. The error increased for the initial part of 5 s frequency sweeping, which explains the difference in theoretical and experimental correlations in Fig. 11. This initial PLL error increased when the test was run for a 3 s sweep. The PLL worked for the 3 s, 100 Hz frequency swept signal but exhibited a large phase error at the beginning that resulted in a decrease in amplitude. The effect can be explained by the movable sleeve's inertia. A more powerful motor was needed to move the heavy sleeve at the necessary speed. Nevertheless, after 1–1.5 s of acceleration, the sleeve reached the necessary position and speed, and it worked well during the last part of the signal. Based on our analysis and experimental research, we conclude that, with a light sleeve made of composite carbon-fiber materials, the system can provide frequency sweeping in a 100 Hz band for a fraction of a second. After testing in Seneca Lake, two deep-water sound sources were successfully used in the SPICE04 and LOAPE (2004) experiments.

VI. CONCLUSION

Computer simulation and field testing of a low-frequency sound source with a variable resonant frequency shows that this design is highly efficient and exhibits significant ability for fast frequency change. A resonant tube sound source with a computer-controlled resonant frequency can be used for radiating broadband swept frequency signals at the rate of 100 Hz for a few seconds and with a lighter sleeve for a fraction of a second. Computer control enables holding the resonant frequency in compliance with the instantaneous signal frequency with a very small error. The correlation function of the linear frequency modulated signal is very close to the theoretical one. The frequency bandwidth of such a projector can reach a value of 0.7 to 0.8 of the central frequency. This sound-projector system is easily deployed and can op-

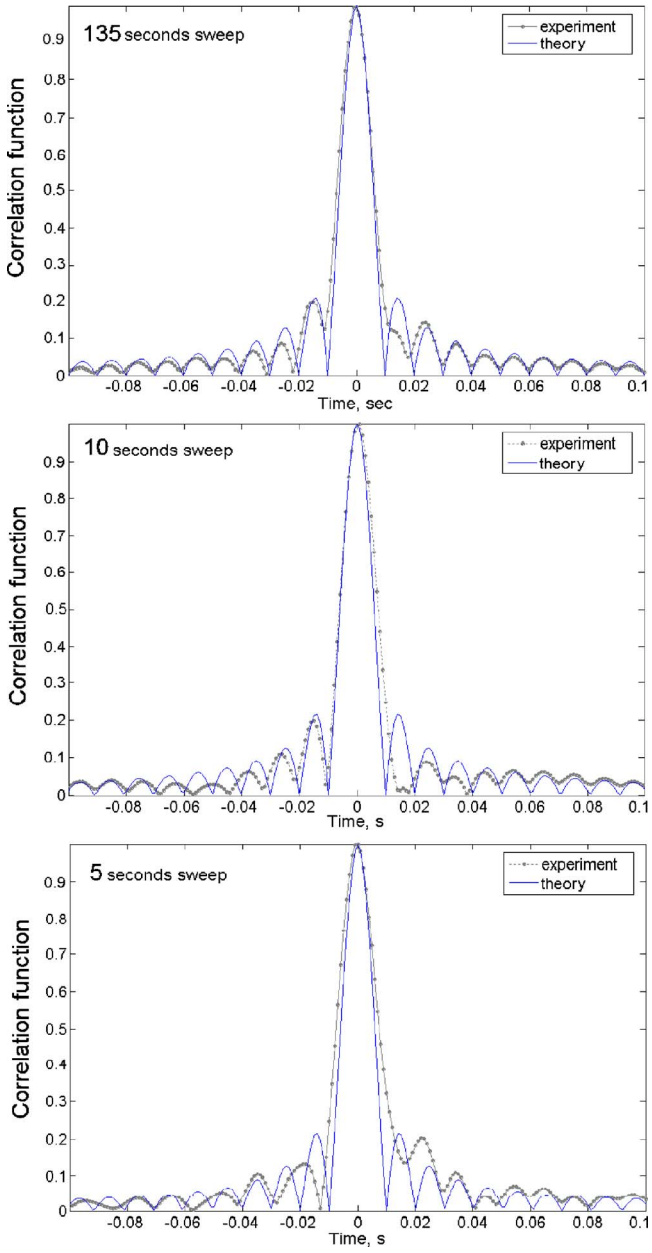


FIG. 11. (Color online) Correlation function for 135, 10, and 5 s, 225–325 Hz frequency sweep. Seneca Lake Sonar Test Facility, 3 May 2004.

erate on alkaline batteries for a long-term period. It is recommended for deep-water research, such as ocean-acoustic tomography and deep-penetration seismic profiling.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support and help of Dr. P. Worcester, Dr. W. Munk from Scripps Institution of Oceanography, and T. Ensign, the president of Engineering Acoustics, Inc., Orlando, FL. The work was supported by ONR.

[1]W. H. Munk and C. Wunsch, "Ocean acoustic tomography: A scheme for large-scale monitoring," Deep-Sea Res., Part A **26**, 123–161 (1979).
[2]P. F. Worcester, R. C. Spindel, and B. W. Howe, "Reciprocal acoustic transmissions: Instrumentation for mesoscale monitoring of ocean currents," IEEE J. Ocean. Eng. **10**, 123–137 (1985).
[3]G. R. Potty, J. H. Miller, J. F. Lynch, and K. B. Smith, "Tomographic inversion for sediment parameters in shallow water," J. Acoust. Soc. Am. **108**, 973–986 (2000).
[4]T. H. Ensign and D. C. Webb, "Electronic performance modeling of the gas-filled bubble projector," *Proceedings of the Third International Workshop on Transducers for Sonic and Ultrasonics*, 6–8 May, Orlando, FL, pp. 268–275, 1992.
[5]R. S. Woollett, "Basic problems caused by depth and size constraints in low-frequency underwater transducers," J. Acoust. Soc. Am. **68**, 1031–1037 (1980).
[6]O. B. Wilson, *Introduction to the Theory and Design of Sonar Transducers* (Peninsula, Los Altos, 1988).
[7]G. W. McMahon, "Performance of open ferroelectric ceramic rings in underwater transducers," J. Acoust. Soc. Am. **36**, 528–533 (1964).
[8]J. B. Lee, "Low-frequency resonant-tube projector for underwater sound," in *Proceedings IEEE Ocean'74*, Nova Scotia, Halifax, 21–23 August, Vol. **2**, pp. 10–15, 1974.
[9]T. J. Rossby, J. Ellis, and D. C. Webb, "An efficient sound source for wide area RAFOS navigation," J. Atmos. Ocean. Technol. **10**, 397–403 (1993).
[10]G. D. Larson, P. H. Rogers, and W. Munk, "State switched transducers: A new approach to high-power, low-frequency, underwater projectors," J. Acoust. Soc. Am. **103**, 1428–1441 (1998).
[11]H. A. B. Alwi, J. R. Carey, and B. V. Smith, "Chirp response of an active-controlled thickness-drive tunable transducer," J. Acoust. Soc. Am. **107**, 1363–1373 (2000).
[12]G. A. Steel, B. V. Smith, and B. K. Gazey, "Tunable sonar transducer," Electron. Lett. **22**, 758–759 (1986).
[13]G. A. Steel, B. V. Smith, and B. K. Gazey, "Active electronic-control of the response of a sonar transducer," Proc. Inst. Acoust. **9**, 79–87 (1987).
[14]S. K. Jain and B. V. Smith, "Tunable sandwich transducer," Electron. Lett. **24**, 311–312 (1988).
[15]W. Chenghao and Z. Zheying, "Principle of piezoelectric-tunable transducer," Chin. J. Acoust. **2**, 16–24 (1983).
[16]B. A. Kasatkin and N. Y. Pavin, "Piezoelectric transducer with controlled response characteristics," Sov. Phys. Acoust. **29**, 418–419 (1983).
[17]H. A. B. Alwi, B. V. Smith, and J. R. Carey, "Tunable transducers," Proc. Inst. Acoust. **17**, 173–182 (1995).
[18]H. A. B. Alwi, B. V. Smith, and J. R. Carey, "Factors which determine the tunable frequency range of tunable transducers," J. Acoust. Soc. Am. **100**, 840–847 (1996).
[19]D. C. Webb, A. K. Morozov, and T. H. Ensign, "A new approach to low frequency wide-band projector design," *Proceedings of Oceans*, 2002, pp. 2342–2349.
[20]A. K Morozov and D. C. Webb, "A sound projector for acoustic tomography and global ocean monitoring," IEEE J. Ocean. Eng. **28**, 174–185 (2003).
[21]A. K. Morozov and D. C. Webb, "Underwater sound source with tunable resonator for ocean acoustic tomography," J. Acoust. Soc. Am. **116**, 2635 (2004).
[22]L. D. Ambs and J. J. Sallas, "Marine seismic source," J. Acoust. Soc. Am. **110**, 651 (2001).
[23]F. Liu, S. B. Horowitz, T. Nishida, L. N. Cattafesta, and M. Sheplak, "A tunable electromechanical Helmholtz resonator," Ninth AIAA/CEAS Aeroacoustics Conference and Exhibit 2003.
[24]M. Sheplak, L. Cattafesta, T. Nishida, and S. B. Horowitz, U. S. Patent No. 6,782,109, 2004.
[25]C. B. Birdsong and C. J. Radcliffe, "A compensated acoustic actuator for systems with strong dynamic pressure coupling," J. Vibr. Acoust. **121**, 89–94 (1999).
[26]K. Nagaya, Y. Hano, and A. Suda, "Silencer consisting of two-stage Helmholtz resonator with auto-tuning control," J. Acoust. Soc. Am. **110**, 289–295 (2001).
[27]B. L. Fanning and G. W. McMahon, U.S. Patent No. 4,855,964, 8 July 1988.

Design guidelines of 1-3 piezoelectric composites dedicated to ultrasound imaging transducers, based on frequency band-gap considerations

M. Wilm,[a] A. Khelif, V. Laude, and S. Ballandras
Institut FEMTO-ST, Department LPMO, CNRS UMR 6174, 32 avenue de l'Observatoire, 25044 Besançon Cedex, France

(Received 22 May 2006; revised 21 May 2007; accepted 23 May 2007)

Periodic piezoelectric composites are widely used for imaging applications such as biomedical imaging or nondestructive evaluation. In this paper such structures are considered as phononic crystals, and their properties are investigated with respect to periodicity. This approach is based on the investigation of band gaps, that strongly depend on the properties of the considered composites (geometry, size, nature of materials). It is motivated by the fact that band gaps in principle allow one to excite the thickness mode without exciting other parasitic propagating waves. The used plane-wave-expansion method has already been applied to periodic piezoelectric composites, but, in contrast to previous approaches, not only waves propagating in the symmetry plane of the composite are considered, but also waves propagating with a nonzero angle of incidence with this plane. The method is applied to a representative 1-3 connectivity piezocomposite in order to demonstrate its potentialities for design purposes. The evolution of band gaps is explored with respect to the wave vector component parallel to piezoelectric transducer-rod axis. All bulk waves that contribute to the setting up of plate modes in the vicinity of the thickness mode are found and identified. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2749462]

PACS number(s): 43.38.Ar, 43.35.Cg, 43.38.Hz [AJZ] Pages: 786–793

I. INTRODUCTION

Modern ultrasound probes for biomedical imaging or nondestructive evaluation are currently based on piezoelectric composites, such as well-known 1-3 connectivity piezocomposites.[1] 1-3 piezocomposites are two-dimensional arrays of piezoelectric ceramic rods embedded in a polymer matrix. They are known to overcome some limitations thanks to the combination of the physical properties of their different component materials.[2,3] They exhibit a lower acoustic impedance than full piezoelectric ceramic plates thanks to the polymer phase, allowing for an easier and more efficient impedance matching with media like water or organic tissues. The bandwidth is improved with the increase of the electromechanical coupling beyond that of the full ceramic plate. The polymer phase finally allows one to reduce cross talks between elements, and to shape the piezoelectric active material on curved surfaces to focus the ultrasound beam. Composites based on single crystals are also currently studied, and are capable of exhibiting higher electromechanical coupling factor yielding an increased bandwidth when in operating conditions.[4]

Due to their heterogeneous—namely massively periodic—structure, piezocomposites exhibit Bragg-diffraction modes, also called lateral modes.[5–7] Considering thick composite plates, compared to the period of the composite, lateral modes are due to transverse waves, polarized along the rod axis, that propagate along the symmetry plane of the structure, and that partially reflect on ceramic/resin interfaces. Lateral modes appear when the wavelength becomes of the order of the composite period. By decreasing the thickness-over-width ratio, they couple with the pure longitudinal compression mode, so that high aspect ratios of piezoelectric rods are required to prevent the appearance of lateral modes in the operation bandwidth. Lamb-like waves can also propagate along composites and may be excited. Composite properties intrinsically depend on size and shape of their active inclusions and on materials. In particular, composites exhibit frequency band gaps—i.e., frequency ranges where no wave can propagate without vanishing—which width highly depends on the composite physical characteristics. These specific properties have to be taken into account during their conception, especially to take advantage of band gaps in order to avoid or to lower parasitic modes.

We have recently proposed an extended plane-wave-expansion (PWE) method, able to take into account piezoelectric and anisotropic materials.[8] Usual dispersion curves and harmonic admittance of 1-3 connectivity piezocomposites were computed by this mean. Such dispersion curves only account for waves propagating parallel to the symmetry plane of the composite without normal component. A finite-element analysis coupled to a boundary-element method has then been proposed to compute harmonic responses of any periodic ultrasonic transducers, as well as mutual responses related to cross-talk phenomena between elementary cells. This method has been applied to 1-3 piezocomposites[9] as well as to micromachined ultrasonic transducers.[10] Assuming plane radiation surfaces, it allows one to obtain qualitative and quantitative information about the expected electrome-

[a]Author to whom correspondence should be addressed. Now with Imasonic in Besançon, France. Electronic mail: m.wilm@orange.fr

chanical response of a specific periodic transducer, by accounting for a backing, for matching layers, or even for a radiation medium. Nevertheless, each analysis of a given transducer takes relatively long computation times, especially two-dimensional periodic transducers that require to consider three-dimensional geometries. Prior to the comprehensive study of the response of a specific composite-based ultrasound probe, it is helpful to make use of simulation tools allowing for a systematic investigation of composite properties related to its physical and geometrical characteristics.

Considering piezocomposites as phononic crystals,[11,12] we propose in this paper a different use of the plane-wave-expansion approach to investigate theoretically the influence of geometries and materials on the composite behavior, in particular on the existence and width of phononic band-gaps. Practically, it consists in studying the evolution of dispersion curves by taking into account a nonzero normal component of waves—considering the symmetry plane of piezocomposites—propagating in an infinite composite structure. Lateral modes, experimentally observed in composite plates dedicated to ultrasound probes, have always been considered as purely transverse waves propagating in the symmetry plane of the composite and polarized along the inclusion axes, when studied with a plane-wave expansion. Actual observed lateral modes in piezocomposites have a longitudinal component along the rod axes, and their behavior, depending on this component, is presented.

Section II summarizes the plane-wave-expansion analysis, extended to account for nonzero normal component of the wave vector. The method is then applied to an usual biperiodic 1-3 piezocomposite in Secs. III and IV, for which the evolution of band gaps is shown and an analysis of the piezoelectrically coupled waves is performed.

II. BRIEF REVIEW OF THE PLANE-WAVE-EXPANSION METHOD

A detailed description of the so-called PWE method for piezoelectric materials was reported in Refs. 8 and 13.

According to the Bloch-Floquet theory, any electromechanical field $h(\boldsymbol{r},t)$ propagating in periodic structures can be expressed as infinite series whatever the dimension of the periodicity

$$h(\boldsymbol{r},t)=\sum_{\boldsymbol{G}} h_{\boldsymbol{G}}(\boldsymbol{k},\omega)\exp(j(\omega t-\boldsymbol{k}\cdot\boldsymbol{r}-\boldsymbol{G}\cdot\boldsymbol{r})), \tag{1}$$

where $\boldsymbol{r}=(x_1,x_2,x_3)^T$, $\boldsymbol{k}$ is the wave vector and $\boldsymbol{G}$ are the vectors of the reciprocal lattice,[14] h stands for either the displacements u_i, the stresses T_{ij}, the electric potential ϕ, or the electric displacement D_i. Similarly, material constants (density, elastic, piezoelectric, and dielectric tensors) are expanded as Fourier series. Each elementary cell of the composite material can consist of several inclusions with different sizes and shapes, trapped in a general matrix.[15]

Separately inserting Bloch-Floquet and Fourier expansions in the usual constitutive relations of piezoelectricity, and in the fundamental equation of dynamics and Poisson's equation for insulating media, yields two very compact systems

$$j\tilde{\boldsymbol{T}}_i(\boldsymbol{k},\omega)=\tilde{A}_{ij}\Gamma_j\tilde{\boldsymbol{U}}(\boldsymbol{k},\omega)\quad(i=1,2,3), \tag{2}$$

$$\omega^2\tilde{R}\tilde{\boldsymbol{U}}(\boldsymbol{k},\omega)=\Gamma_i(j\tilde{\boldsymbol{T}}_i(\boldsymbol{k},\omega)), \tag{3}$$

where $\tilde{R}$ and $\tilde{A}_{ij}$ are the spectral mass-density and material-constant matrices, respectively. The diagonal matrices Γ_i contain the components of the wave vector and of the reciprocal-lattice vectors. $\tilde{\boldsymbol{T}}_i(\boldsymbol{k},\omega)$ and $\tilde{\boldsymbol{U}}(\boldsymbol{k},\omega)$ are vectors containing successively spectral coefficients $\boldsymbol{T}_{i\boldsymbol{G}}(\boldsymbol{k},\omega)$ and $\boldsymbol{u}_{\boldsymbol{G}}(\boldsymbol{k},\omega)$ of $\boldsymbol{T}_i(\boldsymbol{r},t)=(T_{i1},T_{i2},T_{i3},D_i)^T$ and $\boldsymbol{u}(\boldsymbol{r},t)=(u_1,u_2,u_3,\phi)^T$, after truncation of series to a finite number of terms.

From these equations, it is possible to obtain, after some algebra, an algebraic system similar to that of Fahmy-Adler[16,17] and Peach[18] in the case of finite- and semi-infinite-thickness stratified structures.[8] In the case of an infinite composite material, Eqs. (2) and (3) directly yield the generalized eigenvalue problem

$$\omega^2\tilde{R}\tilde{\boldsymbol{U}}(\boldsymbol{k},\omega)=\Gamma_i(\boldsymbol{k})\tilde{A}_{ij}\Gamma_j(\boldsymbol{k})\tilde{\boldsymbol{U}}(\boldsymbol{k},\omega). \tag{4}$$

Finally, generalized displacement vectors can be computed for each solution (r) of Eq. (4) following

$$\boldsymbol{u}^{(r)}(\boldsymbol{r},t)=B^{(r)}\mathrm{e}^{j(\omega^{(r)}t-\boldsymbol{k}\cdot\boldsymbol{r})}\sum_{l=1}^{L}\mathrm{e}^{-j\boldsymbol{G}^{\ell}\cdot\boldsymbol{r}}\boldsymbol{u}_{\boldsymbol{G}^{\ell}}^{(r)}(\boldsymbol{k},\omega^{(r)}), \tag{5}$$

where the coefficients $B^{(r)}$ depend on the normalization used in the solver. All these developments still are available whatever the dimension of the periodicity (one-, two-, or three-dimensional periodic composite material).

In next sections, we deal with the case of an usual (1-3)-connectivity piezocomposite, as represented in Fig. 1. Computations have been performed considering 16×16 terms in each of the Fourier and Bloch-Floquet series, ensuring a satisfying convergence.

III. "IN-PLANE" DISPERSION CURVES

The considered piezoelectric composite, representative of active materials used in modern ultrasound probes, consists of square-section piezoelectric ceramic (PZT) rods (P1-88 from Saint-Gobain Quartz et Silice) embedded in an epoxy matrix. The period d_1 of the composite is arbitrarily fixed to 100 μm for a PZT-rod width equal to 70 μm. Practically, the important parameter is the filling fraction of PZT rods equal to 0.49.

The so-called "in-plane" dispersion curves are first discussed. Usual dispersion curves, used to compute frequency positions of lateral modes,[6] correspond to waves propagating parallel to the plane of the structure (plane (x_1,x_2)) as represented in Fig. 1(a), meaning that k_3—the wave-vector-$\boldsymbol{k}$ component normal to the piezocomposite plane—is zero. Considering lines of high symmetry of the studied composite, such dispersion curves are represented in the first Brillouin zone[14] reported in Fig. 2.

This kind of representation is sufficient to exhibit absolute frequency band gaps, in which no wave can propagate whatever the polarization. Figure 3 shows such in-plane dispersion curves for $k_3=0$ (Fig. 3(a)), and also dispersion

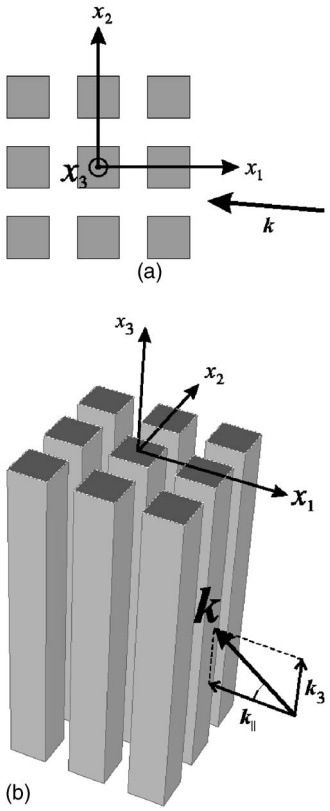


FIG. 1. 1-3 composite consisting in infinite square-section rods embedded in a matrix. A wave—wave vector $\boldsymbol{k}$—propagates (a) along the symmetry plane (x_1,x_2) of the composite, (b) with a nonzero incidence angle.

curves for nonzero values of k_3, corresponding to waves propagating with an incidence angle with the composite plane as indicated in Fig. 1(b). The interest in such dispersion curves is that piezoelectric composites principally vibrate along x_3 in usual conditions of operation.

Because results only depend on the filling fraction, we define a normalized wave vector $\gamma=\boldsymbol{k}d_1/2\pi$. Figure 3 shows the frequency evolution of modes for γ_3 equal to 0, 0.12, 0.16, 0.18, 0.19, 0.2, and then 0.4 and 0.8. One can recognize in Fig. 3(a), along the $\Gamma-X$ path (propagation along x_1), the well-known first three modes, namely the shear horizontal (SH), shear vertical (SV), and longitudinal (L) modes (polarized along x_2, x_3, and x_1, respectively). The fourth branch denoted (T) is similar to a torsional mode, for which PZT rods exhibit a rotating displacement around their axes. These modes are also identified for the $\Gamma-M$ path (propagation along the diagonal). The same kind of identification is more difficult for the $X-M$ path, located on the edge of the first Brillouin zone, since waves are stationary along x_1 with two neighboring cell lines vibrating in opposite phase. One can also comment on the existence of absolute band gaps which move as k_3 increases. Their behavior is described in Sec. IV A.

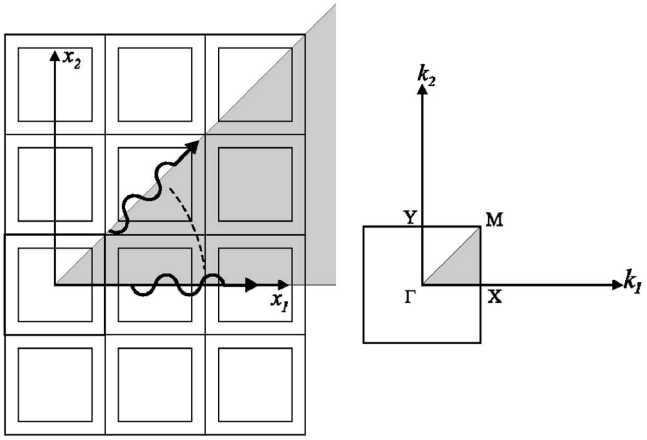


FIG. 2. Wave propagation in the first Brillouin zone, by taking into account composite and material symmetries. The path $\Gamma-X-M-\Gamma$ accounts for phononic band gaps of the structure.

Following the evolution of modes in Figs. 3(b)–3(f) between $\gamma_3=0.1$ and $\gamma_3=0.2$, one can observe crosses between branches with a particular behavior at points X and M. Exchanges of modes can indeed occur between two different paths, for instance, $\Gamma-X$ and $X-M$, when two branches cross each other at point X. Two of these crossing points are reported in Figs. 3(c) and 3(d), and one can follow the inversions of modes numbered from 1 to 3. Such a behavior of dispersion curves is caused by changes in the principal polarization of modes related to their propagation direction. Waves first propagate along a certain direction in the (x_1,x_2) plane for $k_3=0$, and finally propagate principally along x_3 for enough high values of k_3, so that the polarization direction of modes changes when k_3 increases.

For $\gamma_3=0.4$ (Fig. 3(g)), the propagation direction is principally along x_3. The first two modes correspond to quasi-transverse modes, polarized along x_2 (F2) and x_1 (F1), respectively, when traveling through path $\Gamma-X$. They correspond to a flexural behavior of the PZT rods. SH and SV modes for $\gamma_3=0$ become flexural modes, i.e., shear waves propagating along the rod axes. The third and fourth branches are torsional (T) and quasi-longitudinal (L) modes, respectively. The longitudinal mode (L) propagating along the plane of the composite becomes a quasi-longitudinal mode propagating along rod axes. One can note that the quasi-longitudinal mode is located in an absolute band gap. Finally, for $\gamma_3=0.8$ (Fig. 3(h)), the quasi-longitudinal mode disappears due to coupling with lateral modes which is explained in Sec. IV B. The operation mode in ultrasound imaging or nondestructive testing is the longitudinal vibration along rod axes with the requirement to avoid or minimize parasitic contributions such that Lamb-like modes propagating along the composite plane or lateral modes due to Bragg diffraction.

IV. "OUT-OF-PLANE" DISPERSION CURVES

A. General structure

In designing composites for imaging applications, we intend to isolate the thickness mode from any other parasitic mode in order to minimize cross coupling. In that way, we first investigate the evolution of absolute frequency band gaps, observed previously, when the propagation parameter γ_3 along the normal to the piezocomposite plane varies. One performs a projection of in-plane dispersion curves into the plane (f,k_3), as illustrated in Fig. 4.

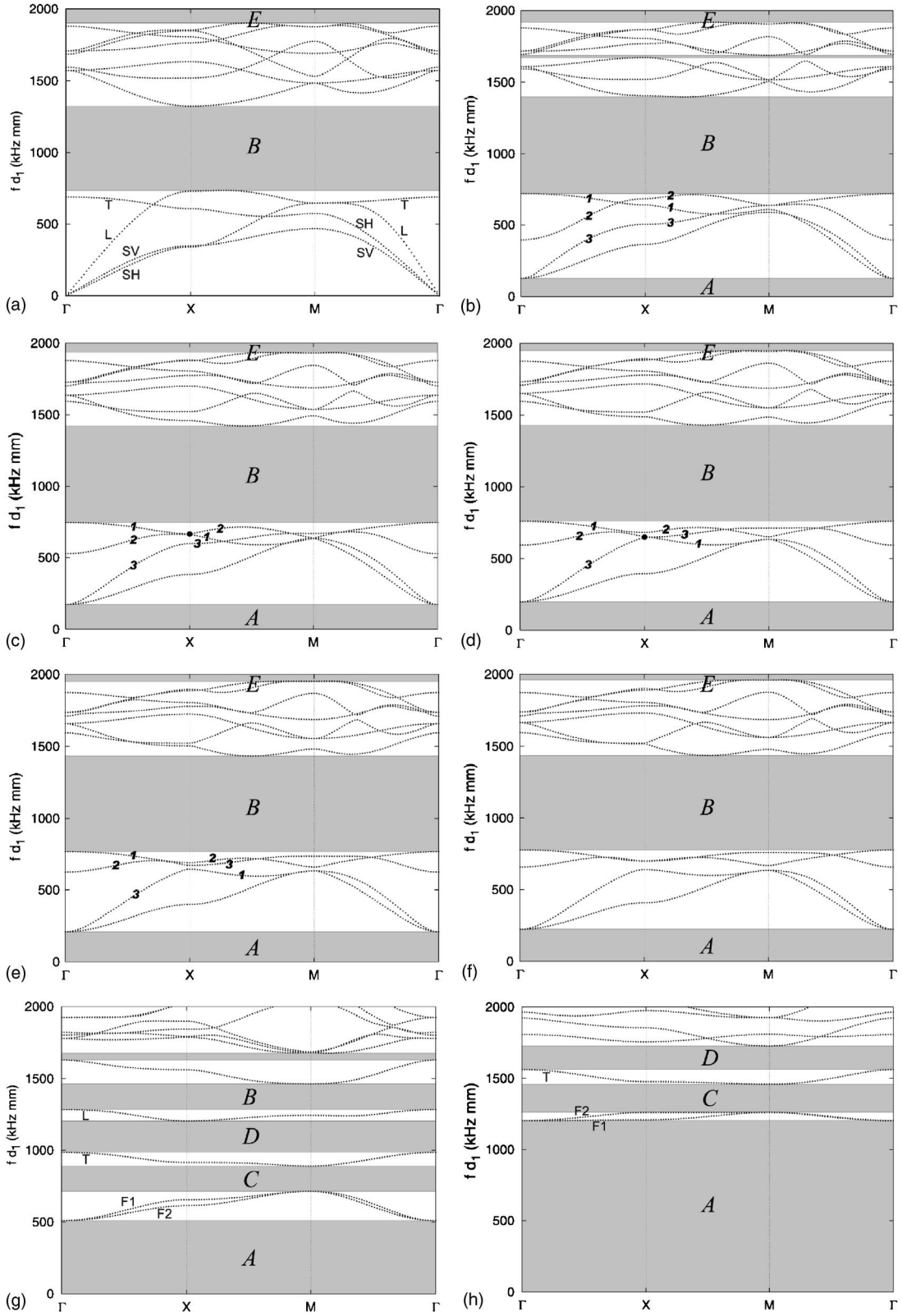


FIG. 3. In-plane dispersion curves for different values of γ_3: (a) 0, (b) 0.12, (c) 0.16, (d) 0.18, (e) 0.19, (f) 0.2, (g) 0.4, and (h) 0.8. Gray areas indicate band gaps.

Results are reported in Fig. 5, and a close-up in Fig. 6. For this diagram only, 10×10 terms in series have been used in computations, because of time consumption. Band gaps labeled from *A* to *E* in Fig. 3 are indicated. Considering all possible incidence angles of propagation and not only propagation in the plane of the composite, one first notes that there are not absolute frequency band gaps any more, i.e., no frequency range where no mode propagates whatever the value

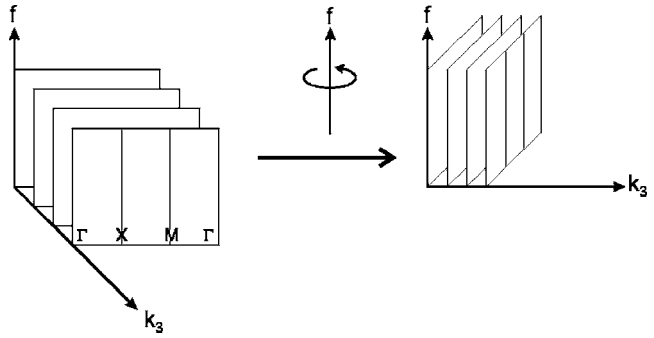


FIG. 4. Projection of in-plane dispersion curves into the plane (f,k_3) in order to obtain out-of-plane dispersion curves.

of γ_3. Nevertheless, we continue to call "band gaps" labeled zones where no wave can propagate for a given value of γ_3, whatever γ_1 and γ_2.

In the specific case of the studied composite, the velocity of the bulk-epoxy shear wave is the minimal velocity that a wave can reach either in epoxy or in PZT. Due to the nature of the composite, band gap *A* extends beyond this intrinsic limit. The other band gaps move in frequency, vary in width, and finally disappear when γ_3 increases.

Figures 3(g) and 3(h) allow one to distinguish branches numbered from 1 to 4, which correspond to projected flexural, longitudinal, and torsional modes, respectively. Since the longitudinal mode is the one used for imaging applications, one can consider that it is partly isolated when located between band gaps *B* and *D*. Furthermore, the torsional mode is not piezoelectrically coupled, and coupling of the longitudinal mode can occur only with flexural modes 1/2 when considering the normalized-frequency range between 750 and 1300 Hz m. By locating the longitudinal mode between frequency gaps, one can choose an operation point and fix a value for γ_3, namely a value for the ratio d_1/λ_3 where λ_3 is the wavelength along x_3, and finally for the pitch-over-thickness ratio of the composite.

From this approach, it can be seen that it is possible to perform a systematic analysis considering materials and geometries of composites in order to obtain configurations that optimize band gaps in which the longitudinal mode has to be located.

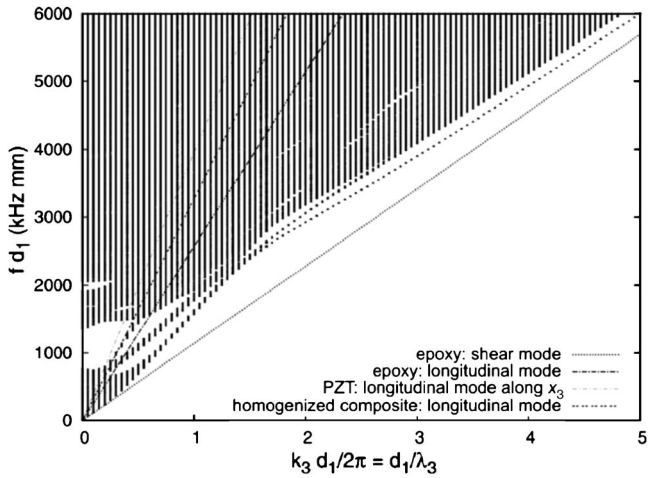


FIG. 5. Out-of-plane dispersion curves for all values of components k_1 and k_2 of the wave vector. White areas correspond to couples (f,k_3) for which no wave propagates.

B. Modal environment of the longitudinal mode

In order to precise the "modal" environment of the thickness mode when fixing an operation frequency (corresponding here to the antiresonance), we plot out-of-plane dispersion curves for fixed values of $\boldsymbol{k}_{\parallel}=(k_1,k_2)^T$, in particular for $\boldsymbol{k}_{\parallel}=\boldsymbol{0}$ (point Γ), and for $(\gamma_1=0.25,\gamma_2=0)$ which corresponds to waves propagating along x_1 (midpoint on the path $\Gamma-X$). Results are shown in Fig. 7.

By computing the generalized displacement according to Eq. (5), mode shapes allow one to identify potentially coupled modes, meaning modes that can be excited regarding the symmetry of the structure and of the considered excitation (this latter related to parameters γ_1 and γ_2). In particular, the so-called lateral modes, due to Bragg diffraction, are investigated. Most of the proposed developments to investigate lateral modes by the use of PWE analysis considered in-plane propagation (see, for instance, Ref. 6). Nevertheless, lateral modes in piezocomposites have a normal component in terms of wave vector.

Figure 7(a) corresponds to the usual case of a synchronous vibration (excitation) of the whole composite, for which lateral modes have been investigated—one has to note that subelements inside an electric pixel of a phased array, for instance, also vibrate synchronously and are liable to exhibit lateral modes. Potentially coupled modes are plotted in solid lines. For γ_3 lower than 0.5, the first coupled mode is the longitudinal one, whereas the other modes correspond to the numerous lateral modes. Considering the longitudinal mode and the first lateral mode around $\gamma_3=0.6$, one first observes a coupling between them when the branches are close, and then an inversion in vibration, in other words the branch of the longitudinal mode becomes that of the first lateral mode beyond 0.6. Rigorously, the one of the first lateral mode should become that of the longitudinal mode, but this kind of phenomenon also occurs between the numerous lateral modes successively, so that pure longitudinal mode does not exist anymore for γ_3 greater than 0.5.

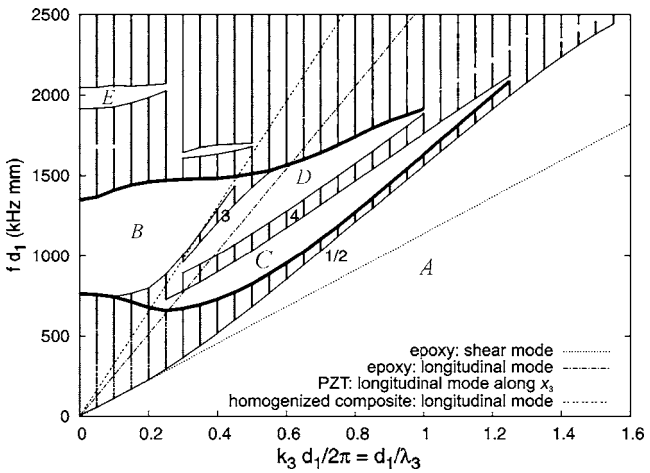


FIG. 6. Close-up of Fig. 5.

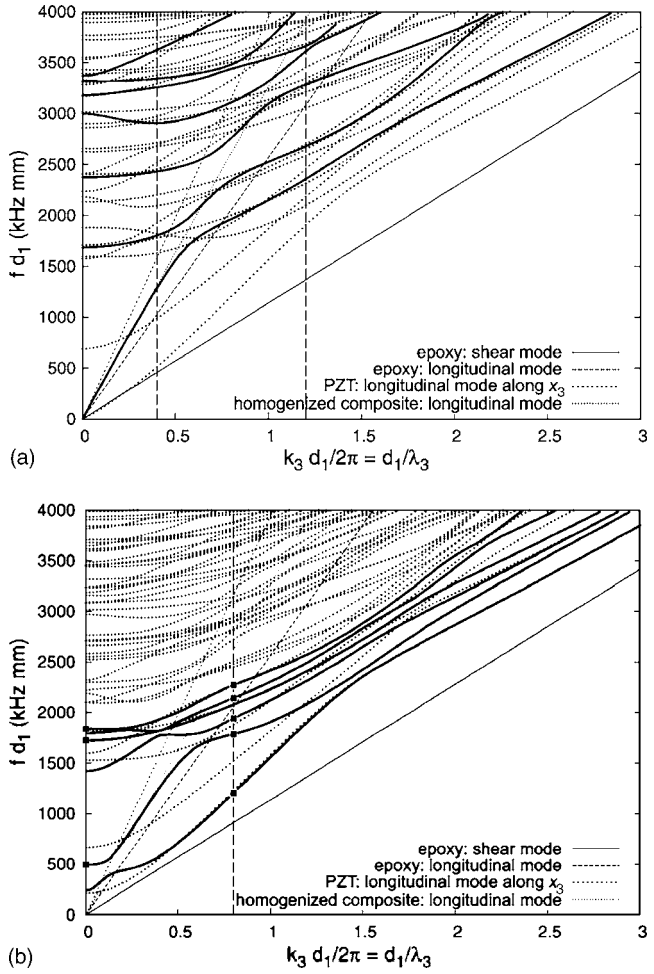


FIG. 7. Out-of-plane dispersion curves for two points on the Γ-*X* path: (a) point Γ, and (b) $\gamma_1=0.25$. Solid lines correspond to potentially coupled modes for (a). In Fig. (b), squares correspond to potentially coupled modes for γ_3 equal to 0 and 0.8, and solid lines are their respective branches.

Let us now assume that the operation point is $\gamma_3=0.4$, then fixing the anti-resonance frequency and locating the longitudinal mode between band gaps. The arbitrary chosen period of 100 μm, corresponds to a wavelength λ_3 of 250 μm, namely a thickness of the composite equal to 125 μm.

For this chosen value, Fig. 8 gives the displacement fields of modes when the composite vibrates synchronously, including the fundamental longitudinal mode, the first, second, and higher-order lateral modes. First and second lateral modes arise for $\gamma_1=\gamma_2=1$ and for $\gamma_1=\gamma_2=2$, respectively, but are located at point Γ when representing dispersion curves in the first Brillouin zone.[14] They are due to the appearance of band gaps considering the SV mode, for $\gamma_3=0$, or the L mode, for $\gamma_3=0.4$, for which the polarization is principally along x_3. At point Γ, the wave-vector components k_1 and k_2 are equal to $2\pi n/d_1$ where n is an integer. It means that the wavelengths λ_1 and λ_2 are equal to d_1/n, depending on the considered mode. When n is not zero, the wavelength becomes equal or a fraction of the period, resulting in the so-called lateral modes due to Bragg diffraction in the periodic structure analogous to a crystal. Considering, for instance, a wavelength equal to the period, vibration nodes can

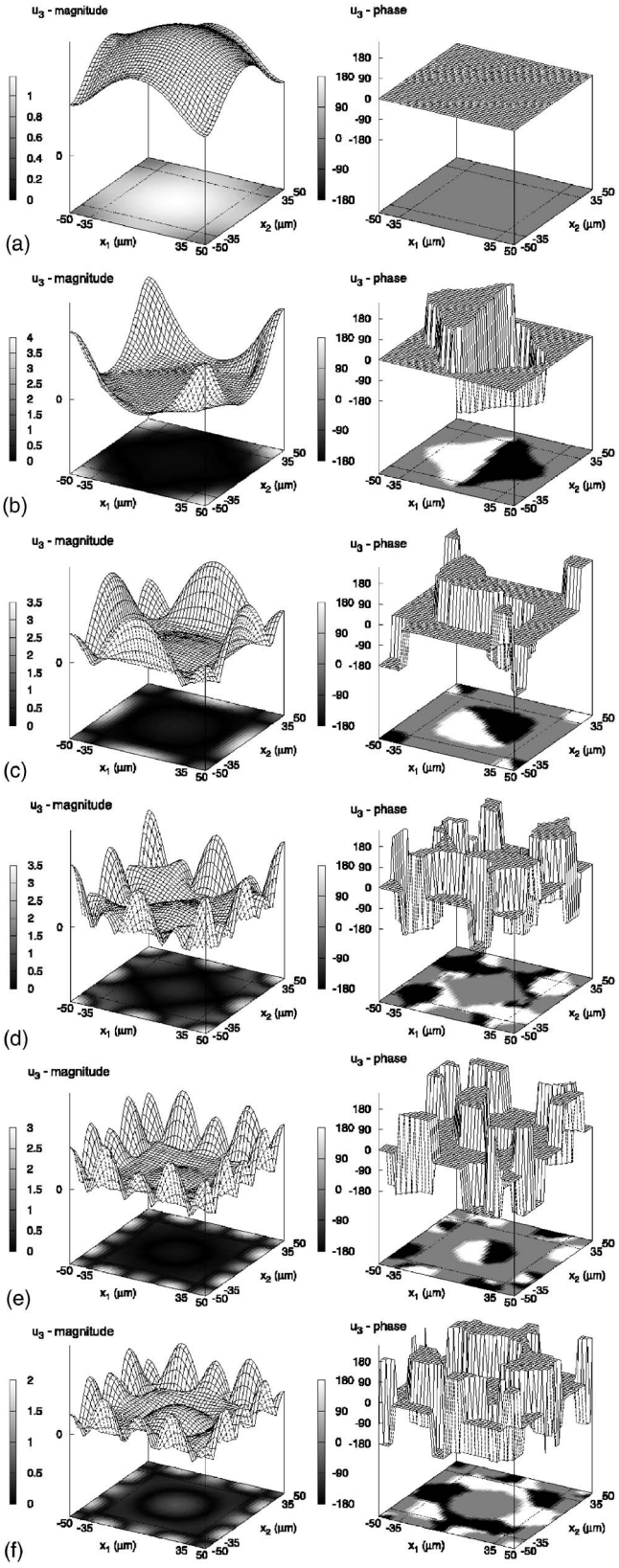


FIG. 8. Magnitude and phase of the normal displacement u_3 of an elementary cell when all cells vibrate synchronously ($\gamma_1=\gamma_2=0$) with a normal component $\gamma_3=0.4$. Reported modes correspond to potentially coupled modes in Fig. 7(a) in increasing-frequency order. Mode (a) is the fundamental longitudinal mode, whereas modes (b) and (c) are the first and second lateral modes, respectively.

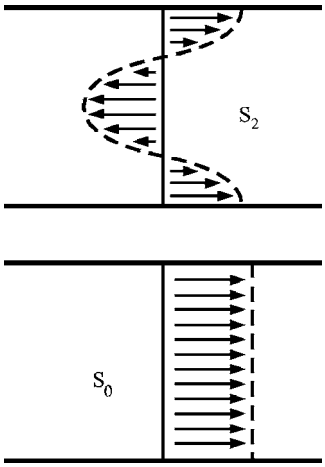


FIG. 9. Mode shape for the S_0 and S_2 Lamb waves.

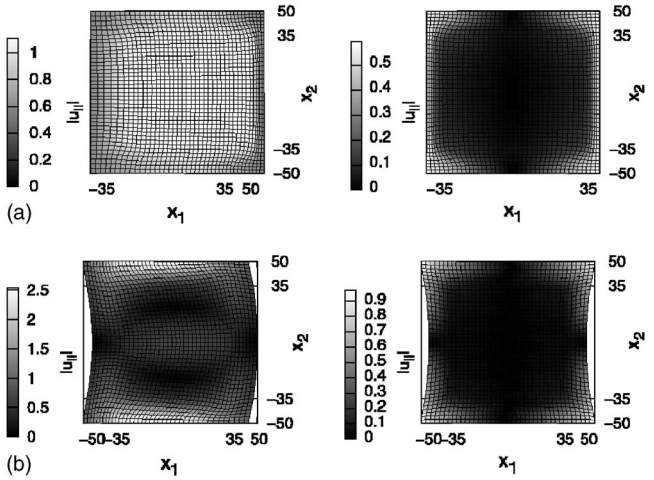


FIG. 10. In-plane displacement field $\boldsymbol{u}_{\parallel}$ of potentially coupled modes for $\gamma_3=0$ (in-plane propagation) at the middle point of path $\Gamma-X$ ($\gamma_1=0.25$ et $\gamma_2=0$). Two neighboring cells vibrate in quadrature so that real and imaginary parts of the displacement correspond to the vibration of two neighboring cells at a given time. Mode (a) is the longitudinal mode along x_1 and yields the S_0 Lamb-like wave for a plate.

be located at the center and the edges of the elementary cell, or alternatively at $\lambda/4$ compared to the previous position, resulting in two different vibration frequencies for the same wavelength, so that a frequency band gap is created. In Fig. 8(b), only the second case is represented, corresponding to the vibration compatible with the symmetry of the composite, when considering the electric excitation used in operation. Higher order modes in Fig. 8 correspond to higher values of n, and to more complex vibration modes due to multiple reflections of waves between interfaces.

If one considers dispersion curves for $\gamma_1=0.25$ in Fig. 7(b), the two first solid lines, corresponding to SV and L modes, respectively, are close together around $\gamma_3=0.15$, and the L mode for $\gamma_3>0.15$ is located in the continuation of the SV mode for $\gamma_3<0.15$. It is due to a transition from SV mode to L mode, both polarized along x_3, when γ_3 increases.

For a given thickness, the wavelength of fundamental modes is $\lambda_3^{(1)}=2h$, but such plate composites can also vibrate with wavelengths of higher order such as $\lambda_3^{(3)}=2h/3$, namely $\gamma_3=1.2$ in our case, as indicated in Fig. 7(a). Practically, at frequencies higher than that of the pure thickness mode, one obtains a superimposition of fundamental lateral modes with their harmonics, which wavelength $\lambda_3^{(3)}$ is a third of that of the fundamental modes, but which frequencies are not threefold in comparison with the fundamental ones. In particular, that means that, when mapping the normal displacement fields of a composite, for instance, with an interferometric laser probe, one cannot know if the observed lateral mode is the fundamental or the harmonic.

Finally, waves can propagate along the plane of the composite for other values of $\boldsymbol{k}$, such as Lamb-like waves S_0 or S_2 according to the definition given in Ref. 19. In that case, the normal wavelength is infinite ($k_3=0$), or equal to the thickness, as can be seen in Fig. 9, meaning that γ_3 is 0 or 0.8 for our specific configuration. The corresponding coupled modes are reported in Fig. 7(b) by means of square-shaped dots for $(\gamma_1=0.25, \gamma_2=0)$. The first indicated bulk modes at 0 and 0.8 yield S_0 and S_2 Lamb waves, respectively, also corresponding to modes labeled (L) and (F1) in Figs. 3(a) and 3(h), respectively. The S_2 frequency is close to the thickness-mode one, as it has been observed experimentally and theoretically—by a finite-element method—in Ref. 9. The transverse displacement along the thickness, considering the S_2-like mode, is due to flexural vibrations of PZT inclusions. Other modes, located at higher frequencies, have similar polarizations, but a wavelength λ_2 along x_2 equal to one period or even smaller, exhibiting Bragg-diffraction phenomena similarly to lateral modes. Displacement fields are given in Fig.10 for the two first modes at $\gamma_3=0$.

Because waves in plate composites practically result from multiple reflexions of bulk modes at surfaces of the plate, this analysis is qualitative and not quantitative, since mechanical and electric boundary conditions at surfaces are not taken into account, resulting in a frequency shift for some modes—especially Lamb-like modes—between the infinite-thickness case and the plate case. Nevertheless, it provides an efficient tool to investigate the width of band gaps depending on geometries and materials, and the capacity of a composite to vibrate without parasitic coupling within a wide frequency range, which is an essential criterion for imaging applications and, for instance, harmonic imaging. Furthermore, the branch of the longitudinal (thickness) mode is similar to that of the corresponding plate at the anti-resonance. Since the inflexion of the dispersion curve gives an indication regarding propagation of energy along the plate, one also obtains information about cross talks due to the thickness mode itself. This analysis is available with composites consisting in only a few cells. Subelements of only two periods already exhibit lateral modes, and Lamb-like modes indicate the frequency range of radial modes of real finite-width composite plates, even consisting of less than ten cells in one direction of periodicity. Once this analysis is achieved, yielding choices of geometries and materials, a more quantitative analysis can be performed by means of other methods such as finite-element analysis[9] coupled to a boundary-element method.[10]

V. CONCLUSION

We have presented a numerical analysis of periodic piezoelectric composites based on the plane-wave-expansion method, for imaging-application purposes. This approach consists in considering composites as phononic crystals, and hence in computing the dispersion curves of waves propagating in the composite whatever the value of the wave-vector component normal to the plane of structure. Thanks to its nonprohibitive computation time, band gaps can systematically be investigated as a function of the filling fraction of the active material, of the geometry of the inclusions, and of the properties of the materials themselves. Locating the thickness mode in band gaps indeed allows one to minimize coupling with parasitic modes such as lateral modes in electric pixels or in mono-element transducers, or even propagation of Lamb waves in phased array or excitation of plate modes.

The presented approach has been applied to the case of a representative 1-3 connectivity piezocomposite with square-section PZT rods in an epoxy matrix. Bulk waves, that *in fine* result into finite-thickness composite modes, have all been identified, especially waves that yield either lateral modes or Lamb-like waves.

Due to the generic nature of this approach and to its efficiency, it is complementary to other approaches such as finite-element methods and it can be used as a first level of conception prior to more complicated and more specific computations. Any size, shape, and arrangement of inclusions can be considered, within the limitations set by machining capabilities.

[1]T. R. Gururaja, W. A. Schulze, L. E. Cross, R. E. Newnham, B. A. Auld, and Y. J. Wang, "Piezo-electric composite materials for ultrasonic transducer applications. Part i: resonant modes of vibration of PZT rod-polymer composites," IEEE Trans. Sonics Ultrason. **32**, 481–498 (1985).

[2]W. A. Smith, "The role of piezocomposites in ultrasonic transducers," in *Proc. of the IEEE Ultrasonics Symposium*, Montreal, Quebec, Canada, 755–766 (1989).

[3]W. Smith and B. Auld, "Modeling 1-3 composite piezoelectrics: Thickness-mode oscillations," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **38**, 40–47 (1991).

[4]T. Ritter, X. Geng, K. K. Shung, P. D. Lopath, S.-E. Park, and T. R. Shrout, "Single crystal PZN/PT-polymer composites for ultrasound transducer applications," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 792–800 (2000).

[5]B. A. Auld, H. A. Kunkel, Y. A. Shui, and Y. Wang, "Dynamic behavior of periodic piezoelectric composites," in *Proc. of the IEEE Ultrasonics Symposium*, Atlanta, Georgia, 554–558 (1983).

[6]D. Certon, F. Patat, F. Levassort, G. Feuillard, and B. Karlsson, "Lateral resonances in 1-3 piezoelectric periodic composite: Modeling and experimental results," J. Acoust. Soc. Am. **101**, 2043–2051 (1997).

[7]D. Certon, O. Casula, F. Patat, and D. Royer, "Theoretical and experimental investigations of lateral modes in 1-3 piezocomposites," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 643–651 (1997).

[8]M. Wilm, S. Ballandras, V. Laude, and T. Pastureaud, "A full 3D plane-wave-expansion model for 1-3 piezoelectric composite structures," J. Acoust. Soc. Am. **112**, 943–952 (2002).

[9]M. Wilm, R. Armati, W. Daniau, and S. Ballandras, "Cross-talk phenomena in a 1-3 connectivity piezoelectric composite," J. Acoust. Soc. Am. **116**, 2948–2955 (2004).

[10]M. Wilm, A. Reinhardt, V. Laude, R. Armati, W. Daniau, and S. Ballandras, "Three-dimensional modeling of micromachined-ultrasonic-transducer arrays operating in water," Ultrasonics **43**, 457–465 (2005).

[11]M. M. Sigalas and E. N. Economou, "Band structure of elastic waves in two dimensional systems," Solid State Commun. **86**, 141 (1993).

[12]M. S. Kushwaha, P. Halevi, L. Dobrzynski, and B. Djafari-Rouhani, "Acoustic band structure of periodic elastic composites," Phys. Rev. Lett. **71**, 2022 (1993).

[13]M. Wilm, A. Khelif, S. Ballandras, V. Laude, and B. Djafari-Rouhani, "Out-of-plane propagation of elastic waves in two-dimensional phononic band-gap materials," Phys. Rev. E **67**, 065602 (2003).

[14]L. Brillouin, *Wave Propagation in Periodic Structures* (Dover, New York, 1953).

[15]J. Vasseur, B. Djafari-Rouhani, L. Dobrzynski, M. Kushwaha, and P. Halevi, "Complete acoustic band gaps in periodic fiber reinforced composite materials: The carbon/epoxy composite and some metallic systems," J. Phys.: Condens. Matter **6**, 8759–8770 (1994).

[16]A. Fahmy and E. Adler, "Propagation of surface acoustic waves in multilayers: A matrix description," Appl. Phys. Lett. **22**, 495–497 (1973).

[17]E. L. Adler, "Matrix methods applied to acoustic waves in multilayers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**, 485–490 (1990).

[18]R. Peach, "A general Green function analysis for SAW devices," in *Proc. of the IEEE Ultrasonic Symposium*, Seattle, Washington, 221–225 (1995).

[19]D. Royer and E. Dieulesaint, *Elastic Waves in Solids 1* (Springer-Verlag, Berlin, 2000).

Dynamic response of an insonified sonar window interacting with a Tonpilz transducer array

Andrew J. Hull[a)]
Autonomous Systems and Technology Department, Naval Undersea Warfare Center Division, Newport, Rhode Island 02841

(Received 7 March 2007; revised 24 May 2007; accepted 24 May 2007)

This paper derives and evaluates an analytical model of an insonified sonar window in contact with an array of Tonpilz transducers operating in receive mode. The window is fully elastic so that all wave components are present in the analysis. The output of the model is a transfer function of a transducer element output voltage divided by input pressure versus arrival angle and frequency. This model is intended for analysis of sonar systems that are to be built or modified for broadband processing. The model is validated at low frequency with a comparison to a previously derived thin plate model. Once this is done, an example problem is studied so that the effects of higher order wave interaction with acoustic reception can be understood. It was found that these higher order waves cause multiple nulls in the region where the array detects acoustic energy and that their locations in the arrival angle-frequency plane can be determined. The effects of these nulls in the beam patterns of the array are demonstrated.
 [DOI: 10.1121/1.2749702]

PACS number(s): 43.38.Hz, 43.20.Tb, 43.40.Dx [DF] Pages: 794–803

I. INTRODUCTION

One very typical sonar design consists of a soft material that is in contact with an array of Tonpilz transducers. This soft material is usually called a window (or sometimes a screen). The window serves several purposes: it protects the transducers from water, impact, and debris; it attenuates nonacoustic energy; it provides a safe covering to the sonar system when it is shipped or handled; and it helps to minimize turbulence and hydrodynamic drag. Tonpilz transducers are a reliable transducer design that has been refined for many years. Single resonant Tonpilz transducers typically consist of a head and tail mass separated by piezoelectric stack that emits a voltage when it is subjected to an applied force. Tonpilz transducers are a useful design because they can operate in transmit (active) and receive (passive) mode. A typical Tonpilz array will transmit energy into the water, wait a short period of time, and then receive signals back based on the echo of an object(s) in the water. Because they can operate in both modes, they are frequently used in situations where higher signal-to-noise ratio is needed than can be obtained from passive sonar only. This paper is specifically interested in understanding their behavior in the receive mode.

Tonpilz transducers have been studied in the literature for many years. Basic design guidelines exist in textbooks.[1] General transducer modeling techniques have been previously developed using a number of numerical methods.[2] A finite element model of the transducer that yields the transmit voltage response of a 2-2 mode piezocomposite when it is in contact with a heavy fluid has been investigated.[3] Another finite element model studied the elastic response of the head and tail mass of a Tonpilz transducer in contact with air.[4] Equivalent circuit models have been derived and analyzed where they model a Tonpilz transducer for use as an underwater horn,[5] compare Tonpilz response to a flexural disk, Helmholtz resonator, moving coil, and a dual array of piezoelectric disks and squares,[6] and a Mason equivalent circuit representation has been optimized to provide a model of broadband and high-power response simultaneously.[7]

Sonar window models have been previously studied. Insertion loss and echo reduction measurements are common in the literature where a general goal is to have no insertion loss in the frequency and wave numbers of interest. In one study, glass microspheres and phenolic plastic microspheres were added to fluorinated epoxies to approach this design goal.[8] A model has been derived and experimentally verified to predict transmission and reflection coefficients for n layers of plane parallel plates.[9] Acoustic measurements of numerous single materials is also present in the literature.[10] The dynamic analysis of multilayer composite plates with an emphasis on wave dispersion has been documented.[11] Acoustic transmission of energy in sandwich construction has been analytically studied and experimentally verified.[12] Structurally stiffening the window for hull applications has also been investigated.[13] The first set of references[1–7] are transducer modeling papers and the second set of references[8–13] are window modeling and testing. There is no analytical model that couples Tonpilz transducers to a fully elastic sonar window.

This paper derives and evaluates an analytical model of a fully elastic sonar window in contact with an array of Tonpilz transducers. The window is insonified by a plane wave at varying arrival angles and frequencies. This model is intended for broadband frequency analysis of a sonar system when there is significant interaction between the window and the array of transducers. The formulation of the problem begins with elasticity theory, which models the motion in the

a)Electronic mail: hullaj@npt.nuwc.navy.mil

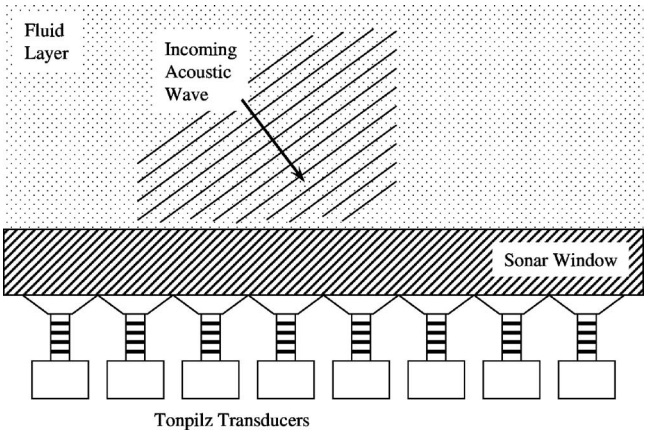


FIG. 1. Sonar window with Tonpilz transducers.

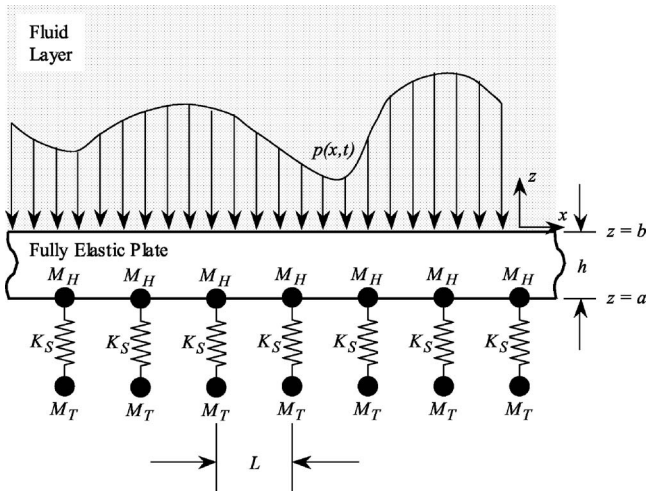


FIG. 2. Model of sonar window with Tonpilz transducers.

window as a combination of dilatational and shear waves. From this theory, expressions for plate displacements in the normal and tangential direction are obtained. The displacements are then inserted into stress relationships and these equations are set equal to the forces acting on the structure by the transducers and the pressure of the incoming acoustic wave. The problem is then written as an algebraic system of equations, in matrix from, where the left-hand terms represent the zero-order window dynamics and are equal to an infinite number of right-hand terms that represent the forces acting on the structure. Rewriting this zero-order dynamic term, by increasing and decreasing the index, results in an expression for the higher-order modes interacting with the applied forces. The integer shift property is then applied to the right-hand side of all of the terms, resulting in an infinite set of equations that model the wave propagation coefficients of all the modes of the structure. This set of equations is truncated to a finite number of terms, and solutions to the displacement fields are calculated. The transducer output is written as a function of the displacement field at the bottom of the window, and this term is calculated as a transfer function of voltage divided by applied pressure versus arrival angle and frequency. A numerical example is included where the array beam patterns are generated and the results are discussed.

II. SYSTEM MODEL

The system model is that of a sonar window attached to an array of Tonpilz transducers, as shown in Fig. 1. This mechanical problem is analytically modeled by assuming the sonar window is a fully elastic plate and the Tonpilz transducers are discrete mass-spring-mass systems, as shown in Fig. 2. The plate (or sonar window) has a thickness of h (m) and is loaded on the top surface with a normal (pressure) forcing function. The transducers on the bottom of the window are equally spaced at a distance of L (m) in the x direction and each has a head mass per unit length M_H (kg/m), tail mass per unit length M_T (kg/m), and stiffness per unit length K_S (N/m^2). The model uses the following assumptions: (1) the forcing function acting on the plate is a plane wave at a definite wave number and frequency; (2) motion is normal and tangential to the plate in one direction (two-dimensional system); (3) the plate has infinite spatial extent in the x direction; (4) the head mass has translational degrees of freedom in the x and z directions; (5) the tail mass and the spring have translational degrees of freedom in the z direction; (6) the particle motion is linear; and (7) the fluid medium is lossless.

The motion of the elastic plate is governed by

$$\mu\nabla^2\mathbf{u}(x,y,z,t) + (\lambda+\mu)\,\nabla\,\nabla\cdot\mathbf{u}(x,y,z,t) = \rho\frac{\partial^2\mathbf{u}(x,y,z,t)}{\partial t^2}, \tag{1}$$

where ρ is the density (kg/m^3), λ and μ are the complex Lamé constants (N/m^2), t is time (s), $\cdot$ denotes a vector dot product, and $\mathbf{u}(x,y,z,t)$ is the three-dimensional Cartesian coordinate displacement vector and is written as

$$\mathbf{u}(x,y,z,t) = \begin{Bmatrix} u(x,y,z,t) \\ v(x,y,z,t) \\ w(x,y,z,t) \end{Bmatrix} = \nabla\phi(x,y,z,t) + \nabla \times \begin{Bmatrix} \psi_x(x,y,z,t) \\ \psi_y(x,y,z,t) \\ \psi_z(x,y,z,t) \end{Bmatrix}, \tag{2}$$

where ϕ is a dilatational scalar potential, ∇ is the gradient operator, $\times$ denotes a vector cross product, and $\vec{\psi}$ is an equivoluminal vector potential. The formulation is now condensed into a two-dimensional problem; thus, $\nu \equiv 0$ and $\partial(\cdot)/\partial y \equiv 0$, where $(\cdot)$ denotes any function. Expanding Eq. (2) and breaking the displacement vector into its individual nonzero terms yields

$$u(x,z,t) = \frac{\partial\phi(x,z,t)}{\partial x} - \frac{\partial\psi_y(x,z,t)}{\partial z} \tag{3}$$

and

$$w(x,z,t) = \frac{\partial\phi(x,z,t)}{\partial z} + \frac{\partial\psi_y(x,z,t)}{\partial x}. \tag{4}$$

Equations (3) and (4) are next inserted into Eq. (1), which results in two decoupled wave equations given by

$$c_d^2 \nabla^2 \phi(x,z,t) = \frac{\partial^2 \phi(x,z,t)}{\partial t^2} \tag{5}$$

and

$$c_s^2 \nabla^2 \psi_y(x,z,t) = \frac{\partial^2 \psi_y(x,z,t)}{\partial t^2}, \tag{6}$$

where Eq. (5) corresponds to the dilatational component and Eq. (6) corresponds to the shear component of the displacement field. Correspondingly, the constants c_d and c_s are the complex dilatational and shear wave speeds (m/s), respectively, and are determined by

$$c_d = \sqrt{\frac{\lambda + 2\mu}{\rho}} \tag{7}$$

and

$$c_s = \sqrt{\frac{\mu}{\rho}}. \tag{8}$$

The equations of motion are formulated as a boundary value problem using four equations of stress written in terms of the plates' displacements and corresponding forcing functions. The plate is loaded by a normal (pressure) forcing function as shown in Fig. 2; thus, the normal stress at $z=b$ is written using a force balance between the pressure in the fluid and the plate as

$$\tau_{zz}(x,b,t) = (\lambda + 2\mu)\frac{\partial w(x,b,t)}{\partial z} + \lambda \frac{\partial u(x,b,t)}{\partial x} = -p(x,b,t), \tag{9}$$

where $p(x,b,t)$ is the pressure field in contact with the top of the plate ($\mathrm{N/m^2}$). The tangential stress on the top of the plate is modeled as a free boundary condition and is written as

$$\tau_{zx}(x,b,t) = \mu\left[\frac{\partial u(x,b,t)}{\partial z} + \frac{\partial w(x,b,t)}{\partial x}\right] = 0. \tag{10}$$

The plate is loaded by the forces in the Tonpilz transducers' head masses acting on the bottom of the plate, thus, the normal stress at $z=a$ is

$$\tau_{zz}(x,a,t) = (\lambda + 2\mu)\frac{\partial w(x,a,t)}{\partial z} + \lambda \frac{\partial u(x,a,t)}{\partial x} = \sum_{n=-\infty}^{n=+\infty} f_z(a,t)\,\delta(x - nL), \tag{11}$$

where $f_z(a,t)$ is the force per unit length that each Tonpilz transducer exerts on the plate in the z direction and $\delta(x - nL)$ is the Dirac delta function that distributes the transducer forces discretely and periodically. Similarly, the tangential stress on the bottom of the plate is

$$\tau_{zx}(x,a,t) = \mu\left[\frac{\partial u(x,a,t)}{\partial z} + \frac{\partial w(x,a,t)}{\partial x}\right] = \sum_{n=-\infty}^{n=+\infty} f_x(a,t)\,\delta(x - nL), \tag{12}$$

where $f_x(a,t)$ is the force per unit length that each Tonpilz transducer exerts on the plate in the x direction.

The acoustic pressure in the fluid medium is governed by the two-dimensional wave equation and is written in Cartesian coordinates as

$$\frac{\partial^2 p(x,z,t)}{\partial z^2} + \frac{\partial^2 p(x,z,t)}{\partial x^2} - \frac{1}{c_f^2}\frac{\partial^2 p(x,z,t)}{\partial t^2} = 0, \tag{13}$$

where $p(x,z,t)$ is the pressure ($\mathrm{N/m^2}$) and c_f is the real valued compressional wave speed in the fluid (m/s). The interface between the fluid and solid surface at $z=b$ satisfies the linear momentum equation, which relates the acceleration of the plate surface to the spatial gradient of the pressure field by

$$\rho_f \frac{\partial^2 w(x,b,t)}{\partial t^2} = -\frac{\partial p(x,b,t)}{\partial z}, \tag{14}$$

where ρ_f is the density of the fluid ($\mathrm{kg/m^3}$).

The system output is a transfer function of a transducer voltage divided by incident pressure, commonly called receive voltage sensitivity (rvs). This is equal to

$$\mathrm{rvs}(x,t) = g_{33} t f_s(x,t), \tag{15}$$

where g_{33} is the material constant that relates the mechanical force to electrical output (volts m/N), t is the thickness of each piezoelectric piece from the transducer stack, and $f_s(x,t)$ is the force in the transducer, which, in the case of this model, is the force across the spring.

III. ANALYTICAL SOUTION

The displacements are now written in the spatial-frequency domain by using the functional form where the field variables are equal to a sum of unknown functions in the z direction multiplied by spatially indexed harmonic exponential functions in the x direction multiplied by an exponential function in time. The displacements become

$$u(x,z,t) = \sum_{m=-\infty}^{m=+\infty} U_m(z)\exp(\mathrm{i}k_m x)\exp(-\mathrm{i}\omega t) \tag{16}$$

and

$$w(x,z,t) = \sum_{m=-\infty}^{m=+\infty} W_m(z)\exp(\mathrm{i}k_m x)\exp(-\mathrm{i}\omega t), \tag{17}$$

where $\mathrm{i} = \sqrt{-1}$, ω is frequency (rad/s), and

$$k_m = k + \frac{2\pi m}{L}, \tag{18}$$

where k is the wave number with respect to the x axis (rad/m), and it is noted that $k_0 \equiv k$.

Inserting Eqs. (16) and (17) into Eqs. (2)–(6) and solving the differential equations gives the unknown displacement field term $U_m(z)$ as

$$U_m(z) = A_m(k,\omega)\mathrm{i}k_m \exp(\mathrm{i}\alpha_m z) + B_m(k,\omega)\mathrm{i}k_m \times \exp(-\mathrm{i}\alpha_m z) - C_m(k,\omega)\mathrm{i}\beta_m \exp(\mathrm{i}\beta_m z) + D_m(k,\omega)\mathrm{i}\beta_m \exp(-\mathrm{i}\beta_m z) \tag{19}$$

and the unknown displacement field term $W_m(z)$ as

$$W_m(z) = A_m(k,\omega)\mathrm{i}\alpha_m \exp(\mathrm{i}\alpha_m z) - B_m(k,\omega)\mathrm{i}\alpha_m \times\exp(-\mathrm{i}\alpha_m z) + C_m(k,\omega)\mathrm{i}k_m \exp(\mathrm{i}\beta_m z) + D_m(k,\omega)\mathrm{i}k_m \exp(-\mathrm{i}\beta_m z), \quad (20)$$

where $A_m(k,\omega)$, $B_m(k,\omega)$, $C_m(k,\omega)$, and $D_m(k,\omega)$ are unknown complex wave propagation coefficients of the plate, α_m is the modified wave number (rad/m) associated with the dilatational wave and is expressed as

$$\alpha_m = \sqrt{k_d^2 - k_m^2}, \quad (21)$$

where k_d is the dilatational wave number and is equal to ω/c_d; β_m is the modified wave number (rad/m) associated with the shear wave and is expressed as

$$\beta_m = \sqrt{k_s^2 - k_m^2}, \quad (22)$$

where k_s is the shear wave number (rad/m) and is equal to ω/c_s.

The pressure field consists of two terms: an outgoing pressure field caused by the plate displacement and an incoming applied incident pressure field (the forcing function) acting on the structure. A convenient way to express this field, after solving Eq. (13), is

$$p(x,z,t) = \sum_{m=-\infty}^{m=+\infty} P_m(z)\exp(\mathrm{i}k_m x)\exp(-\mathrm{i}\omega t), \quad (23)$$

where

$$P_m(z) = M_m(k,\omega)\exp(\mathrm{i}\gamma_m z) + \delta_{m0}P_I(\omega)\exp(-\mathrm{i}\gamma_m z), \quad (24)$$

where δ_{m0} is the Kronecker delta function and $P_I(\omega)$ is the magnitude of the applied pressure $\mathrm{N/m^2}$. In Eq. (24), γ_m is the modified wave number (rad/m) associated with the fluid and is expressed as

$$\gamma_m = \sqrt{(\omega/c_f)^2 - k_m^2} = \sqrt{k_f^2 - k_m^2}, \quad (25)$$

where γ_m is purely real or imaginary, depending on the sign of the argument under the radical. When $m=0$ and the sign of the argument is positive, the analysis is in the acoustic region; when $m=0$ and the sign of the argument is negative, the analysis is in the nonacoustic region. This acoustic region is frequently called the acoustic cone as it is the interior of a V shape on a wave-number-frequency plot. For acoustic sonar response, the analysis is typically studied in the acoustic cone. The relationship between the arrival angle of an acoustic wave and its wave number is

$$k = (\omega/c_f)\sin(\theta), \quad (26)$$

where θ is the arrival angle of an incoming acoustic wave (rad) with respect to the x axis and a value of 0 corresponds to broadside excitation.

The forces exerted by the Tonpilz transducers can be determined with a dynamical model of a mass-spring-mass system. The force per unit length in the z direction for each transducer is

$$f_z(a,t) = \left[\frac{\omega^4 M_H M_T - \omega^2 K_S(M_T + M_H)}{K_S - \omega^2 M_T}\right] w(x,a,t) = F_z(\omega)w(x,a,t), \quad (27)$$

and the force per unit length in the x direction for each transducer is

$$f_x(a,t) = -\omega^2 M_H u(x,a,t) = F_x(\omega)u(x,a,t), \quad (28)$$

as the transducer spring constant is zero in the horizontal direction. It is noted that if the transducer is a double or triple resonant type, the expressions given in Eqs. (27) and (28) can be changed to reflect these dynamic effects.

Finally, the output of the transducer can be found by substituting the spring dynamics into Eq. (15). This yields the receive voltage sensitivity in terms of the displacement at the bottom of the sonar window as

$$\mathrm{rvs}(x,t) = \left(\frac{\omega^2 M_T}{K_S - \omega^2 M_T}\right) g_{33} t(K_S/d) w(x,a,t), \quad (29)$$

where d is the width of the transducer head in the x direction (m).

The four boundary value equations [Eqs. (9)–(12)] are now rewritten using Eq. (14) with the displacements and pressure terms inserted into their respective variables. They become

$$(\lambda + 2\mu)\sum_{m=-\infty}^{m=+\infty} \frac{\partial W_m(b)}{\partial z}\exp(\mathrm{i}k_m x) + \mathrm{i}k_m \lambda \sum_{m=-\infty}^{m=+\infty} U_m(b)\exp(\mathrm{i}k_m x) + \left(\frac{\omega^2\rho_f}{\mathrm{i}\gamma_m}\right)\sum_{m=-\infty}^{m=+\infty} W_m(b)\exp(\mathrm{i}k_m x) = -2P_I(\omega)\exp(\mathrm{i}kx), \quad (30)$$

$$\mu\left[\sum_{m=-\infty}^{m=+\infty} \frac{\partial U_m(b)}{\partial z}\exp(\mathrm{i}k_m x) + \mathrm{i}k_m \sum_{m=-\infty}^{m=+\infty} W_m(b)\exp(\mathrm{i}k_m x)\right] = 0, \quad (31)$$

$$(\lambda + 2\mu)\sum_{m=-\infty}^{m=+\infty} \frac{\partial W_m(a)}{\partial z}\exp(\mathrm{i}k_m x) + \mathrm{i}k_m \lambda \sum_{m=-\infty}^{m=+\infty} U_m(a)\exp(\mathrm{i}k_m x) = F_z(\omega)\sum_{n=-\infty}^{n=+\infty}\left[\sum_{m=-\infty}^{m=+\infty} W_m(a)\exp(\mathrm{i}k_m x)\right]\delta(x - nL), \quad (32)$$

and

$$\mu\left[\sum_{m=-\infty}^{m=+\infty}\frac{\partial U_m(a)}{\partial z}\exp(\mathrm{i}k_m x)+\mathrm{i}k_m\sum_{m=-\infty}^{m=+\infty}W_m(a)\exp(\mathrm{i}k_m x)\right]$$
$$=F_x(\omega)\sum_{n=-\infty}^{n=+\infty}\left[\sum_{m=-\infty}^{m=+\infty}U_m(a)\exp(\mathrm{i}k_m x)\right]\delta(x-nL). \quad (33)$$

The Dirac delta comb function that is present in Eqs. (32) and (33) obeys the relationship

$$\sum_{n=-\infty}^{n=+\infty}\delta(x-nL)=\frac{1}{L}\sum_{n=-\infty}^{n=+\infty}\exp(\mathrm{i}2\pi nx/L), \quad (34)$$

and using this equation, Eqs. (33) and (34) become

$$(\lambda+2\mu)\sum_{m=-\infty}^{m=+\infty}\frac{\partial W_m(a)}{\partial z}\exp(\mathrm{i}k_m x)$$
$$+\mathrm{i}k_m\lambda\sum_{m=-\infty}^{m=+\infty}U_m(a)\exp(\mathrm{i}k_m x)$$
$$=\frac{F_z(\omega)}{L}\sum_{n=-\infty}^{n=+\infty}\left[\sum_{m=-\infty}^{m=+\infty}W_m(a)\exp(\mathrm{i}k_m x)\right]\exp(\mathrm{i}2\pi nx/L) \quad (35)$$

and

$$\mu\left[\sum_{m=-\infty}^{m=+\infty}\frac{\partial U_m(a)}{\partial z}\exp(\mathrm{i}k_m x)+\mathrm{i}k_m\sum_{m=-\infty}^{m=+\infty}W_m(a)\exp(\mathrm{i}k_m x)\right]$$
$$=\frac{F_x(\omega)}{L}\sum_{n=-\infty}^{n=+\infty}\left[\sum_{m=-\infty}^{m=+\infty}U_m(a)\exp(\mathrm{i}k_m x)\right]\exp(\mathrm{i}2\pi nx/L). \quad (36)$$

Because both the n and m summations run from minus infinity to plus infinity, the following relationship must hold true:

$$\sum_{n=-\infty}^{n=+\infty}\left[\sum_{m=-\infty}^{m=+\infty}W_m(a)\exp(\mathrm{i}k_m x)\right]\exp(\mathrm{i}2\pi nx/L)$$
$$=\left[\sum_{n=-\infty}^{n=+\infty}W_n(a)\right]\sum_{m=-\infty}^{m=+\infty}\exp(\mathrm{i}k_m x). \quad (37)$$

Equation (37) also applies to the $U_m(a)$ term, and inserting these results into Eqs. (35) and (36) yields

$$(\lambda+2\mu)\sum_{m=-\infty}^{m=+\infty}\frac{\partial W_m(a)}{\partial z}\exp(\mathrm{i}k_m x)$$
$$+\mathrm{i}k_m\lambda\sum_{m=-\infty}^{m=+\infty}U_m(a)\exp(\mathrm{i}k_m x)$$
$$=\frac{F_z(\omega)}{L}\left[\sum_{n=-\infty}^{n=+\infty}W_n(a)\right]\sum_{m=-\infty}^{m=+\infty}\exp(\mathrm{i}k_m x) \quad (38)$$

and

$$\mu\left[\sum_{m=-\infty}^{m=+\infty}\frac{\partial U_m(a)}{\partial z}\exp(\mathrm{i}k_m x)+\mathrm{i}k_m\sum_{m=-\infty}^{m=+\infty}W_m(a)\exp(\mathrm{i}k_m x)\right]$$
$$=\frac{F_x(\omega)}{L}\left[\sum_{n=-\infty}^{n=+\infty}U_n(a)\right]\sum_{m=-\infty}^{m=+\infty}\exp(\mathrm{i}k_m x). \quad (39)$$

Equations (30), (31), (38), and (39) are now all multiplied by $\exp(-\mathrm{i}k_p x)$ and integrated from $[0,L]$. This results in an orthogonal relationship, and the series terms of the equations will decouple into individual m indexed equations written as

$$(\lambda+2\mu)\frac{\partial W_m(b)}{\partial z}+\mathrm{i}k_m\lambda U_m(b)+\left(\frac{\omega^2\rho_f}{\mathrm{i}\gamma_m}\right)W_m(b)$$
$$=\begin{cases}-2P_I(\omega) & m=0\\ 0 & m\neq 0,\end{cases} \quad (40)$$

$$\mu\frac{\partial U_m(b)}{\partial z}+\mathrm{i}k_m\mu W_m(b)=0, \quad (41)$$

$$(\lambda+2\mu)\frac{\partial W_m(a)}{\partial z}+\mathrm{i}k_m\lambda U_m(a)=\frac{F_z(\omega)}{L}\left[\sum_{n=-\infty}^{n=+\infty}W_n(a)\right], \quad (42)$$

and

$$\mu\frac{\partial U_m(a)}{\partial z}+\mathrm{i}k_m\mu W_m(a)=\frac{F_x(\omega)}{L}\left[\sum_{n=-\infty}^{n=+\infty}U_n(a)\right]. \quad (43)$$

Next the functional form of the displacements from Eqs. (19) and (20) are inserted into Eqs. (40)–(43) and the following algebraic matrix equation is obtained:

$$[\mathbf{A}^{(0)}(k)]\{\mathbf{y}^{(0)}(k)\}=\sum_{n=-\infty}^{n=+\infty}[\mathbf{F}^{(n)}(k_n)]\{\mathbf{y}^{(n)}(k_n)\}+\mathbf{p}, \quad (44)$$

where $[\mathbf{A}^{(0)}(k)]$ is a four by four matrix that models the dynamics of the plate for $m=0$, $\{\mathbf{y}^{(0)}(k)\}$ is the four by one vector of wave propagation coefficients for $m=0$, $[\mathbf{F}^{(n)}(k_n)]$ is the four by four matrix that represents the periodic transducer loading on the structure for nth mode, $\{\mathbf{y}^{(n)}(k_n)\}$ is the four by one vector of wave propagation coefficients for nth mode, and $\mathbf{p}$ is the four by one vector that models the plane wave excitation. The entries of the matrices and vectors in Eq. (44) are listed in the Appendix. To facilitate a solution to the problem, index shifting is employed. The integer shift property of an infinite summation is applied to Eq. (44), which, because of the summation running from minus infinity to positive infinity, results in

$$[\mathbf{A}^{(m)}(k_m)]\{\mathbf{y}^{(m)}(k_m)\}=\sum_{n=-\infty}^{n=+\infty}[\mathbf{F}^{(n+m)}(k_{n+m})]\{\mathbf{y}^{(n+m)}(k_{n+m})\}$$
$$+\begin{cases}\mathbf{p}, & m=0\\ \mathbf{0}, & m\neq 0\end{cases}=\sum_{m=-\infty}^{m=+\infty}[\mathbf{F}^{(m)}(k_m)]$$
$$\times\{\mathbf{y}^{(m)}(k_m)\}+\begin{cases}\mathbf{p}, & m=0\\ \mathbf{0}, & m\neq 0.\end{cases} \quad (45)$$

where the **0** term is a four by one vector whose entries are zeros. Once the $[\mathbf{A}^{(m)}]$ matrix is integer-indexed and the transducer load matrix indexes have been shifted, the system equations can be rewritten using all the n indexed modes as

$$\mathbf{A}\mathbf{y} = \mathbf{F}\mathbf{y} + \mathbf{P}, \tag{46}$$

where $\mathbf{A}$ is a block-diagonal matrix and is equal to

$$\mathbf{A} = \begin{bmatrix} \ddots & & \vdots & & \iddots \\ & [\mathbf{A}^{(-1)}(k_{-1})] & \mathbf{0} & \mathbf{0} & \\ \cdots & \mathbf{0} & [\mathbf{A}^{(0)}(k)] & \mathbf{0} & \cdots \\ & \mathbf{0} & \mathbf{0} & [\mathbf{A}^{(1)}(k_1)] & \\ \iddots & & \vdots & & \ddots \end{bmatrix}, \tag{47}$$

$\mathbf{F}$ is a rank deficient, block-partitioned matrix and is written as

$$\mathbf{F} = \begin{bmatrix} \ddots & & \vdots & & \iddots \\ & [\mathbf{F}^{(-1)}(k_{-1})] & [\mathbf{F}^{(0)}(k)] & [\mathbf{F}^{(1)}(k_1)] & \\ \cdots & [\mathbf{F}^{(-1)}(k_{-1})] & [\mathbf{F}^{(0)}(k)] & [\mathbf{F}^{(1)}(k_1)] & \cdots \\ & [\mathbf{F}^{(-1)}(k_{-1})] & [\mathbf{F}^{(0)}(k)] & [\mathbf{F}^{(1)}(k_1)] & \\ \iddots & & \vdots & & \ddots \end{bmatrix}, \tag{48}$$

$\mathbf{P}$ is the plane wave load vector

$$\mathbf{P} = [\cdots \mathbf{0}^T \ \mathbf{p}^T \ \mathbf{0}^T \ \cdots]^T, \tag{49}$$

and $\mathbf{y}$ is the wave propagation coefficient vector that contains all the unknown indexed coefficients as

$$\mathbf{y} = [\cdots \{\mathbf{y}^{(-1)}(k_{-1})\}^T \{\mathbf{y}^{(0)}(k)\}^T \{\mathbf{y}^{(1)}(k_1)\}^T \cdots]^T, \tag{50}$$

where the unknown zero indexed wave propagation coefficients are contained in the equations as

$$\begin{aligned} \{\mathbf{y}^{(0)}(k)\} &= \{A(k,\omega)\ B(k,\omega)\ C(k,\omega)\ D(k,\omega)\}^T \\ &\equiv \{A_0(k,\omega)\ B_0(k,\omega)\ C_0(k,\omega)\ D_0(k,\omega)\}^T. \end{aligned} \tag{51}$$

The **0** term in Eq. (47) is a four by four matrix whose entries are all zeros and the **0** term in Eq. (49) is a four by one vector whose entries are all zeros. Equation (46) is assembled, and the wave-propagation coefficients that reside in the $\mathbf{y}$ vector are found by

$$\mathbf{y} = [\mathbf{A} - \mathbf{F}]^{-1}\mathbf{P}. \tag{52}$$

When the coefficients are determined, the displacements of the system, in the spatial domain, can be calculated using Eqs. (16) and (17).

For analytical problems that model sonar systems, it is frequently desirable to transform the solution into the wave-number-frequency (k, ω) domain for analysis. For a function that is periodic on the interval $[0, L]$, the Fourier transform into the wave-number domain is

$$\hat{\mathbf{U}}(k) = \frac{1}{L}\int_0^L \mathbf{u}(x,z,t)\exp(-\mathrm{i}kx)dx. \tag{53}$$

Inserting Eqs. (16) and (17) into Eq. (53) results in the integrand for all the $m \neq 0$ terms equaling zero, and this gives

$$\begin{aligned} \hat{u}(k,z,t) = [&A_0(k,\omega)\mathrm{i}k\exp(\mathrm{i}\alpha_0 z) + B_0(k,\omega)\mathrm{i}k\exp(-\mathrm{i}\alpha_0 z) \\ &- C_0(k,\omega)\mathrm{i}\beta_0\exp(\mathrm{i}\beta_0 z) + D_0(k,\omega)\mathrm{i}\beta_0 \\ &\times\exp(-\mathrm{i}\beta_0 z)]\exp(-\mathrm{i}\omega t) \end{aligned} \tag{54}$$

and

$$\begin{aligned} \hat{w}(k,z,t) = [&A_0(k,\omega)\mathrm{i}\alpha_0\exp(\mathrm{i}\alpha_0 z) - B_0(k,\omega)\mathrm{i}\alpha_0 \\ &\times\exp(-\mathrm{i}\alpha_0 z) + C_0(k,\omega)\mathrm{i}k\exp(\mathrm{i}\beta_0 z) \\ &+ D_0(k,\omega)\mathrm{i}k\exp(-\mathrm{i}\beta_0 z)]\exp(-\mathrm{i}\omega t), \end{aligned} \tag{55}$$

where the caret denotes the displacement function in the wave-number domain.

IV. MODEL VALIDATION

The sonar window—Tonpilz transducer interaction model can be compared and validated for a thin plate at low frequencies using a Bernoulli-Euler thin plate model that has been previously developed.[14–16] The stiffeners in these models are replaced by Tonpilz transducer dynamics so that the validation example corresponds to the thick plate model developed in Secs. II and III. The thin plate model has one degree of freedom that is the displacement in the z direction. This equation is written as

$$\frac{\hat{w}(k,\omega)}{P_I(\omega)} = 2T(k,\omega) \times \left[\frac{1 + \dfrac{F_z(\omega)}{L}T(k,\omega) - \dfrac{F_z(\omega)}{L}\displaystyle\sum_{n=-\infty}^{n=+\infty} T^{(n)}(k_n,\omega)}{1 - \dfrac{F_z(\omega)}{L}\displaystyle\sum_{n=-\infty}^{n=+\infty} T^{(n)}(k_n,\omega)}\right], \tag{56}$$

where

$$T^{(n)}(k_n,\omega) = \frac{-1}{Dk_n^4 - \rho h\omega^2 + \left(\dfrac{\rho_f\omega^2}{\mathrm{i}\gamma_n}\right)}, \tag{57}$$

and

$$D = \frac{Eh^3}{12(1-\upsilon^2)}, \tag{58}$$

where E is Young's modulus $(\mathrm{N/m^2})$ and υ is Poisson's ratio (dimensionless). Once the normal displacement is known, the transducer output is determined using Eq. (29).

Figure 3 is a plot of the transfer function of transducer voltage divided by input pressure versus wave number at a frequency of 100 Hz. This extremely low frequency was chosen because it is a value at which the two models should theoretically agree. Additionally, the evaluation is made versus wave number rather than arrival angle so that the higher wave number dynamics are included in the comparison. This example was generated with the following system parameters: window thickness h is 0.005 m, window density ρ is 1200 $\mathrm{kg/m^3}$, Lamé constant λ is 9.31×10^8 $\mathrm{N/m^2}$, Lamé constant μ is 1.03×10^8 $\mathrm{N/m^2}$, fluid density ρ_f is 1000 $\mathrm{kg/m^3}$, fluid compressional wave speed c_f is

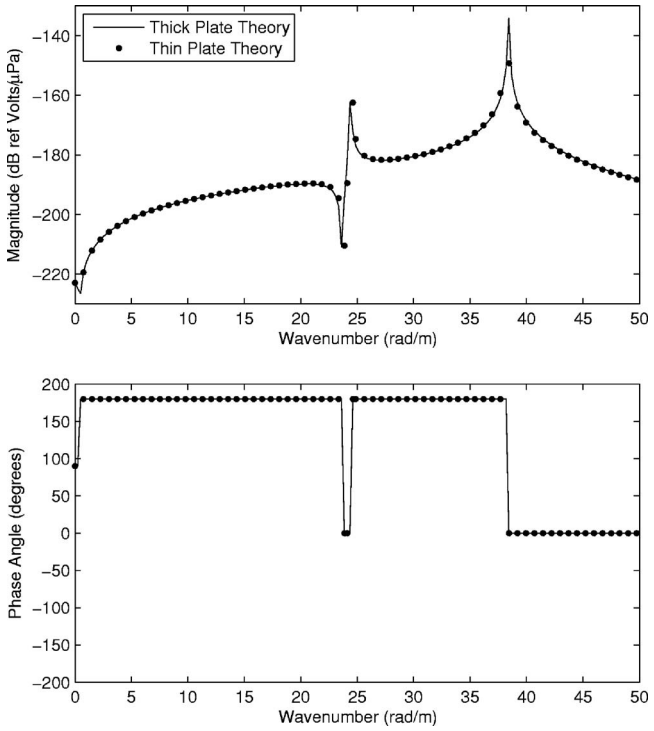


FIG. 3. Transducer voltage divided by input pressure vs wave number at a frequency of 100 Hz.

1500 m/s, transducer head mass M_H is 2.5 kg/m, transducer tail mass M_T is 10.0 kg/m, transducer stiffness K_S is 1×10^6 N/m^2, transducer separation distance L is 0.1 m, transducer face width d is 0.095 m, transducer stack height t is 0.01 m, and transducer constant g_{33} is 0.025 (V m)/N. In Fig. 3, the solid line is the elastic plate theory developed in Secs. II and III and corresponds to Eq. (55); and the black dot symbols are the Bernoulli-Euler plate theory and correspond to Eq. (56). The fully elastic plate model was calculated using eleven modes ($-5 < n < 5$) that produced a 20-by 20-element system matrix while the thin plate model was calculated using 51 modes ($-25 < n < 25$). Note that there is agreement between the two models over the entire wave number region.

V. A NUMERICAL EXAMPLE

A numerical example to illustrate the dynamics of a sonar window interacting with the Tonpilz transducer array is now presented. To understand fully the model results, it is necessary at this point to slightly digress. The features present in the Tonpilz transducer output are best understood if the dispersion curve for the system is studied along with the element output. The dispersion curve is found by setting

$$\det[\mathbf{A} - \mathbf{F}] = 0, \tag{59}$$

and this determines the location in the wave-number-frequency plane where free wave propagation can exist. Each of these free waves is related to a specific dynamic motion of the system. In Fig. 4, the dispersion curve is shown (in x's) plotted on the transducer response (in color) versus arrival angle and frequency. The scale shown above the image is a colorbar and corresponds to the transducer output in units of dB referenced to V/μPa. The following system parameters were used for this example: window thickness h is 0.1 m, window density ρ is 1200 kg/m^3, Lamé constant λ is 1.43×10^9 N/m^2, Lamé constant μ is 3.57×10^8 N/m^2, fluid density ρ_f is 1000 kg/m^3, fluid compressional wave speed c_f is 1500 m/s, transducer head mass M_H is 1 kg/m, transducer tail mass M_T is 4 kg/m, transducer stiffness K_S is 1×10^7 N/m^2, transducer separation distance L is 0.1 m, transducer face width d is 0.095 m, transducer stack height t is 0.01 m, and transducer constant g_{33} is 0.025 (V m)/N. The noticeable feature from Fig. 4 is that the system is very rich with free wave propagation above 2490 Hz. This frequency value corresponds to the first antisymmetric Lamb wave of the system. Without the mass-loading, this would occur at

$$f = \frac{c_s}{2h} = 2730 \text{ Hz}; \tag{60}$$

however, because the transducers mass-load the plate, this location is shifted downward in frequency. It is noted that the multiple free waves correspond to higher order plate waves, fluid/structure interaction waves, and the first set of spatially periodic waves, i.e., $n = \pm 1$. These periodic waves are related to the spacing of the individual transducers and occur in wave number at integer multiples of

$$k = \frac{2\pi}{L}. \tag{61}$$

These periodic waves are frequently called Floquet[17] and/or Bloch[17,18] waves. An additional note to the dispersion curve is included. Although Eq. (59) may predict a specific free wave, it may not necessarily propagate in the structure. (An acoustic load may not excite a specific free wave, or the excitation may result in an extremely small response.) The propagation of a wave is dependent on the interaction of the structural load with the physics of the free wave. The response of the system shown in Fig. 4 is extremely complex, however, most of the features that are present are nulls. Nulls frequently exist between branches of the dispersion curve that have split or separated. This separation is due to the fluid loading on the plate. Figure 5 is three cuts of Fig. 4 at 1500, 2610, and 4030 Hz, respectively, and is shown to illustrate the magnitude (solid line) and phase angle (dashed line) values versus arrival angle at three frequencies. The last two frequencies are shown to illustrate the nulling features (or system zero dynamics) that are present at higher frequencies.

Finally, the beamformed array response of a linear array can be found using

$$B(k, \omega) = \sum_{n=1}^{N} \text{rvs}(k, \omega) \exp[\text{i}(k - k_s)x_n], \tag{62}$$

where $\text{rvs}(k, \omega)$ is the receive voltage sensitivity in the wave-number-frequency domain, N is the number of sensors in the array, k_s is the steered wave number (rad/m), and x_n is the location of the nth sensor (m). For the following analysis, a 16-sensor (element) array is used. The beam patterns are displayed as polar plots, with the solid lines in the plots corresponding to array response based on the theory devel-

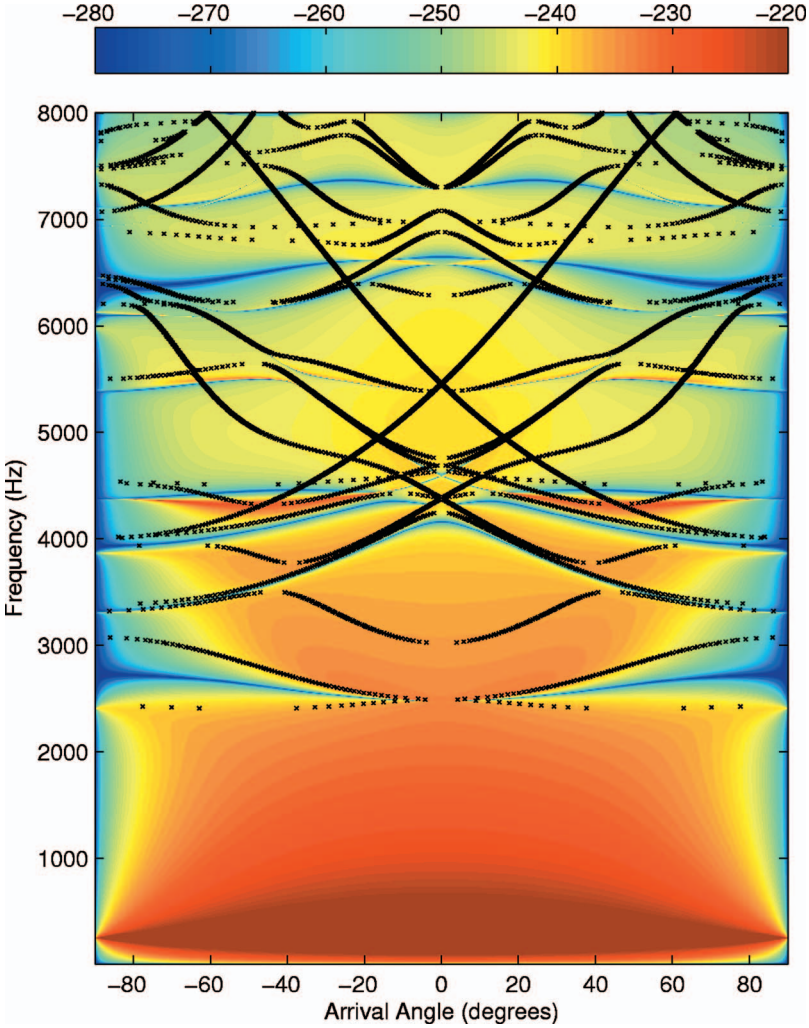


FIG. 4. Transducer voltage divided by input pressure vs arrival angle and frequency (color) overlaid with system dispersion curve (x's). Units in dB ref. V/μPa.

oped previously [Eqs. (1)–(55)], and the dashed lines are the response of the array to unity input at all wave numbers. Each short dashed concentric half circle represents 10 dB of energy. Because these beam patterns are being displayed as polar plots, the wave numbers in Eq. (62) have been converted into arrival angles using Eq. (26). Figure 6 is a plot of the beamformed response at 1500 Hz with a steer angle of 0° (top) and 30° (bottom). It is noted that in these cases, the sonar window improves the response of the beamformer. This is primarily due to the receive energy drop off at large arrival angles. Figure 7 is a plot of the beamformed response at 2610 Hz with a steer angle of 0° (top) and 45.1° (bottom), an angle that corresponds to a null in the system response. For this frequency and 45.1° steer angle, the array beamformer response has significantly degraded. Figure 8 is a plot of the beamformed response at 4030 Hz with a steer angle of 0° (top) and 12.5° (bottom). Steering into the null at 12.5° does not seriously degrade the beamformer response. Ultimately, any frequency and steer angle can be modeled to determine if there are detrimental effects at the operating parameters of interest.

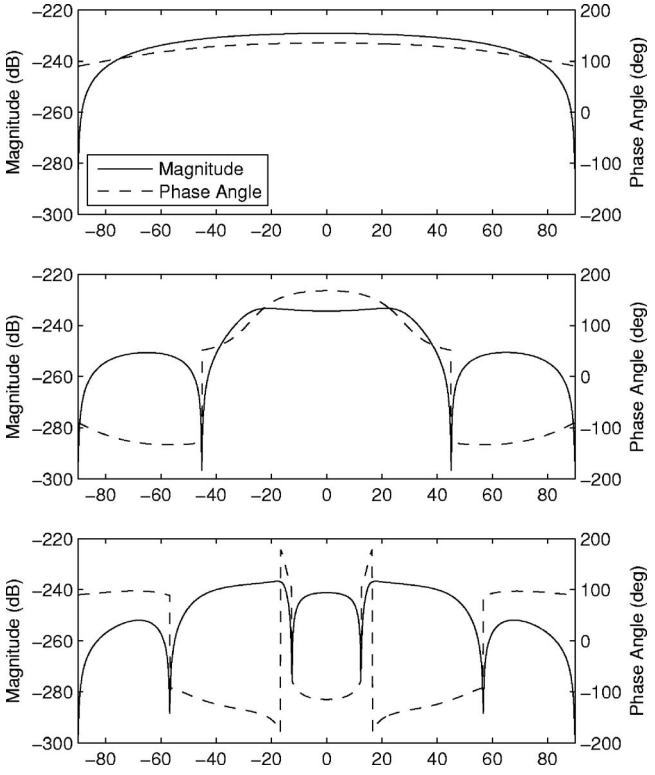


FIG. 5. Transducer voltage divided by input pressure vs arrival angle at 1500 Hz (top), 2610 Hz (middle), and 4030 Hz (bottom).

VI. CONCLUSIONS

This paper has derived the equations of motion of an insonified fully elastic sonar window that is attached to an

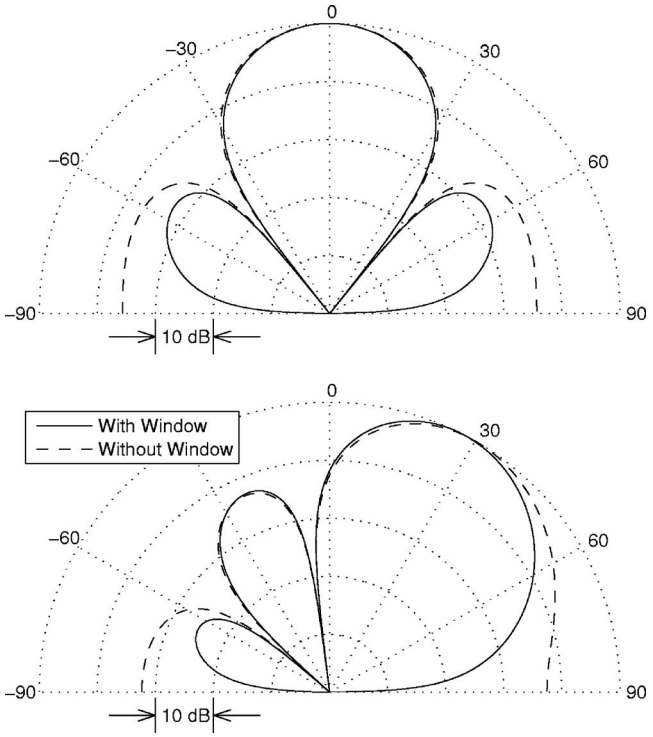


FIG. 6. Beamformed array response at 1500 Hz with a steer angle of 0° (top) and 30° (bottom).

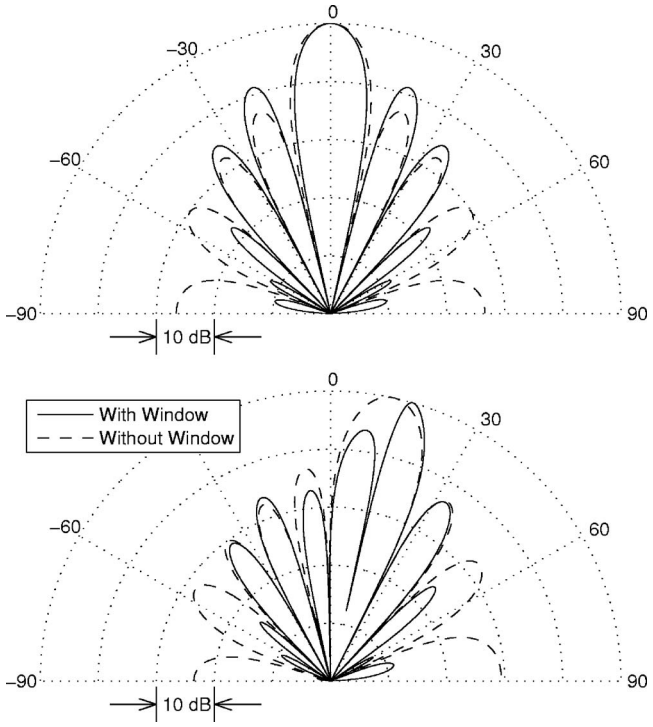


FIG. 8. Beamformed array response at 4030 Hz with a steer angle of 0° (top) and 12.5° (bottom).

array of single resonant Tonpilz transducers. The equations were then manipulated so that the displacement field of the sonar window and the electrical output of the transducers could be determined. Once this is done, the results can be beamformed to predict the response of an array of sensors for all frequencies and wave numbers or arrival angles. This truncated analytical model was compared to a single wave, low frequency model of the system and found to be in close agreement. This new fully elastic model specifically shows how higher order plate waves interact with the beam pattern of an array. Furthermore, the model predicts the location of Floquet waves and how they enter into the analysis. This is useful as sonar systems are built or modified for broadband processing. The model predicts the locations in the acoustic cone that are smooth for optimum sonar processing and the locations where the effects of null responses will enter into the processing.

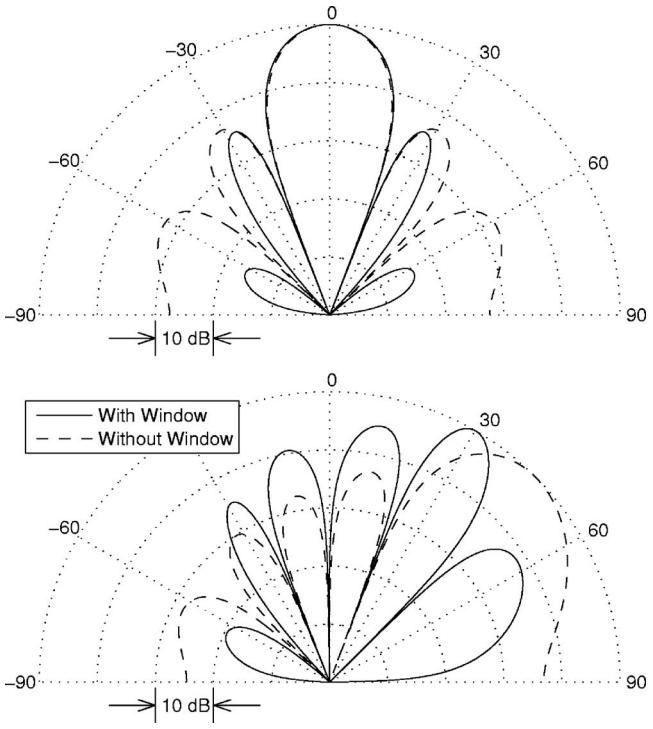


FIG. 7. Beamformed array response at 2610 Hz with a steer angle of 0° (top) and 45.1° (bottom).

ACKNOWLEDGMENTS

The work described in this paper was sponsored by program officer David M. Drumheller (Code 333) at the Office of Naval Research (ONR). The author is indebted to Stephen C. Butler and John B. Blottman, both of the Naval Undersea Warfare Center, for discussions on Tonpilz transducer behavior.

APPENDIX: MATRIX AND VECTOR ENTRIES

The entries of the matrixes and vectors in Eq. (52) are listed in the following. Without loss of generality, the top of the plate is defined as $z=b=0$. For the $[\mathbf{A}^{(n)}(k_n)]$ matrix, the nonzero entries are

$$a_{11} = -\alpha_n^2\lambda - 2\alpha_n^2\mu - \lambda k_n^2 + \frac{\alpha_n\omega^2\rho_f}{\gamma_n}, \tag{A1}$$

$$a_{12} = -\alpha_n^2\lambda - 2\alpha_n^2\mu - \lambda k_n^2 - \frac{\alpha_n\omega^2\rho_f}{\gamma_n}, \tag{A2}$$

$$a_{13} = -2\mu k_n \beta_n + \frac{k_n \omega^2 \rho_f}{\gamma_n}, \tag{A3}$$

$$a_{14} = 2\mu k_n \beta_n + \frac{k_n \omega^2 \rho_f}{\gamma_n}, \tag{A4}$$

$$a_{21} = -2\mu k_n a_n, \tag{A5}$$

$$a_{22} = 2\mu k_n \alpha_n, \tag{A6}$$

$$a_{23} = \mu(\beta_n^2 - k_n^2), \tag{A7}$$

$$a_{24} = \mu(\beta_n^2 - k_n^2), \tag{A8}$$

$$a_{31} = (-\alpha_n^2 \lambda - 2\alpha_n^2 \mu - \lambda k_n^2)\exp(\mathrm{i}\alpha_n a), \tag{A9}$$

$$a_{32} = (-\alpha_n^2 \lambda - 2\alpha_n^2 \mu - \lambda k_n^2)\exp(-\mathrm{i}\alpha_n a), \tag{A10}$$

$$a_{33} = -2\mu k_n \beta_n \exp(\mathrm{i}\beta_n a), \tag{A11}$$

$$a_{34} = 2\mu k_n \beta_n \exp(-\mathrm{i}\beta_n a), \tag{A12}$$

$$a_{41} = -2\mu k_n \alpha_n \exp(\mathrm{i}\alpha_n a), \tag{A13}$$

$$a_{42} = 2\mu k_n \alpha_n \exp(-\mathrm{i}\alpha_n a), \tag{A14}$$

$$a_{43} = \mu(\beta_n^2 - k_n^2)\exp(\mathrm{i}\beta_n a), \tag{A15}$$

and

$$a_{44} = \mu(\beta_n^2 - k_n^2)\exp(-\mathrm{i}\beta_n a). \tag{A16}$$

For the $[\mathbf{F}^{(n)}(k_n)]$ matrix, the nonzero entries are

$$f_{31} = \frac{F_z(\omega)}{L}(\mathrm{i}\alpha_n)\exp(\mathrm{i}\alpha_n a), \tag{A17}$$

$$f_{32} = \frac{F_z(\omega)}{L}(-\mathrm{i}\alpha_n)\exp(-\mathrm{i}\alpha_n a), \tag{A18}$$

$$f_{33} = \frac{F_z(\omega)}{L}(\mathrm{i}k_n)\exp(\mathrm{i}\beta_n a), \tag{A19}$$

$$f_{34} = \frac{F_z(\omega)}{L}(\mathrm{i}k_n)\exp(-\mathrm{i}\beta_n a), \tag{A20}$$

$$f_{41} = \frac{F_x(\omega)}{L}(\mathrm{i}k_n)\exp(\mathrm{i}\alpha_n a), \tag{A21}$$

$$f_{42} = \frac{F_x(\omega)}{L}(\mathrm{i}k_n)\exp(-\mathrm{i}\alpha_n a), \tag{A22}$$

$$f_{43} = \frac{F_x(\omega)}{L}(-\mathrm{i}\beta_n)\exp(\mathrm{i}\beta_n a), \tag{A23}$$

and

$$f_{44} = \frac{F_x(\omega)}{L}(\mathrm{i}\beta_n)\exp(-\mathrm{i}\beta_n a). \tag{A24}$$

The **y** vector entries are

$$\mathbf{y} = \{\ldots\ A_{-1}\ B_{-1}\ C_{-1}\ D_{-1}\ A_0\ B_0\ C_0\ D_0\ A_1\ B_1\ C_1\ D_1\ \ldots\}^T. \tag{A25}$$

The **p** vector entries are

$$\mathbf{p} = \{-2P_I(\omega)\ 0\ 0\ 0\}^T. \tag{A26}$$

[1] D. Stansfield, *Underwater Electroacoustic Transducers, A Handbook for User and Designers* (Bath University Press and Institute of Acoustics, Bath, United Kingdom, 1991), pp. 179–195.

[2] M. P. Johnson, "Equivalent modal impedance matrix of multiple degree of freedom electroelastic structures," J. Acoust. Soc. Am. **88**, 1–6 (1990).

[3] Y. Roh and X. Lu, "Design of an underwater Tonpilz Transducer with 2-2 mode piezocomposite materials," J. Acoust. Soc. Am. **119**, 3734–3740 (2006).

[4] C. Desilets, G. Wojcik, L. Nikodym, and K. Mesterton, "Analysis and measurements of acoustically matched, air-coupled Tonpilz transducers," Proc.-IEEE Ultrason. Symp. **2**, 1045–1048 (1999).

[5] M. B. Moffett, J. M. Powers, and M. D. Jevnager, "A Tonpilz projector for use in an underwater horn," J. Acoust. Soc. Am. **103**, 3353–3361 (1998).

[6] J. C. Piquette, "Applications of the method for transducer transient suppression to various transducer types," J. Acoust. Soc. Am. **94**, 646–651 (1993).

[7] M. Van Crombrugge and W. Thompson, Jr., "Optimization of the transmitting characteristics of a Tonpilz-type transducer by proper choice of impedance matching layers," J. Acoust. Soc. Am. **77**, 747–752 (1985).

[8] C. M. Thompson, "Development of a structurally rigid, acoustically transparent plastic," J. Acoust. Soc. Am. **87**, 1138–1143 (1990).

[9] D. L. Folds and C. D. Loggins, "Transmission and reflection of ultrasonic waves in layered media," J. Acoust. Soc. Am. **62**, 1102–1109 (1977).

[10] E. E. Mikeska and J. A. Behrens, "Evaluation of transducer window materials," J. Acoust. Soc. Am. **59**, 1294–1298 (1976).

[11] M. Kim and Y. F. Hwang, "An analysis of wave dispersion in coarsely laminated symmetric composite plates," J. Acoust. Soc. Am. **100**, 1981–1991 (1996).

[12] J. S. Hickman, D. E. Risty, and E. S. Stewart, "Properties of sandwich-type structures as acoustic windows," J. Acoust. Soc. Am. **29**, 858–864 (1957).

[13] J. O. R. Blake, R. A. Shenoi, J. House, and T. Turton, "Strength modeling in stiffened FRP structures with viscoelastic inserts for ocean structures," Ocean Eng. **29**, 849–869 (2002).

[14] B. A. Cray, "Acoustic radiation from periodic and sectionally aperiodic rib-stiffened plates," J. Acoust. Soc. Am. **95**, 256–264 (1994).

[15] B. R. Mace, "Periodically stiffened fluid-loaded plates. I Response to convected harmonic pressure and free wave propagation," J. Sound Vib. **73**, 473–486 (1980).

[16] B. R. Mace, "Periodically stiffened fluid-loaded plates. II Response to line and point forces," J. Sound Vib. **73**, 487–504 (1980).

[17] M. Tran-Van-Nhieu, "Scattering from a ribbed finite cylindrical shell with internal axisymmetric oscillators," J. Acoust. Soc. Am. **112**, 402–410 (2002).

[18] D. M. Photiadis, "The effect of irregularity on the scattering of acoustic waves from a ribbed plate," J. Acoust. Soc. Am. **91**, 1897–1903 (1992).

Pseudo-damping in undamped plates and shells

A. Carcaterra[a)]
Department of Mechanics and Aeronautics, University of Rome, "La Sapienza," Via Eudossiana, 18, 00184, Rome, Italy and Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

A. Akay
Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

F. Lenti
Department of Mechanics and Aeronautics, University of Rome, "La Sapienza," Via Eudossiana, 18, 00184, Rome, Italy

(Received 22 June 2006; revised 27 April 2007; accepted 4 May 2007)

Pseudo-damping is a counter-intuitive phenomenon observed in a special class of linear structures that exhibit an impulse response characterized by a decaying amplitude, even in the absence of any dissipation mechanism. The conserved energy remains within but designated parts of the system. Pseudo-damping develops when the natural frequency distribution of the system includes condensation points. The recently formulated theoretical foundation of this phenomenon, based on mathematical properties of special trigonometric series, makes it possible to describe a class of mechanical systems capable of displaying pseudo-damping characteristics. They include systems with discrete oscillators and one-dimensional continuous beamlike structures already reported by the authors in recent studies. This paper examines development of pseudo-damping phenomenon in two-dimensional structures, using plates and shells as examples, and shows how a preloaded plate on an elastic foundation can lead to pseudo-damping. Moreover, in the case of curved shell elements examined here, pseudo-damping can result due to the curvature of the structure, which naturally introduces condensation points in the modal density.
[DOI: 10.1121/1.2747093]

PACS number(s): 43.40.At, 43.40.Kd, 43.40.Jc, 43.40.Dx, 43.40.Ey [ADP] Pages: 804–813

I. PSEUDO-DAMPING AND NEAR-IRREVERSIBILITY: THEORETICAL BACKGROUND

The impulse response of a mechanical system is known to have an asymptotically vanishing amplitude only in the presence of dissipation effects. Although in most cases this observation is valid, it has been shown that this expectation is not the general rule. In fact, the observed, or apparent, damping in a master structure does not depend only on its inherent dissipation effect but also on the capability of attached additional systems to absorb its energy.[1–3] This last effect can be so important that an apparent damping, leading to an asymptotically damped motion of the master, can be exhibited from structures even in the absence of dissipation effects as shown in Ref. 4. In this case, the system attached to the master has an infinite number of degrees of freedom but a finite total mass. In Refs. 5–7, a transient damping effect has been studied also for a finite number of degrees of freedom system. It appears in these cases the damping effect induced on the master is only temporary, but after a characteristic return time[7] the energy is transferred back to the master. The apparent damping has been also experimentally demonstrated.[8] Recent studies point to the existence of a class of conservative mechanical systems, with a finite number of degrees of freedom and having a particular frequency distribution, that exhibit an apparent damping with a near-irreversible decaying trend in their impulse response.[9] A similar phenomenon appears also in continuous structures, as in the beam illustrated in Ref. 10. In these cases, the spatial redistribution of the conserved energy within a conservative linear system produces appearance of damping in the system, for redistribution takes place with near irreversibility, without recurrence, or return to its original form. This amounts to a completely new phenomenon we refer to here as pseudo-damping. In such systems, energy, although conserved, is transported away from the point at which it is induced, without returning to it. A general explanation for these phenomena has been recently formulated in Ref. 11.

The present paper, as those described in Refs. 10 and 11, is about conservative systems that do not exhibit recurrence, or energy return, due to certain properties of their natural frequency distributions; specifically, frequency distributions that contain one or more condensation points, as it happens in certain two-dimensional structures described later.

The relationship between a condensation point in the natural frequency distribution of a conservative linear system and its ability to spatially redistribute its vibratory energy with near irreversibility emerges when the discrete Fourier series that represents an impulse response is compared with its integral counterpart.

The general expression for impulse response $h(t)$ of a linear elastic structure has the form

[a)]Electronic mail: a.carcaterra@dma.ing.uniroma1.it

$$h(t) = \sum_{i=1}^{M} G(\omega_i) \sin \omega_i t, \tag{1}$$

where ω_i represents the structure's natural frequencies and $G(\omega_i)$ is its modal participation factors. The continuous form, or integral counterpart, of $h(t)$ becomes

$$h_{\text{int}}(t) = \int_0^1 G[\omega(\xi)] \sin \omega(\xi) t \, d\xi, \tag{2}$$

where $\omega(\xi)$ represents the distribution of frequency and ξ is a dummy variable. Under conditions specified in Ref. 1, $h_{\text{int}}(t)$ has the asymptotic property

$$h_{\text{int}} \to 0 \quad \text{for} \quad t \to \infty, \tag{3}$$

which shows that the integral associated with the impulse response vanishes asymptotically, even without any dissipation in the system, thus exhibiting the pseudo-damping described above. However, this property does not extend to the actual impulse response, $h(t)$, even though it is an approximation of the integral $h_{\text{int}}(t)$. Notwithstanding, the two expressions h and h_{int} are related through a remainder term $\Re_M(t)$:

$$\int_0^1 G[\omega(\xi)] \sin \omega(\xi) t \, d\xi = \Delta\xi \sum_{i=1}^{M} G(\omega_i) \sin \omega_i t + \Re_M(t), \tag{4}$$

where $\Delta\xi = 1/M$, $\xi_M(t) \in [0,1]$, and $\Re_M(t)$ depends on the number of modes M considered in the modal expansion. The magnitude of the remainder term $\Re_M(t)$ relates to the choice of the frequency distribution $\omega(\xi)$ as[7,10]

$$|\Re_M(t)| \propto \max\left\{\left|\frac{d\omega}{d\xi}\right|\right\} \quad \text{for} \quad \xi \in [0,1]. \tag{5}$$

It follows that those distributions having one or more stationary points, $d\omega/d\xi = 0$, within the interval $\xi \in [0,1]$, engender a smaller remainder term within that interval. Accordingly, in such cases $h(t)$ approaches $h_{\text{int}}(t)$ and thus develops similar pseudo-damping characteristics.

A similar relationship to that between condensation points in a natural frequency distribution and pseudo-damping also appears between the ratio $d\omega/d\xi$ and the modal density $\nu(\omega)$ of the structure: $d\omega/d\xi \propto 1/\nu(\omega)$. The condensation points, which represent singularities in the modal density, also correspond to group velocities that vanish at certain frequencies, and this amounts to a filtering of the wave energy propagation,[10] producing a pseudo-damping effect by spatially redistributing energy.

A recent study reports a closed-form expression for the natural frequency distribution of a conservative linear system in order for it to possess pseudo-damping characteristics.[11] The theory, based on the requirement to minimize the remainder term $\Re_M(t)$, or the distance between the functions $h(t)$ and $h_{\text{int}}(t)$, leads to a functional problem that depends on the frequency distribution $\omega(\xi)$. The resulting closed-form solution, again, establishes the correspondence between the pseudo-damping and the presence of condensation points as described above.

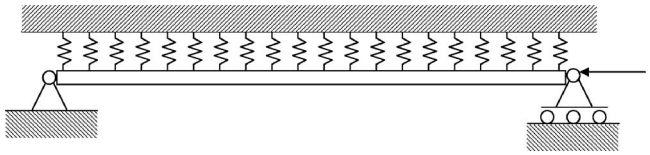


FIG. 1. Schematic description of a simply supported beam on an elastic foundation with an axial compressive force applied at each end.

The natural frequency distributions of many structures do not inherently contain condensation points but can be introduced to the system. Such an example, depicted in Fig. 1, consists of a simply supported beam on an elastic foundation and, under axial preload, demonstrates pseudo-damping in a conservative linear system.[10] The preload as well as the stiffness of the elastic foundation can be selected *a priori* to suitably modify the natural frequencies of the beam to produce a condensation around a desired frequency, resulting in a typical impulse response at the middle point of the beam as shown in Fig. 2.

This paper, inspired by the observation that certain two-dimensional continuous structures, such as shells, frequently exhibit condensation points in their natural frequency distributions,[12–14] also investigates pseudo-damping properties that may naturally exist in such structures.

II. PSEUDO-DAMPING IN TWO-DIMENSIONAL SYSTEMS

The equation of motion of a two-dimensional shell-like linear undamped structure can be written in the general form:

$$\boldsymbol{\alpha}\mathbf{D}\mathbf{w} + \boldsymbol{\beta}\mathbf{D}\ddot{\mathbf{w}} = \mathbf{0}, \tag{6}$$

where the vector displacement $\mathbf{w}(x_1, x_2, t)$ has two tangential and one normal component with respect to the surface structure as the reference. An initial impulsive velocity applied at

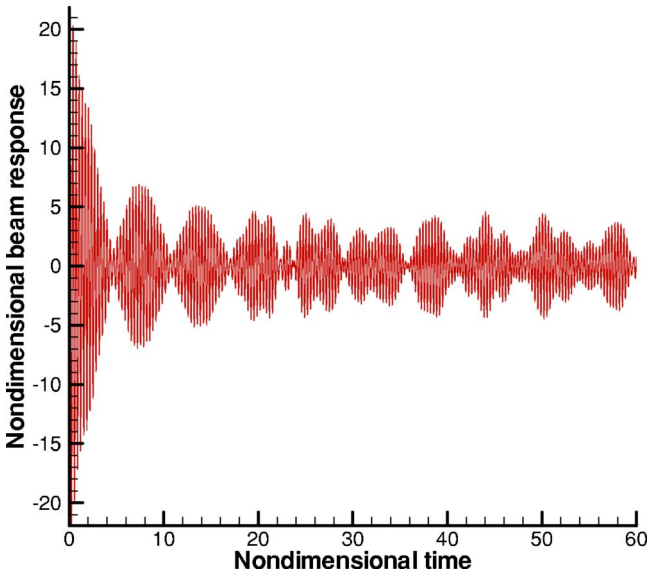


FIG. 2. (Color online) A typical impulse response of a beam with no internal dissipation shows a decay when its natural frequency distribution has a condensation point provided by the spring on which it rests and the axial compressive force.

a point $\mathbf{x}_0=(x_{10},x_{20})$ can be represented in terms of the initial conditions:

$$\mathbf{w}(x_1,x_2,0)=0,$$

$$\dot{\mathbf{w}}(x_1,x_2,0)=\mathbf{V}_0L_1L_2\delta(x_1-x_{10})\,\delta(x_2-x_{20}),$$

where x_1 and x_2 are generalized coordinates along the shell reference surface $x_1\in[0,L_1]$, $x_2\in[0,L_2]$, $\mathbf{V}_0$ is the initial velocity at $\mathbf{x}_0$, and δ represents the Dirac distribution. Tensors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ represent elastic and mass properties of the structure and $\mathbf{D}$ is a tensor of derivatives:

$$\boldsymbol{\alpha}=[\alpha_{lm}^{rs}],\quad \boldsymbol{\beta}=[\beta_{lm}^{rs}],\quad \mathbf{D}=\left[\frac{\partial^{(l+m)}}{\partial x_1^l\,\partial x_2^m}(\cdot)\right],$$
$$\mathbf{w}=[w^r],\quad r,s=1,2,3. \tag{7}$$

Superscripts r, s refer to the coordinate axes. The minimum value each of the derivation-related superscripts l and m take is 0, and their maximum values depend on the type of structure. For example, the value 8 is used in the examples considered in this paper.

The dispersion relationship, which for these structures[15–17] can be expressed starting with solutions of the kind $\mathbf{w}=\mathbf{A}g(x_1,x_2,t)$, with $g(x_1,x_2,t)=e^{-jk_1x_1}e^{-jk_2x_2}e^{j\omega t}$, where t is time, ω is the frequency, and $\mathbf{A}$ is the displacement amplitude vector, produces

$$\mathbf{D}(\mathbf{w})=\mathbf{D}(g)\mathbf{A}=\mathbf{p}\mathbf{A}g,\quad \mathbf{p}=[p_{lm}]=[(-j)^{l+m}k_1^lk_2^m], \tag{8}$$

which, when substituted into the equation of motion (6), yields a linear homogeneous set of equations in terms of $\mathbf{A}$:

$$[(\boldsymbol{\alpha}-\omega^2\boldsymbol{\beta})\mathbf{p}]\mathbf{A}=\mathbf{0}. \tag{9}$$

Existence of nontrivial solution to this set of equation requires that

$$\det(\boldsymbol{\alpha}-\boldsymbol{\beta}\omega^2)=\det\mathbf{C}=0,\quad \mathbf{C}=[C^{rs}], \tag{10}$$

which in explicit form becomes

$$\varepsilon_{hnq}C^{2h}C^{3n}C^{1q}=0,$$

$$\varepsilon_{hnq}p_{lm}p_{l'm'}p_{l''m''}(\alpha_{lm}^{2h}-\omega^2\beta_{lm}^{2h})(\alpha_{l'm'}^{3n}-\omega^2\beta_{l'm'}^{3n})\times(\alpha_{l''m''}^{1q}-\omega^2\beta_{l''m''}^{1q})=0, \tag{11}$$

where ε_{hnq} is the Ricci tensor. The dispersion expression follows as

$$-A\omega^6+B\omega^4-C\omega^2+D=0, \tag{12}$$

where

$$A=\beta_{lm}^{2h}\beta_{l'm'}^{3h}\beta_{l''m''}^{1q}\varepsilon_{hnq}p_{lm}p_{l'm'}p_{l''m''},$$

$$B=(\alpha_{lm}^{2h}\beta_{l'm'}^{3h}\beta_{l''m''}^{1q}+\beta_{lm}^{2h}\beta_{l'm'}^{1q}\alpha_{l''m''}^{3n}+\beta_{lm}^{2h}\beta_{l'm'}^{3h}\alpha_{l''m''}^{1q})\varepsilon_{hnq}p_{lm}p_{l'm'}p_{l''m''},$$

$$C=(\alpha_{lm}^{2h}\alpha_{l'm'}^{3h}\beta_{l''m''}^{1q}+\alpha_{lm}^{2h}\alpha_{l'm'}^{1q}\beta_{l''m''}^{3n}+\beta_{lm}^{2h}\alpha_{l'm'}^{3h}\alpha_{l''m''}^{1q})\varepsilon_{hnq}p_{lm}p_{l'm'}p_{l''m''},$$

$$D=\alpha_{lm}^{2h}\alpha_{l'm'}^{3h}\alpha_{l''m''}^{1q}\varepsilon_{hnq}p_{lm}p_{l'm'}p_{l''m''}.$$

From the dispersion relation, an explicit relation between frequency and wave numbers results as

$$\omega=f(k_1,k_2). \tag{13}$$

For the purpose of illustration, assume a displacement field that vanishes at the boundary:

$$\Phi_{nm}=A_{nm}\sin k_{n1}x_1\sin k_{m2}x_2,\quad k_{1n}=\frac{\pi n}{L_1},\quad k_{2m}=\frac{\pi m}{L_2}, \tag{14}$$

where $\Phi_{nm}(x)$ represents the normal modes. The choice of these particular boundary conditions, and the corresponding mode shapes, does not affect the results obtained in terms of modal density, which are, in general, independent of the boundary conditions. Substitution of these natural wave numbers into the dispersion relation produces a corresponding set of natural frequencies of the structure:

$$\omega_{nm}=f(k_{1n},k_{2m}). \tag{15}$$

Finally, in nondimensional form, the response of the structure, expressed with a finite number of M modes, follows as

$$W(T)=\sum_{n,m=1}^{M}\frac{1}{\Omega_{nm}}\sin^2\left(\frac{\pi n}{2}\right)\sin^2\left(\frac{\pi m}{2}\right)\sin 2\pi\Omega_{nm}T, \tag{16}$$

where $W=w\omega^*/4V_0$, $T=t\omega^*/2\pi$, $\Omega_{nm}=\omega_{nm}/\omega^*$, and ω^* is a reference frequency.

The system response consists of a combination of harmonic functions of the type discussed in Sec. I.

Pseudo-damping occurs when the frequency distribution $\Omega_{nm}=\omega_{nm}/\omega^*$ has a singularity point in its modal density. This general mathematical property, demonstrated rigorously in Ref. 11, has a physical counterpart based on the concept of group velocity. In some cases, a singularity in the modal density at a particular frequency corresponds to a vanishing group velocity or to a decrease of a directional average of the group velocity component normal to the wavefront at that frequency. The result is an inhibition of energy propagation at that frequency followed by a reduction of the amount of energy reflected back at the boundaries. This leads to an impulse response with a sharp decaying trend, even in the absence of any dissipation. Energy leaves the location at which the excitation force is applied, but inhibition of energy propagation towards the boundaries does not permit the impulse response to build up, thus producing the apparent damping effect.

III. MODAL DENSITY SINGULARITY AND ENERGY PROPAGATION

The section demonstrates the relationship between singularity in the modal density and some characteristic of the group velocity of a structure using polar coordinates in the k_1, k_2 plane to simplify the analysis. The vector expression for the wave number $\mathbf{k}=(k_1,k_2)$ takes the form

$$\mathbf{k} = k\mathbf{e}_k, \quad \mathbf{e}_k = (\cos\vartheta, \sin\vartheta), \quad k = \sqrt{k_1^2 + k_2^2},$$
$$\vartheta = \arctan\frac{k_2}{k_1}. \tag{17}$$

The corresponding expression for the dispersion relation suggests that, in the general case of anisotropic structures, frequency depends both on wave number and direction:

$$\omega = f(k, \vartheta). \tag{18}$$

The expressions for the modal distribution $N(\omega)$ and the modal density $\nu(\omega)$ follow as[12]

$$N(\omega) = \frac{1}{\Delta}\int_{\vartheta_1}^{\vartheta_2} k^2\, d\vartheta, \quad \Delta = \frac{\pi^2}{L_1 L_2},$$
$$\nu(\omega) = \frac{dN}{d\omega} = \frac{2}{\Delta}\int_{\vartheta_1}^{\vartheta_2} k\frac{\partial k}{\partial \omega}\, d\vartheta = \frac{2}{\Delta}\int_{\vartheta_1}^{\vartheta_2} \frac{k(\omega,\vartheta)}{\partial f/\partial k}\, d\vartheta, \tag{19}$$

where $[\vartheta_1, \vartheta_2]$ is the interval in which both k_1, k_2 are positive. Since for isotropic structures $\omega = f(k)$ is independent of ϑ, the previous equation simplifies as

$$\nu(\omega) = \frac{2k(\omega)}{\Delta c_g(\omega)}\int_0^{\pi/2} d\vartheta = \frac{\pi k(\omega)}{\Delta c_g(\omega)}, \tag{20}$$

where, in this case, $c_g(\omega) = \partial f/\partial k$ is the group velocity. Thus, for isotropic structures, any frequency ω^* for which $c_g(\omega^*)$ vanishes corresponds, generally, to a singularity point in the modal density, ν. Exceptions are met when $k(\omega^*) = 0$. In this case, ν could not be singular even if $c_g(\omega^*)$ vanishes. This happens, for instance, if the group velocity expression can be factorized in the form $c_g(\omega) = k(\omega)\,\eta(\omega)$, with $\eta(\omega^*) \neq 0$.

In the general case of anisotropic structures, the correspondence between modal density singularity and group velocity is less obvious. However, even in this case, Eq. (19) still provides a useful hint for physical interpretation of the pseudo-damping effect associated with a singularity for ν.

For anisotropic structures, the group velocity in the Cartesian reference system is expressed as vector, with $\mathbf{e}_1$, $\mathbf{e}_2$ as the unit vectors of the axes k_1, k_2, $\mathbf{c_g} = \nabla\omega = (\partial\omega/\partial k_1)\mathbf{e}_1 + (\partial\omega/\partial k_2)\mathbf{e}_2$. Here, $\mathbf{c_g}$ describes transport of energy within a narrow frequency bandwidth and may have a different direction than that of $\mathbf{k}$. The group velocity vector can be also expressed in polar coordinates $\mathbf{c_g} = \nabla\omega = (\partial\omega/\partial k)\mathbf{e}_k + (1/k) \times (\partial\omega/\partial\vartheta)\mathbf{e}_\vartheta$, where $\mathbf{e}_k = \mathbf{e}_1\cos\vartheta + \mathbf{e}_2\sin\vartheta$, $\mathbf{e}_\vartheta = -\mathbf{e}_1\sin\vartheta + \mathbf{e}_2\cos\vartheta$ are the unit vectors in directions parallel and normal to $\mathbf{k} = k\mathbf{e}_k$, respectively.

Considering the component of the group velocity parallel to the direction of wave number,

$$\mathbf{c_g}\cdot\mathbf{k} = k\frac{\partial\omega}{\partial k} \rightarrow \frac{\partial\omega}{\partial k} = \frac{1}{k}\mathbf{c_g}\cdot\mathbf{k},$$

when introduced into Eq. (19) yields

$$\nu(\omega) = \frac{2}{\Delta}\int_{\vartheta_1}^{\vartheta_2} \frac{k^2}{\mathbf{c_g}\cdot\mathbf{k}}\, d\vartheta.$$

The scalar product $\mathbf{c_g}\cdot\mathbf{k}$ represents the component of the group velocity along the normal to the wave front, i.e., it provides a measure of the energy flow transmitted across the wave front itself. Small values for $\mathbf{c_g}\cdot\mathbf{k}$ imply that the wave front does not effectively transport energy, suggesting an interpretation of a singularity in $\nu(\omega)$ as an average ineffectiveness of the structure in propagating wave energy and the relationship between the singularity condition for $\nu(\omega)$ and pseudo-damping.

The physical meaning of $\partial\omega/\partial k$ can also be provided following a different approach.

In general terms, a wave propagating along a two-dimensional structure can be represented as

$$w(\mathbf{x}, t) = \int\int B(k,\vartheta)e^{-j(k\cos\vartheta x_1 + k\sin\vartheta x_2 - \omega(k,\vartheta)t)}\, dk\, d\vartheta, \tag{21}$$

where the integrals are calculated over suitable intervals for k and ϑ and B is a suitable function of k and ϑ. The expression in Eq. (21) represents the superposition of plane waves with different wavelengths and proceeding with different heading angles ϑ. The wave packet associated with each heading angle can be expressed from Eq. (21) by removing the integration over the direction ϑ:

$$w_\vartheta(\mathbf{x}, \vartheta, t) = \int B(k,\vartheta)e^{-j(k\cos\vartheta x_1 + k\sin\vartheta x_2 - \omega(k,\vartheta)t)}\, dk. \tag{22}$$

Expression (22) represents the superposition of all the plane waves with different wavelengths propagating along the same direction ϑ. The Taylor expansion of $\omega(k,\omega)$ with respect to k alone shows that the wave packet travels along the direction ϑ with the speed $\partial\omega/\partial k$.

Thus, for a narrow frequency bandwidth, w can be seen as the superposition of wave packets traveling along any direction ϑ with different speeds $(\partial\omega/\partial k)(k,\vartheta)$.

Returning to the interpretation of the singularity for the modal density in light of the previous remarks, Eq. (19) reveals that the modal density can be represented as a weighted average of $1/(\partial\omega/\partial k)$. Thus, a singularity in $v(\omega)$ at some frequency ω^* can be interpreted as a frequency at which the average of the inverse of the speed $(\partial\omega/\partial k)(k,\vartheta)$ is large, i.e., an energy transport along the structure is slowed down.

Both these approaches provide a physical explanation of the pseudo-damping effect as it relates to the presence of a singularity in the modal density. Singularity in the modal density is also an indicator of the average slowness of the energy propagation in the structure.

The developments in this section help clarify the relationship of singularities in $\nu(\omega)$ with propagation of the energy in a structure, and why the tensors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ play a crucial role in producing pseudo-damping. The next section considers the case of a simply supported isotropic plate, which, with the application of a suitable compression force and a suitable elastic foundation, can also produce coefficients $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ that induce a singularity in its modal density, or, equivalently, the described effects on the group velocity leading to pseudo-damping. Section IV presents the case of a spherical shell with analogous conditions; however, in this case the tensors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ naturally produce a frequency condensation

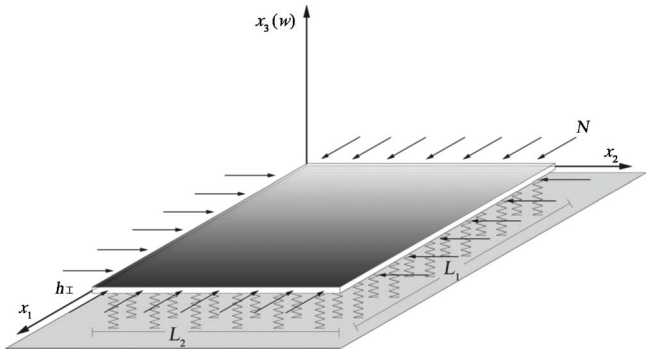


FIG. 3. (Color online) Schematic of a plate that rests on an elastic foundation with in-plane compressive forces applied at each edge.

similar to that of the isotropic plate. In that case, that shell's curvature provides the stiffness effect replacing the role of the elastic foundation on the plate response.[18] The way the energy leaves the input force location for shells or plates on elastic foundation is emphasized, leading to the apparent damping effect. An energy delocalization effect that can have a similar nature has been shown in Ref. 19. Finally, Sec. V demonstrates the case of a cylindrical shell that naturally exhibits pseudo-damping due to its anisotropic structure.

IV. PRELOADED FLAT PLATE ON AN ELASTIC FOUNDATION

Consider a flat plate on an elastic foundation of stiffness γ per unit area and subjected to a compressive preload F (see Fig. 3).

The equation of the motion (6), with the out-of-plane displacement w decoupled from the in-plane displacements, takes the simple form

$$D\nabla^4 w + F\nabla^2 w + \gamma w + \rho h \frac{\partial^2 w}{\partial t^2} = 0, \tag{23}$$

where $\nabla(\cdot) = \mathbf{e_1}(\partial/\partial x_1)(\cdot) + \mathbf{e_2}(\partial/\partial x_2)(\cdot)$, x_1, x_2 are the Cartesian coordinates in the plane of the plate, and $\mathbf{e_1}$, $\mathbf{e_2}$ are the unit vectors along the coordinate axes, D, F, ρ, and h represent the plate bending stiffness, the compressive preload, the material mass density, and the plate thickness, respectively.

A. Effect of the elastic foundation, $F=0$, $\gamma \neq 0$

Following the procedure shown in the previous section, the expression for isotropic dispersion becomes

$$\omega = \sqrt{\frac{Dk^4 + \gamma}{\rho h}} \Rightarrow k = \sqrt{\frac{\rho h \omega^2 - \gamma}{D}}, \tag{24}$$

and the group velocity has the form

$$c_g(\omega) = \frac{2D}{\sqrt{\rho h}} \frac{k^3}{\sqrt{Dk^4 + \gamma}}. \tag{25}$$

In this case, the group velocity vanishes for $k=0$, or, for $\omega = \omega^* = \sqrt{\gamma/\rho h}$, the factorization $c_g(\omega) = k(\omega)\eta(\omega)$ holds, but $\eta(\omega^*) = 0$. Thus, any frequency for which $c_g(\omega)$ vanishes is expected to produce a singularity in $\nu(\omega)$. Expressions for the modal distribution and the modal density can be obtained from Eqs. (24) and (25) as

$$N(\omega) = \frac{\pi}{4\Delta}\sqrt{\frac{\rho h \omega^2 - \gamma}{D}},$$
$$\nu(\omega) = \frac{\pi k(\omega)}{2\Delta c_g(\omega)} = \frac{\pi}{4\Delta}\frac{\rho h}{\sqrt{D}}\frac{\omega}{\sqrt{\rho h \omega^2 - \gamma}}. \tag{26}$$

Equations (26) confirm the singularity of ν at $\omega = \omega^* = \sqrt{\gamma/\rho h}$ at which the modal distribution N vanishes and the modal density has a singularity. The presence of such a singularity in the modal density suggests the possibility of a pseudo-damping effect.

The dispersion relation, modal distribution, and modal density shown, respectively, in Figs. 4(a)–4(c) demonstrate the changes affected by the elastic foundation alone, without a preload. Figures 5(a) and 5(b) compare the impulse response of a flat plate at its center ($x_1 = L_1/2$, $x_2 = L_2/2$) with ($F=0, \gamma \neq 0$) and without elastic support and preload ($F=0, \gamma=0$), respectively, clearly demonstrating the role of elastic foundation in producing pseudo-damping.

B. Effects of preload, $F \neq 0$, $\gamma \neq 0$

The effects of preload, $F \neq 0$, in generating pseudo-damping can be examined explicitly through the isotropic dispersion relationships and the group velocity:

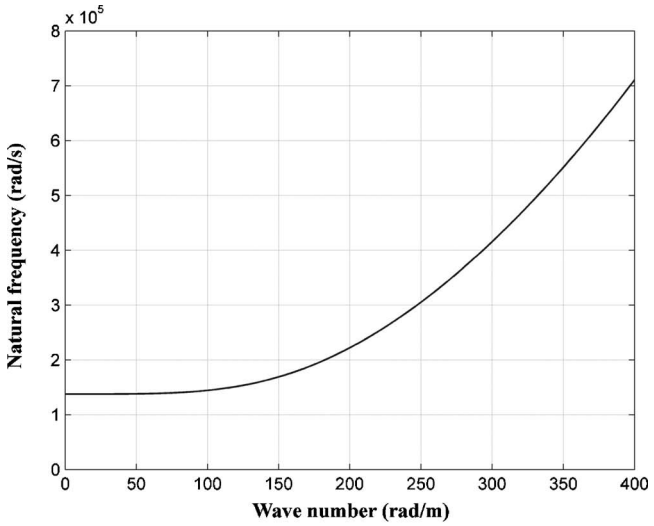
$$\omega = \sqrt{\frac{Dk^4 - Fk^2 + \gamma}{\rho h}},$$
$$k = \sqrt{\frac{\sqrt{F^2/4 - D(\gamma - \rho h \omega^2)} + F/2}{D}}, \tag{27}$$
$$c_g = \frac{1}{\sqrt{\rho h}}\frac{2Dk^3 - Fk}{\sqrt{Dk^4 - Fk^2 + \gamma}}.$$

The group velocity vanishes for $k=0$ and $k = \sqrt{F/2D}$, associated, respectively, with the frequencies $\omega_0^* = \sqrt{\gamma/\rho h}$, and $\omega_1^* = \sqrt{(D\gamma - F^2/4)/\rho h}$, $\omega_1^* < \omega_0^*$. In this case, the factorization of $c_g(\omega) = k(\omega)\eta(\omega)$ holds and $\eta(\omega_0^*) \neq 0$. Thus, the frequency ω_0^* is not expected to generate a singularity in ν. The corresponding modal distribution and the modal density expressions follow as

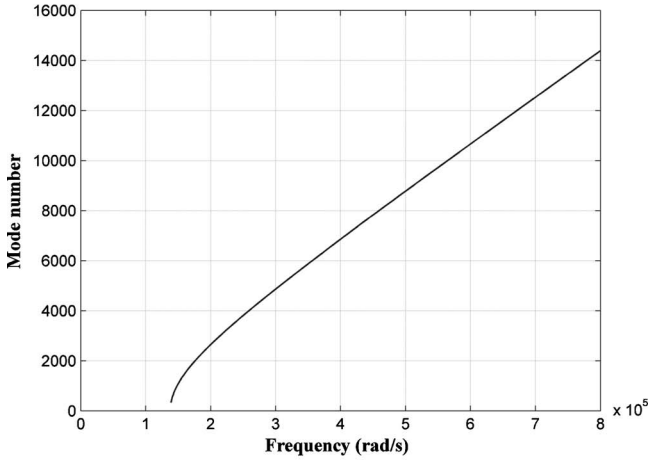
$$N(\omega) = \frac{L_1 L_2}{4\pi}\frac{\sqrt{F^2/4 - D(\gamma - \rho h \omega^2)} + F/2}{D},$$
$$\nu(\omega) = \frac{L_1 L_2 \rho h}{4\pi}\frac{\omega}{\sqrt{F^2/4 - D(\gamma - \rho h \omega^2)}}, \tag{28}$$

where the modal density $\nu(\omega)$ has only one singularity at $\omega_1^* = \sqrt{(D\gamma - F^2/4)/\rho h}$. This result, then, provides the capability to select the condensation frequency ω_1^* through a suitable combination of parameters γ and F.

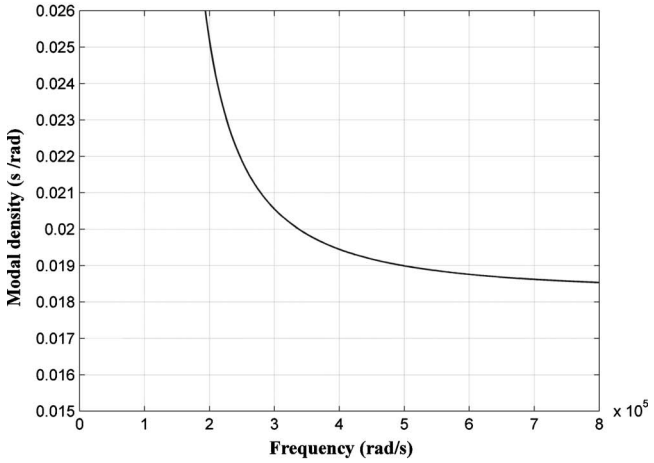
The requirement that the frequency at which condensation develops is real is satisfied by introducing an additional condition that stems from the buckling criteria: $\gamma > F^2/4D$ or $\gamma = \sigma F^2/4D$ with $\sigma > 1$.



(a)



(b)

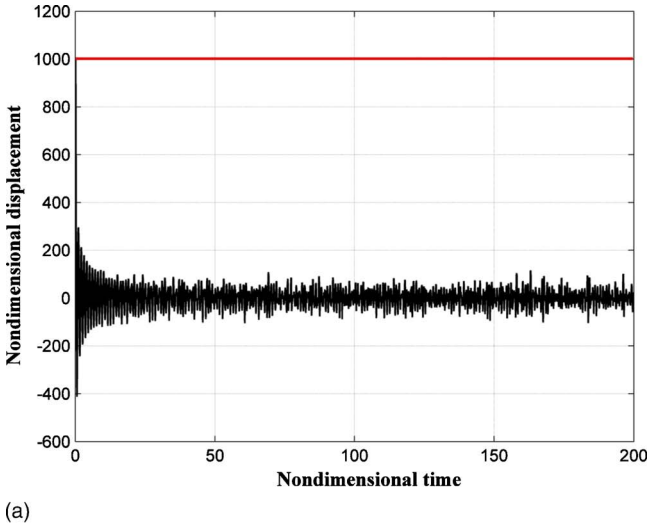


(c)

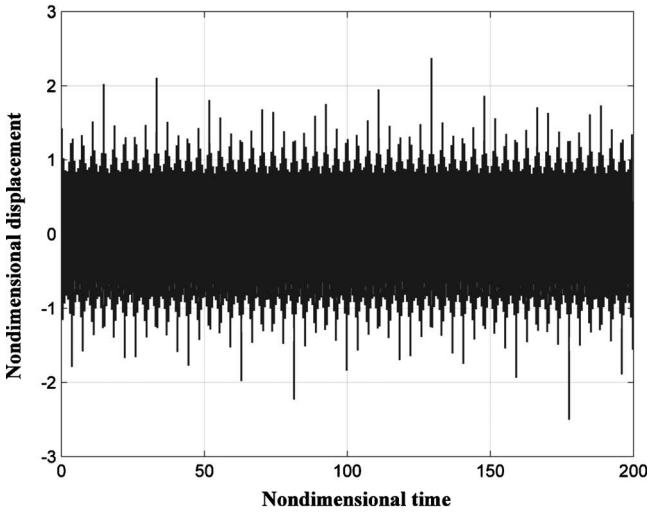
FIG. 4. For a plate resting on an elastic foundation, without the compressive in-plane forces: (a) dispersion relation, (b) modal distribution, and (c) modal density. Elastic foundation introduces a condensation region where the change in natural frequency with respect to wave number nearly vanishes.

Simulation of the impulse response of a plate is made with the same parameters of the previous case, but with a preload that corresponds to $\sigma=5$.

Results for the dispersion relation, modal distribution, and modal density are given in Figs. 6(a) and 6(b), respectively. The corresponding impulse response shown in Fig. 7 again demonstrates the presence of pseudo-damping.



(a)



(b)

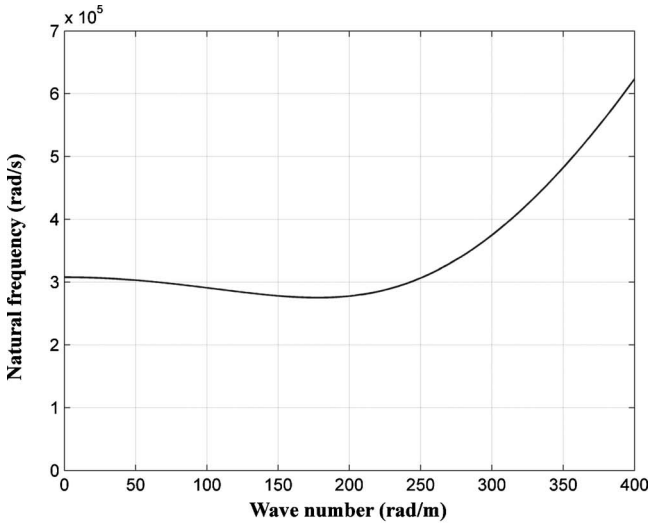
FIG. 5. (Color online) Impulse response of the simply supported plate depicted in Fig. 2 (a) with (horizontal line is initial displacement) and (b) without the elastic foundation. Presence of elastic foundation leads to pseudo-damping of the plate response.

V. SPHERICAL SHELL

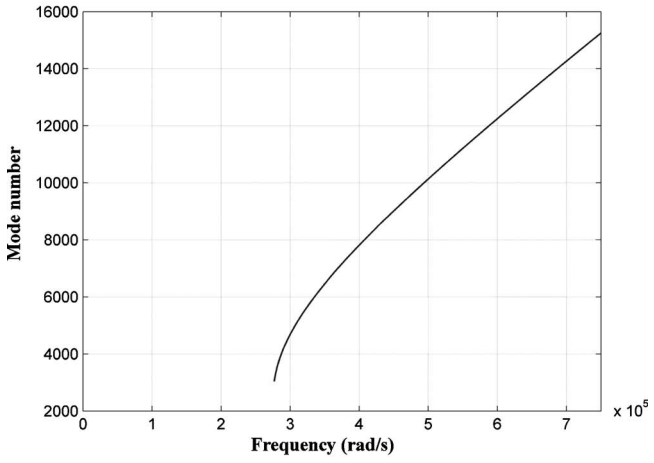
The case of the spherical shell represented in Fig. 8 is analogous to the previous case except that the role of the elastic foundation is replaced by that of shell's curvature. The equation of motion of a spherical shell has the form

$$D\nabla^8 w + \frac{Eh}{R^2}\nabla^4 w + \rho h \frac{\partial^2 \nabla^4 w}{\partial t^2} = 0, \tag{29}$$

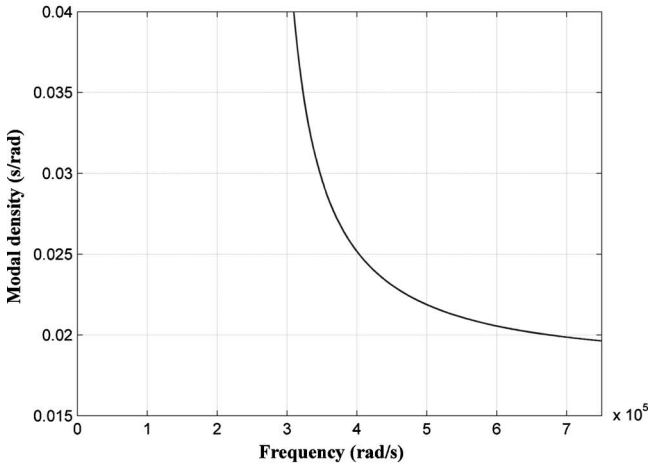
where $\nabla(\cdot)=\mathbf{e_1}(\partial/\partial x_1)(\cdot)+\mathbf{e}_2(\partial/\partial x_2)(\cdot)$, x_1 and x_2 represent spherical coordinates over the spherical surface of radius R, respectively, $\mathbf{e_1}$, $\mathbf{e_2}$ are the associated unit vectors, and E is the Young's modulus of the shell's material. Substituting for $w=Ae^{-jk_1x_1}e^{-jk_2x_2}e^{j\omega t}$, with $k_1=k\cos\vartheta$, $k_2=k\sin\vartheta$ in the previous equation, yields the dispersion relation, which becomes identical to that of the plate on an elastic foundation by sim-



(a)



(b)



(c)

FIG. 6. For a plate resting on an elastic foundation and under compressive in-plane forces: (a) dispersion relation, (b) modal distribution, and (c) modal density. Compressive forces move the frequency at which condensation develops.

ply considering $N=0$ and replacing γ with Eh/R^2. As a result, the condensation frequency for a spherical shell becomes $\omega=\omega^*=(1/R)\sqrt{E/\rho}$.

Simulation of the impulse response of a shell of radius $R=0.2$ m, using the same parameters as in the previous

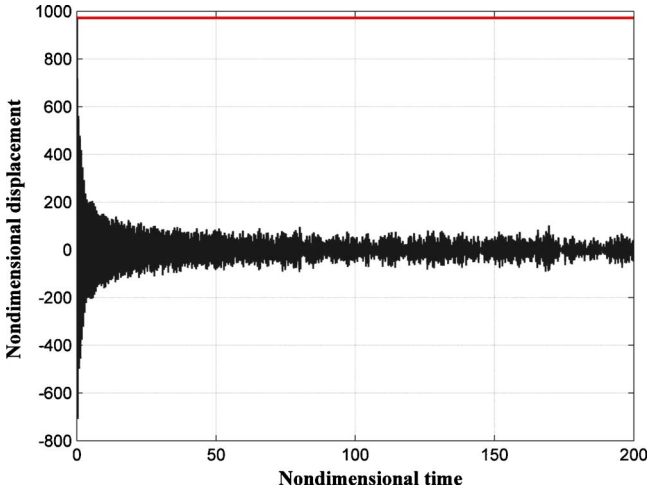


FIG. 7. (Color online) Impulse response of the plate under compressive in-plane forces (horizontal line is initial displacement).

cases, produced the results given in Figs. 9(a)–9(c) for the dispersion relation, modal distribution, modal density, respectively, with the corresponding impulse response in Fig. 10. As before, these results show inhibition of wave propagation for waves with frequencies around ω^*.

VI. CYLINDRICAL SHELL

The case of the cylindrical shell, depicted in Fig. 11, differs from that of the spherical shell because of the dispersion relation dependence on the heading angle. The general Eq. (6) can be specialized for cylindrical shells, adopting one of several theories. Instead of the Donnell theory, which has the two tangential displacements coupled with the normal displacement w, the Donnell-Vlasov approach leads, with some approximations, to the decoupled equation of motion:

$$D\nabla^8 w + \frac{Eh}{R^2}\frac{\partial^4 w}{\partial x_1^4} + \rho h \frac{\partial^2 \nabla^4 w}{\partial t^2} = 0, \tag{30}$$

where $\nabla(\cdot)=\mathbf{e_1}(\partial/\partial x_1)(\cdot)+\mathbf{e}_2(\partial/\partial x_2)(\cdot)$, $\mathbf{e_1}$, $\mathbf{e_2}$ are the associated unit vectors, and x_1 and x_2 are the axial and the circumferential coordinates, respectively. Substituting for $w=Ae^{-jk_1x_1}e^{-jk_2x_2}e^{j\omega t}$, $k_1=k\cos\vartheta$, and $k_2=k\sin\vartheta$ in the previ-

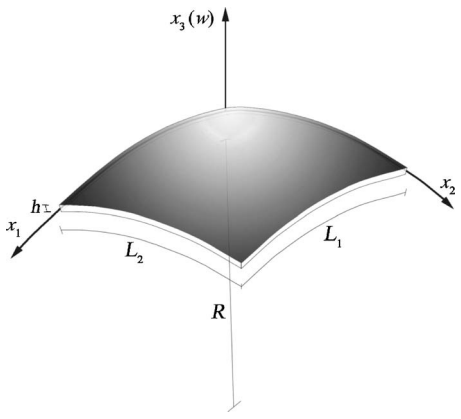
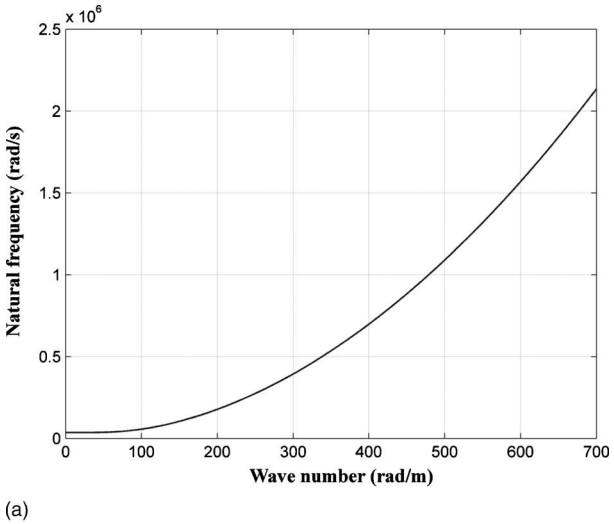
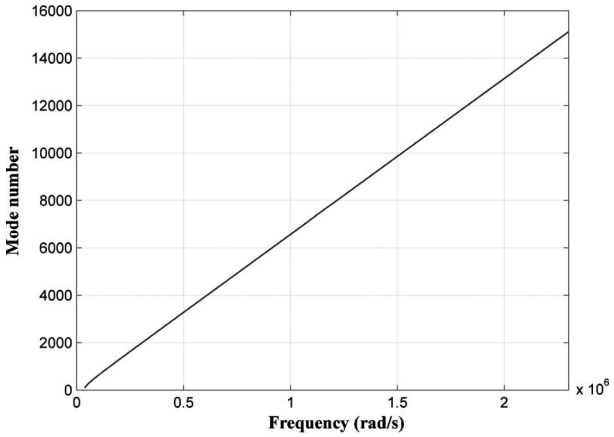


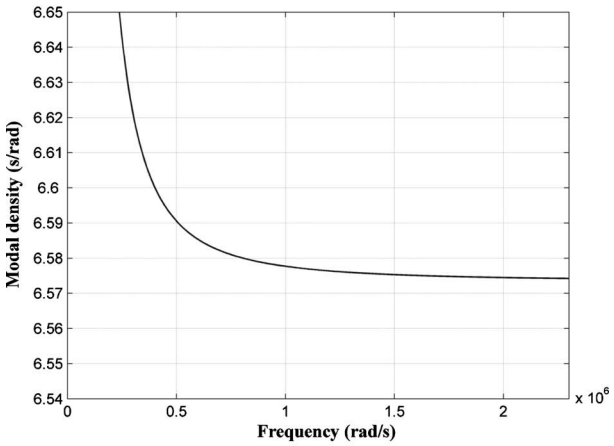
FIG. 8. Coordinate system for a spherical shell element. $E=1.4\times10^{11}$ MPa, $\nu=0.3$, $L_1=0.6$ m, $L_2=0.6$ m, thickness $h=2\times10^{-3}$ m, and radius $R=0.2$ m.



(a)



(b)



(c)

FIG. 9. For the spherical shell element in Fig. 7: (a) dispersion, (b) modal distribution, and (c) modal density.

ous equation yields the anisotropic dispersion relation:

$$\omega = \sqrt{\frac{D}{\rho h}}\sqrt{k^4 + c\cos^4\vartheta}, \quad c = \frac{Eh}{R^2 D}. \tag{31}$$

In this case, anisotropy does not allow a direct correspondence between the singularity in modal density and the group velocity. As explained in Sec. III, the group velocity component along the normal to the wave front must be considered:

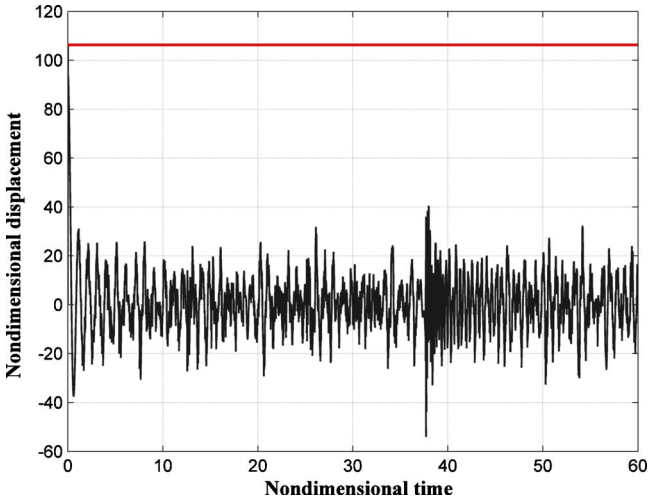


FIG. 10. (Color online) Impulse response of the spherical shell element, which includes 10 000 modes in the simulations (horizontal line is initial displacement). Position of excitation and response at $x_1 = L_1/2$, $x_2 = L_2/2$.

$$\mathbf{c_g}\cdot\mathbf{k} = \frac{\partial\omega}{\partial k} = 4\sqrt{\frac{D}{\rho h}}\frac{k^3}{\sqrt{k^4 + c\cos^4\vartheta}}. \tag{32}$$

The modal density can be expressed in terms of this last quantity:

$$\nu(\omega) = \frac{2}{\Delta}\int_{\vartheta_1}^{\vartheta_2}\frac{k^2}{\mathbf{c_g}\cdot\mathbf{k}}\,d\vartheta = \frac{1}{4\Delta}\sqrt{\frac{\rho h}{D}}\int_{\vartheta_1}^{\vartheta_2}\frac{d\vartheta}{\sqrt{1-(cD/\rho h\omega^2)\cos^4\vartheta}}. \tag{33}$$

The elliptic integral in Eq. (33) can be shown to have a singularity at the condensation frequency $\omega_0^* = (1/R)\sqrt{E/\rho}$.

Simulation of the impulse response for a shell with $R = 0.2$ m is made using the same parameters of the previous cases. The results for the dispersion relation, modal density, are given in Figs. 12(a) and 12(b), respectively, with the corresponding impulse response in Fig. 13. Again, the impulse response shows a decay and reduced amplitude over time.

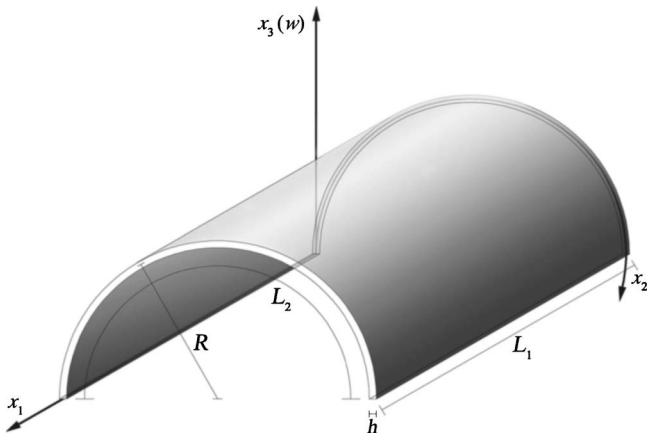
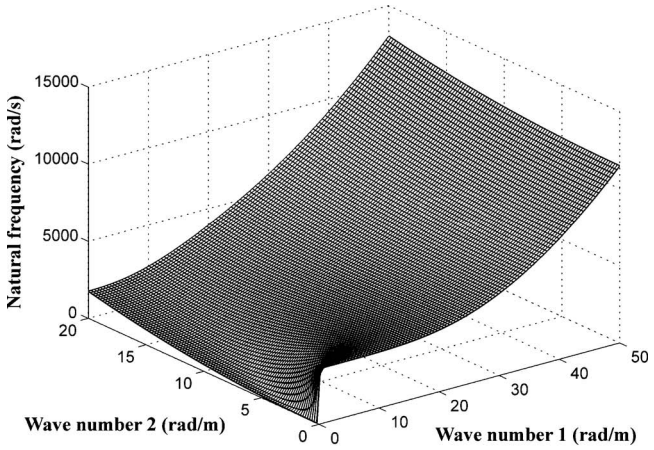
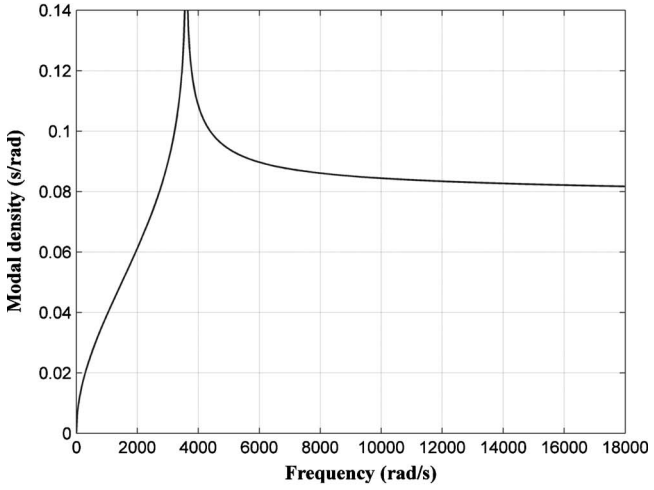


FIG. 11. Cylindrical shell element and the coordinate system that describes it. $E = 1.4\times10^{11}$ MPa, $\nu = 0.3$, with dimensions $L_1 = L_2 = 1$ m, thickness $h = 2\times10^{-3}$ m, and radius $R = 2$ m.



(a)



(b)

FIG. 12. Dispersion (a) and modal density (b) of the cylindrical shell element.

VII. CONCLUDING REMARKS

This paper introduces the concept of pseudo-damping, represented by a decaying impulse response of structures even in the absence of any dissipation. Pseudo-damping develops when the structure has one or more singularities in its modal density or, analogously, condensation points in its natural frequency distribution. In two-dimensional isotropic plates and shells, as well as in one-dimensional structures, a singularity in the modal density, or a condensation in the natural frequency distribution, corresponds to a vanishing group velocity that stops radiation, or propagation, of structure-borne waves following an impulse excitation. Associated with the singularity is a cutoff frequency, which in effect filters those waves with frequencies in its vicinity. In two-dimensional anisotropic structures, the singularity in the modal density corresponds to a singularity in the directional average of the inverse of the group velocity component along the normal to the wave front, again implying an inhibition of the energy propagation along the structure and, consequently, of the energy reflected back from the boundaries.

As in a beam, a plate can also be made to exhibit pseudo-damping by applying in-plane compressive forces while supporting it on an elastic foundation to induce a condensation point in its natural frequency distribution. Shell natural frequencies, on the other hand, naturally exhibit condensation points, which can be significant in the measurements of shell impulse response. The measured results will have decaying characteristics due to both the inherent dissipation and pseudo-damping described in this paper.

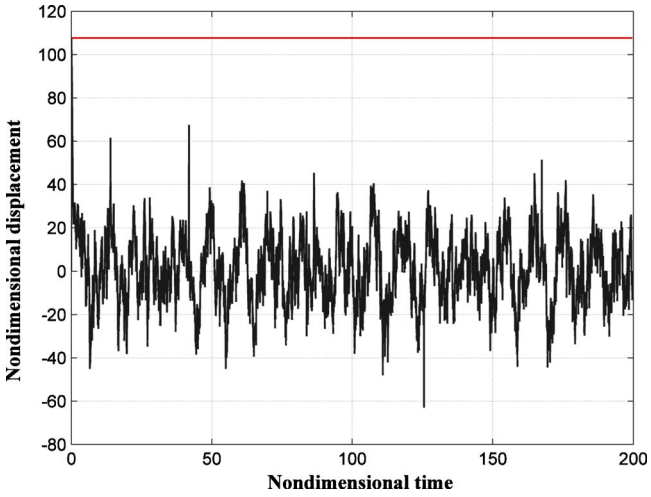


FIG. 13. (Color online) Impulse response of a cylindrical shell (horizontal line is initial displacement). Simulations considered include 10 000 modes. Position of excitation and response $x_1=L_1/2$, $x_2=L_2/2$.

ACKNOWLEDGMENT

The research described in this manuscript is based upon work while one of the authors (AA) was serving at the National Science Foundation.

[1]A. D. Pierce, V. W. Sparrow, and D. A. Russel, "Fundamental structural-acoustic idealization for structure with fuzzy internals," J. Vibr. Acoust. **117**, 339–348 (1995).
[2]M. Strasberg and D. Feit, "Vibration damping of large structures induced by attached small resonant structures," J. Acoust. Soc. Am. **99**, 335–344 (1996).
[3]G. Maidanik, "Induced damping by a nearly continuous distribution of a nearly undamped oscillators: linear analysis," J. Sound Vib. **240**, 717–731 (2001).
[4]R. J. Nagem, I. Veljkovic, and G. Sandri, "Vibration damping by a continuous distribution of undamped oscillators," J. Sound Vib. **207**, 429–434 (1997).
[5]R. L. Weaver, "The effect of an undamped finite degree of freedom 'fuzzy' substructure: numerical solution and theoretical discussion," J. Acoust. Soc. Am. **101**, 3159–3164 (1996).
[6]R. L. Weaver, "Equipartition and mean square response in large undamped structures," J. Acoust. Soc. Am. **110**, 894–903 (2001).
[7]A. Carcaterra and A. Akay, "Transient energy exchange between a primary structure and a set of oscillators: return time and apparent damping," J. Acoust. Soc. Am. **115**, 683–696 (2004).
[8]A. Akay, Z. Xu, A. Carcaterra, and I. Murat Koç, "Experiments on vibration absorption using energy sinks," J. Acoust. Soc. Am. **118** (5), 3043–3049 (2005).
[9]I. Murat Koç, A. Carcaterra, Z. Xu, and A. Akay, "Energy sinks: vibration absorption by an optimal set of undamped oscillators," J. Acoust. Soc. Am. **118** (5), 3031–3042 (2005).
[10]A. Carcaterra, A. Akay, and I. M. Koc, "Near-irreversibility and damped response of a conservative linear structure with singularity points in its modal density," J. Acoust. Soc. Am. **119** (4), 2141–2149 (2006).
[11]A. Carcaterra and A. Akay, "Theoretical foundations of apparent-damping phenomena and nearly irreversible energy exchange in linear conservative systems," J. Acoust. Soc. Am. **121**, 1971–1982 (2007).
[12]R. S. Langley, "The modal density of anisotropic structural components," J. Acoust. Soc. Am. **99** (6), 3481–3487 (1995).

[13]I. Elishakoff, "Distribution of natural frequencies in certain structural elements," J. Acoust. Soc. Am. **57** (2), 361–369 (1975).
[14]J. P. D. Wilkinson, "Modal densities of certain shallow structural elements," J. Acoust. Soc. Am. **43** (2), 245–251 (1968)
[15]X. M. Zhang, G. R. Liu, and K. Y. Lam, "Vibration analysis of thin cylindrical shells using wave propagation approach," J. Sound Vib. **239** (3), 397–403 (2001).
[16]V. L. Bersin, L. Yu Kossovich, and J. D. Kaplunov, "Synthesis of the dispersion curves for a cylindrical shell on the basis of approximate theories," J. Sound Vib. **186** (1), 37–53 (1995).
[17]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953).
[18]M. El-Mously, "A Timoshenko-beam-on-Pasternak-foundation analogy for cylindrical shells," J. Sound Vib. **261**, 635–652 (2003).
[19]G. I. Mikhasev, "Localized families of bending waves in a thin medium-length cylindrical shell under pressure," J. Sound Vib. **253** (4), 833–857 (2002).

An investigation of transmission coefficients for finite and semi-infinite coupled plate structures

Michael B. Skeen[a] and Nicole J. Kessissoglou
School of Mechanical and Manufacturing Engineering, University of New South Wales, Sydney, NSW, 2052, Australia

(Received 15 December 2006; revised 6 May 2007; accepted 10 May 2007)

This paper introduces a method for determining the transmission coefficient for finite coupled plates using an analytical waveguide model combined with a scattering matrix. In the scattering matrix method, the amplitudes of the structural waves impinging on a junction are separated into incident, reflected, and transmitted components. The energy flow due to each of these waves is obtained using a wave impedance method, which is subsequently used to determine the transmission coefficient. Transmission coefficients for semi-infinite and finite L-shaped plates are investigated for single and multiple point force excitations, and for controlled incident wave sources. It is shown that the transmission coefficients can also be calculated from details of the modal transmission coefficients and the modal composition of the energy incident on the junction. Results show that the modal transmission coefficients are largely independent of whether the plates have finite or semi-infinite boundary conditions, and are only dependent on the details of the coupling. Finally, frequency averaged transmission coefficients are compared for semi-infinite and finite structures. In the cases considered, it is found that the semi-infinite system is a good approximation for finite systems after frequency averaging, especially if the system is excited with multiple point force excitation. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2747165]

PACS number(s): 43.40.Dx [DSB] Pages: 814–822

I. INTRODUCTION

The transmission coefficient for a structural junction is defined as the ratio of energy transmitted through the junction to the energy that is incident upon it. Transmission coefficients are used extensively in vibroacoustic analysis, in particular as a parameter in energy based methods such as statistical energy analysis (SEA)[1] and energy flow analysis.[2]

To determine the transmission coefficient of a junction, it is generally required to calculate the energy flow incident on and through the junction and this has been the focus of various energy methods. Energy flow in a plate structure using the Poynting vector method has been presented by Romano *et al.*[3] and accounts for energy transmission due to flexural and in-plane motion. The wave impedance approach has been employed by many authors,[4–8] for example, Wester and Mace[4] use this method to examine the flow of energy through edge connected plates. The wave impedance method is particularly useful as the energy flow can easily be broken into directional components. It is therefore possible to separate the energy at a junction into incoming and outgoing energy, which is an important step toward determining the incident, transmitted, and reflected energy components required to determine transmission and reflection coefficients. In this paper, energy associated with individual waves will be determined using a combination of the principles of the wave impedance approach and the Poynting vector method.

Transmission coefficients for finite structures have not been widely studied due to the general assumption that, at least in the frequency average, the transmission coefficients for semi-infinite and finite structures are equivalent.[9] It has been common practice to use the transmission coefficient derived for a semi-infinite system as an approximation for that of a finite system.[1,5,9,10] Simplified transmission coefficient expressions have been given for L-, T-, and X-shaped plates based on the ratio of the plate thickness by Cremer *et al.*[5] using a wave impedance based approach. Langley and Heron[6] developed expressions for transmission coefficients for an arbitrary number of semi-infinite plates connected at various angles at a beam junction. Langley and Shorter[7] used a wave impedance approach to develop transmission coefficients in point connected structures such as beam-plate connections. Park *et al.*[8] examined "semifinite" structures where either the source or receiver plate was finite. They demonstrated that the finite boundary conditions significantly altered the transmission coefficient from that predicted for the semi-infinite structure, though there were similarities in the transmission coefficients predicted for the semi-infinite and semifinite systems. Another reason for using the transmission coefficient derived from a semi-infinite structure as an approximation to that of a finite system is due to the difficulty involved in separating the transmitted and reflected wave components from the outgoing waves leaving the junction in a reverberant structure. Wester and Mace[4] used a scattering matrix derived from a wave approach to determine the outgoing waves produced by incoming waves incident on a junction of a coupled plate structure and used a combination of these coupled junctions to model an entire plate structure. Shorter and Langley[11] describe a method of separating a reverberant field into components due to the direct field and due to subsequent reflections of the direct field from the boundaries. However, to the authors' knowledge, there has

[a]Electronic mail: michaelskeen78@hotmail.com

 0001-4966/2007/122(2)/814/9/$23.00

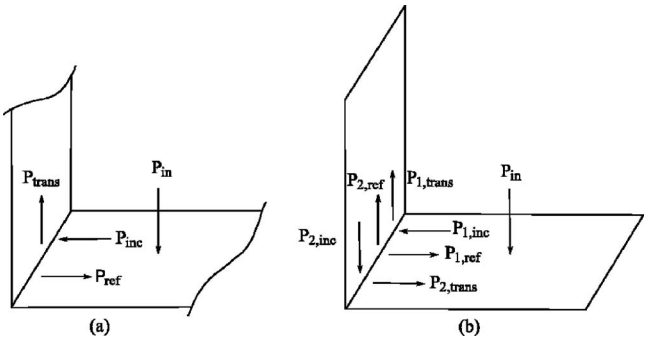


FIG. 1. Power incident, transmitted and reflected on (a) semi-infinite and (b) finite plates.

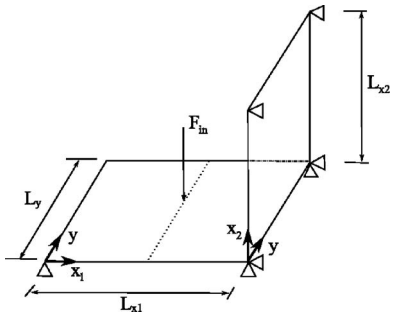


FIG. 2. L-shaped plate dimensions and coordinate system.

been little or no work on separating the reverberant field at the junction of finite coupled structures into components due to incident, reflected, and transmitted waves, from which the transmission and reflection coefficients can be determined.

In this paper, a method for determining the transmission coefficient of finite coupled plates is presented. The scattering matrix method has been employed as a means of separating the incoming and outgoing waves at a coupled finite plate junction. A wave impedance method is then used to determine the energy due to each wave component, from which the energy flow associated with incident, reflected, and transmitted waves can be obtained. Transmission coefficients for semi-infinite and finite L-shaped plates are investigated for both single and multiple point force excitation. A method of calculating the finite transmission coefficient from details of the semi-infinite transmission coefficient and the modal composition of the energy incident on the junction is also presented. Results show that the modal transmission coefficients are largely independent of whether the plates have finite or semi-infinite boundary conditions, and are only dependent on the details of the coupling. Finally, frequency averaged transmission coefficients are presented for semi-infinite and finite systems. In the cases considered, it is found that the semi-infinite system is a good approximation for finite systems after frequency averaging, especially if the system is excited with multiple point force excitation.

II. THEORY

A. Transmission coefficients

The transmission coefficient τ for a junction is defined as the ratio of transmitted power P_{trans} to the incident power P_{inc} and is given by[5]

$$\tau = \frac{P_{\text{trans}}}{P_{\text{inc}}}. \tag{1}$$

Similarly, the reflection coefficient μ and loss coefficients λ are, respectively, defined by

$$\mu = \frac{P_{\text{ref}}}{P_{\text{inc}}}, \tag{2}$$

$$\lambda = 1 - \mu - \tau, \tag{3}$$

where P_{ref} is the reflected power. Figures 1(a) and 1(b) respectively show the waves generated by an external force for a semi-infinite and finite L-shaped plate system. P_{in} is the input power due to the external force. With a single excitation, semi-infinite structures only generate one set of active incident, transmitted, and reflected waves, as shown in Fig. 1(a). In the case of a finite structure [Fig. 1(b)], a reverberant field is generated by reflections at the finite plate edges. For finite structures, it is important to determine what proportions of the outgoing waves from the junction are due to transmitted and reflected waves. To evaluate the transmission coefficient, it is necessary to separate the energy flow leaving the junction into components due to reflection and transmission, respectively. In this paper, this is achieved using a scattering matrix method, which is discussed later in further detail.

A simplified method of calculating the transmission coefficient for L-shaped structures for semi-infinite plate and beam structures was presented by Cremer *et al.*[5] This method is derived from the moment and force equilibrium and displacement and slope continuity requirement across a junction. An approximation for the transmission coefficient for plates with homogeneous properties is given by

$$\tau_{12} = \frac{2}{(\sigma^{5/2} + \sigma^{-5/2})^2}, \tag{4}$$

where $\sigma = h_1/h_2$ is the ratio of the thickness of the two plates. For plates of the same thickness ($\sigma = 1$), the transmission coefficient is $\tau_{12} = 0.5$.

B. Waveguide method

An L-shaped plate system as shown in Fig. 2 is examined, which is simply supported along two parallel edges corresponding to $y = 0$ and $y = L_y$, and free at the other two edges corresponding to $x_1 = 0$ and $x_2 = L_{x2}$. The junction of the two plates corresponds to $x_1 = L_{x1}$ and $x_2 = 0$. The two plates are of the same material and thickness. An external point force excitation of amplitude F_{in} at a location $(x_{\text{in}}, y_{\text{in}})$ is applied to plate 1 and is described by a Dirac delta function. The equation of motion for the flexural motion of a plate w is given by[12]

$$D\nabla^4 w + \rho h \frac{\partial^2 w}{\partial t^2} = F_{\text{in}} \delta(x - x_{\text{in}}) \delta(y - y_{\text{in}}) e^{j\omega t}, \tag{5}$$

where $\nabla^4 = \nabla^2 \nabla^2$ and $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$ is the Laplace operator. $D = \hat{E}h^3/12(1-\nu^2)$ is the plate flexural rigidity, ν is Poisson's ratio, ρ is the density, and h is the plate thickness.

Damping is included using a complex Young's modulus, $\hat{E}=E(1+j\eta)$, where E is the Young's modulus and η is the damping loss factor. Using the analytical waveguide method, the flexural displacement of the plate with two parallel simply supported edges results in a modal solution in the y direction and a traveling wave solution in the x direction. A general solution for the flexural displacement is given by[13]

$$w_p(x,y,t)=\sum_{m=1}^{\infty}(A_{p,1,m}e^{-jk_x x}+A_{p,2,m}e^{jk_x x}+A_{p,3,m}e^{-k_n x}+A_{p,4,m}e^{k_n x})\sin(k_y y)e^{j\omega t}, \tag{6}$$

where $A_{p,q,m}$ are the wave displacement amplitudes, and the subscript p, q, and m refer to the plate number, wave index, and mode number, respectively. The first two waves of amplitude $A_{p,1,m}$ and $A_{p,2,m}$ represent traveling waves in the positive and negative x directions, respectively. The last two waves are, respectively, evanescent waves in the positive and negative x directions. $k_y=m\pi/L_y$ is the wave number in the y direction. $k_x=\sqrt{k_p^2-k_y^2}$ and $k_n=\sqrt{k_p^2+k_y^2}$ are the wave numbers along the x direction for the propagating and evanescent waves, where $k_p=\sqrt[4]{\omega^2\rho h/D}$ is the flexural wave number of the plate. The frequency of transition between the evanescent and propagating waves is known as the modal cut-on frequency.[8] The displacement only has to be considered above the cut-on frequency for a particular mode, limiting the infinite summation in Eq. (6) The modal cut-on frequency is defined to be when $k_y=k_p$ and is given by[8]

$$f_{co}=\frac{\pi}{2}\left(\frac{m}{L_y}\right)^2\sqrt{\frac{D}{\rho h}}. \tag{7}$$

The wave displacement amplitudes in each section of the L-shaped plate are evaluated for each mode number using boundary conditions and coupling equations which are developed from displacement and slope continuity, and moment and force equilibrium equations. The internal forces and moments acting on the plate are given by[12]

$$M_x=-D\left(\frac{\partial^2 w}{\partial x^2}+\nu\frac{\partial^2 w}{\partial y^2}\right), \tag{8}$$

$$Q_x=-D\left(\frac{\partial^3 w}{\partial x^3}+\frac{\partial^3 w}{\partial x\partial y^2}\right), \tag{9}$$

$$M_{xy}=-D(1-\nu)\frac{\partial^2 w}{\partial x\partial y}, \tag{10}$$

where M_x is the bending moment, Q_x is the shear force due to bending, and M_{xy} is the twisting moment. The net vertical shear force is obtained by $V_x=Q_x+\partial M_{xy}/\partial y$.

The boundary conditions at the free edges corresponding to $x_1=0$ and $x_2=L_{x2}$ result in zero bending moment and net vertical shear force.[12] Due to the external force, there are four coupling equations at $x=x_{in}$ corresponding to continuity of displacement and slope, and equilibrium of the moments and shear forces.[13] The coupling equations at the junction of an L-shaped plate are well established[5,14] but details are provided here for completeness and because of their underlying importance to both the Poynting vector energy flow and scattering matrix methods. The coupling equations at the L-shaped junction are also given by continuity of displacement and slope, and equilibrium of moments and shear forces. Continuity of bending moment from plate 1 to 2 plate results in

$$M_{x1}(L_{x1},y)=M_{x2}(0,y). \tag{11}$$

Continuity of slope yields

$$\frac{\partial w_1(L_{x1},y)}{\partial x}=\frac{\partial w_2(0,y)}{\partial x}. \tag{12}$$

At the junction, the transformation of flexural to in-plane energy is treated as a loss. The shear force in plate 1 induces a longitudinal force in plate 2. Similarly, the shear force in plate 2 generates a longitudinal force in plate 1. The amplitude of the flexural displacement at the junction in plate 1 is equal to the amplitude of the longitudinal waves in plate 2, and similarly, the amplitude of the flexural displacement in plate 2 is equal to the amplitude of the longitudinal waves in plate 1. The force and displacement continuity equations can be resolved in terms of the flexural displacement in plates 1 and 2, and are given by[14]

$$\frac{\partial^3 w_1(L_{x1},y)}{\partial x^3}+(2-\nu)\frac{\partial^3 w_1(L_{x1},y)}{\partial x\partial y^2}=\frac{jk_p^4}{k_L}w_1(L_{x1},y), \tag{13}$$

$$\frac{\partial^3 w_2(0,y)}{\partial x^3}+(2-\nu)\frac{\partial^3 w_2(0,y)}{\partial x\partial y^2}=\frac{-jk_p^4}{k_L}w_2(0,y), \tag{14}$$

where $k_L=\sqrt{\rho\omega^2(1-\nu^2)/E}$ is the in-plane wave number. Due to the external point force located at (x_{in},y_{in}), the L-shaped plate is separated into three sections $(0\leqslant x_1\leqslant x_{in})$, $(x_{in}\leqslant x_1\leqslant L_{x1})$ and $(0\leqslant x_2\leqslant L_{x2})$. This results in a total of 12 unknown displacement wave amplitudes to be determined. Substituting the general solution for the plate flexural displacement given by Eq. (6) into the equations for the boundary conditions at the free edge, and continuity equations at the force location and junction, the displacement amplitudes can be determined. This is achieved by arranging the 12 equations in matrix form given by $\boldsymbol{\alpha}\mathbf{A}=\mathbf{F}$, where the matrix $\boldsymbol{\alpha}$ contains details of the boundary and continuity equations, the vector $\mathbf{A}$ contains the unknown wave displacement amplitudes, and the vector $\mathbf{F}$ contains details of the external force applied to the system. The wave displacement amplitudes can be found by $\mathbf{A}=\boldsymbol{\alpha}^{-1}\mathbf{F}$.

C. Modeling of the energy flow

There are a number of ways that energy flow can be evaluated from a waveguide solution including the Poynting[3] and wave impedance[4] methods. Using the Poynting method, the time averaged net energy flow in the x direction per unit width of plate is given by[3,14]

$$P=\frac{1}{L_y}\int_0^{L_y}\mathrm{Re}\left(\dot{w}^*Q_x-\frac{\partial\dot{w}^*}{\partial x}M_x-\frac{\partial\dot{w}^*}{\partial y}M_{xy}\right)\partial y, \tag{15}$$

where $\dot{w}$ denotes derivative of w with respect to time and the asterisk (*) denotes the complex conjugate. When Eq. (15) is fully expanded there are cross-power terms involving inter-

actions between the positive and negative traveling and evanescent waves. These cross-power terms are commonly neglected in wave impedance based methods on the assumption that traveling waves alone are the dominant mode of energy transmission.[4,6] Using this assumption, the energy flow that is associated with a particular wave can be determined from Eq. (15) by substituting only the component of the displacement associated with that wave. The displacement $w_{p,q}$ associated with a traveling wave of amplitude $A_{p,q,m}$ is given by

$$w_{p,q} = \sum_{m=1}^{\infty} A_{p,q,m} \sin(k_y y) e^{\pm jk_x x} e^{j\omega t}. \tag{16}$$

Using this approach, the power flow is broken into positive and negative components associated with the positive and negative traveling waves. This method can be used to obtain the energy flow associated with the incoming and outgoing waves at a junction. The energy flow due to the reflected and transmitted waves at a junction can now be determined using a scattering matrix method which is described in what follows.

D. Scattering matrix method

The scattering matrix is used to separate the outgoing waves generated at a junction in a reverberant system into components due to reflected and transmitted waves. The scattering matrix is developed by rearranging the coupling equations for the plate junction such that the outgoing wave amplitudes are calculated for a given set of incoming wave amplitudes. Hence, the coupling equations can be expressed as

$$\mathbf{a}\mathbf{A}_{\text{out}} = \mathbf{b}\mathbf{A}_{\text{in}}, \tag{17}$$

where $\mathbf{A}_{\text{in}} = \{A_{1,1,m}\ A_{1,3,m}\ A_{2,2,m}\ A_{2,4,m}\}^T$ is a vector containing the amplitudes of waves incoming to the junction, and $\mathbf{A}_{\text{out}} = \{A_{1,2,m}\ A_{1,4,m}\ A_{2,1,m}\ A_{2,3,m}\}^T$ contains the amplitudes of waves produced at and leaving the junction. Both **a** and **b** are matrices simply derived by rearranging the coupling equations in the matrix $\boldsymbol{\alpha}$ and are given by

$$\mathbf{a} = \begin{bmatrix} -jk_x e^{jk_x L_{x1}} & -k_n e^{k_n L_{x1}} & -jk_x & -k_n \\ (k_x^2 + \nu k_y^2) e^{jk_x L_{x1}} & -(k_n^2 - \nu k_y^2) e^{k_n L_{x1}} & -(k_x^2 + \nu k_y^2) & k_n^2 - \nu k_y^2 \\ (-\beta + \psi) e^{jk_x L_{x1}} & (\chi + \psi) e^{k_n L_{x1}} & 0 & 0 \\ 0 & 0 & \beta - \psi & -\chi - \psi \end{bmatrix}, \tag{18}$$

$$\mathbf{b} = \begin{bmatrix} -jk_x e^{-jk_x L_{x2}} & -k_n e^{-k_n L_{x2}} & -jk_x & -k_n \\ -(k_x^2 + \nu k_y^2) e^{-jk_x L_{x2}} & (k_n^2 - \nu k_y^2) e^{-k_n L_{x2}} & k_x^2 + \nu k_y^2 & -k_n^2 + \nu k_y^2 \\ (-\beta - \psi) e^{-jk_x L_{x2}} & (\chi - \psi) e^{-k_n L_{x2}} & 0 & 0 \\ 0 & 0 & \beta + \psi & -\chi + \psi \end{bmatrix}, \tag{19}$$

where $\beta = jk_x(k_x^2 + (2-\nu)k_y^2)$, $\chi = k_n(k_n^2 - (2-\nu)k_y^2)$, and $\psi = jk_p^4/k_L$. The scattering matrix **T** is given by

$$\mathbf{T} = \mathbf{a}^{-1}\mathbf{b} \tag{20}$$

and hence

$$\mathbf{A}_{\text{out}} = \mathbf{T}\mathbf{A}_{\text{in}}. \tag{21}$$

The first step in separating the reflected and transmitted wave components in a finite structure is to evaluate the wave displacement amplitudes using the waveguide method. The incoming waves (both traveling and evanescent) that are incident on one side of the junction are then used as the input to the scattering matrix with the incoming wave from the opposite direction being set to zero. The outgoing wave amplitudes are then calculated and these represent the transmitted and reflected waves. Using Eqs. (15) and (16), the energy flow associated with the transmitted and reflected waves can be determined. Referring to Fig. 1(b), the transmission coefficient for transmission from plates 1 to 2, $\tau_{12} = P_{1,\text{trans}}/P_{1,\text{inc}}$, is then be obtained. It should be noted that the incident energy on the junction from plate 1, $P_{1,\text{inc}}$, is due to waves generated directly by the input force and the waves that are reflected back to the junction by the finite edges on plate 1. The process is repeated for the incoming waves on the opposite side of the junction to determine the transmission coefficient for transmission from plates 2 to 1, $\tau_{21} = P_{2,\text{trans}}/P_{2,\text{inc}}$. In this case, the incident energy on the junction from plate 2, $P_{2,\text{inc}}$, is only due to the reverberant field in plate 2.

E. Modal transmission coefficient relationship

After the incident and transmitted energies have been determined using the Poynting method, the total transmission coefficient τ can be evaluated [Eq. (1)]. This total transmission coefficient accounts for the energy transmitted by all the active modes. Alternatively, the modal transmission coefficient τ_m is calculated by only considering the displacement associated with a single mode m, that is

$$\tau_m = \frac{P_{\text{trans},m}}{P_{\text{inc},m}}. \tag{22}$$

The total transmission coefficient can also be derived in terms of modal power components where the transmitted and

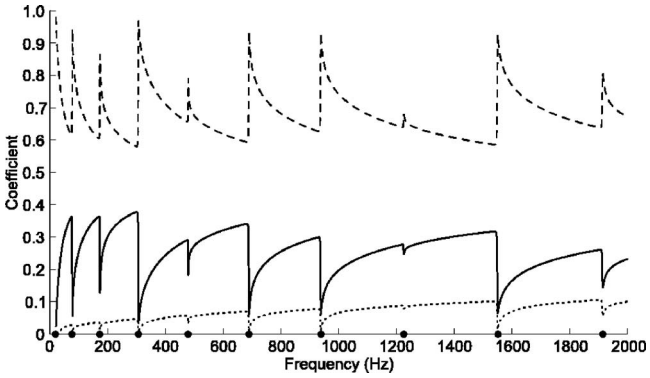


FIG. 3. The total transmission (—), reflection (---), and loss (···) coefficients for two semi-infinite plates coupled at right angles with excitation at $(x_1,y)=(742,309)$. Modal cut-on frequencies (●) are also shown.

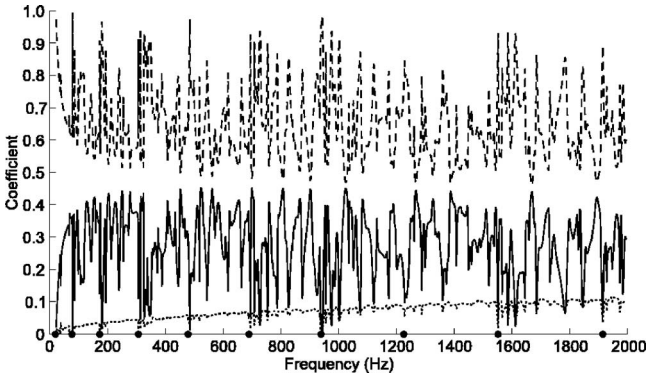


FIG. 4. The total transmission (—), reflection (---), and loss (···) coefficients for waves traveling from plate 1 to 2 for two finite plates coupled at right angles with excitation at $(x_1,y)=(742,309)$. Modal cut-on frequencies (●) are also shown.

incident power terms in Eq. (1) are given in terms of their modal summation. The total transmission efficiency in this form is given by

$$\tau=\frac{\sum_m P_{\text{trans},m}}{\sum_n P_{\text{inc},n}}=\sum_m\left(\frac{P_{\text{trans},m}}{\sum_n P_{\text{inc},n}}\right). \tag{23}$$

Multiplying both the numerator and denominator of the expression inside the brackets in Eq. (23) by the incident modal power $P_{\text{inc},m}$ and using Eq. (22), an expression for the total transmission coefficient can be obtained in terms of a summation of the modal transmission coefficient multiplied by the proportion of incident modal power,

$$\tau=\sum_m\left(\frac{P_{\text{trans},m}}{\sum_n P_{\text{inc},n}}\frac{P_{\text{inc},m}}{P_{\text{inc},m}}\right)=\sum_m\left(\tau_m\frac{P_{\text{inc},m}}{\sum_n P_{\text{inc},n}}\right). \tag{24}$$

The relationship established by Eq. (24) is investigated analytically in the proceeding section.

III. RESULTS

Results are presented for an L-shaped plate shown in Fig. 2 with dimensions of $L_{x1}=1200$ mm, $L_{x2}=600$ mm, and $L_y=500$ mm. Both plates have a thickness of $h=2$ mm. The plates are of aluminum with Young's modulus $E=71$ MPa, density $\rho=2800$ kg m^{-3}, Poisson's ratio $\nu=0.3$, and damping loss factor $\eta=0.001$.

A. Semi-infinite and finite transmission coefficients

The transmission coefficient is initially investigated for an L-shaped plate subject to a single harmonic point force excitation at a location $(x_1,y)=(0.618L_x,0.618L_y)=(742,309)$ (all locations given are in millimeters), at which a large number of modes are excited. Figure 3 shows the total transmission, reflection, and loss coefficients for the semi-infinite L-shaped plate. Coefficients are presented for frequencies above the first modal cut-on frequency. Below this frequency, all waves are evanescent and the analysis is no longer valid as it has been assumed that the traveling waves are the primary transmission path. The coefficient results for the semi-infinite plate show distinct frequencies at which the total transmission coefficient is dramatically reduced. These frequencies are related to the modal cut-on frequencies (also shown in Fig. 3), showing that as a mode becomes active the transmission coefficient decreases. It can also be seen in Fig. 3 that the loss coefficient slightly increases with frequency and appears to follow a square root trend. This behavior was predicted by Cremer *et al.*[5] for losses due to the transformation of flexural to in-plane waves at the L-plate junction, and confirms that the contribution of in-plane motion increases with frequency.[14]

In Fig. 4, the transmission, reflection, and loss coefficients are shown for the finite L-shaped plate generated using the scattering matrix method for incident waves impinging on the L junction from plate 1. The transmission and reflection coefficients for the finite system show much greater variation with frequency. Similarities between the results for the finite and semi-infinite structures can be found. A dramatic reduction in the transmission coefficient occurs at the modal cut-on frequencies. The loss coefficient also follows the same general trend as the semi-infinite results.

The transmission coefficients for the semi-infinite and finite L-shaped plates are compared in Fig. 5. The transmission coefficient as predicted by Cremer *et al.*[5] is also shown. In the low frequency region between the first and second modal cut-on frequencies, the transmission coefficients for

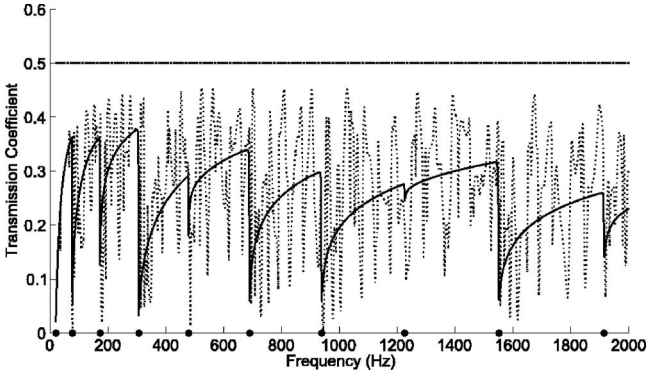


FIG. 5. Comparison of the total transmission coefficients for semi-infinite (—) and finite (···) systems excited at $(x_1,y)=(742,309)$. Also shown is the transmission coefficient predicted by Cremer *et al.*—Ref. 5 (---) and the modal cut-on frequencies (●).

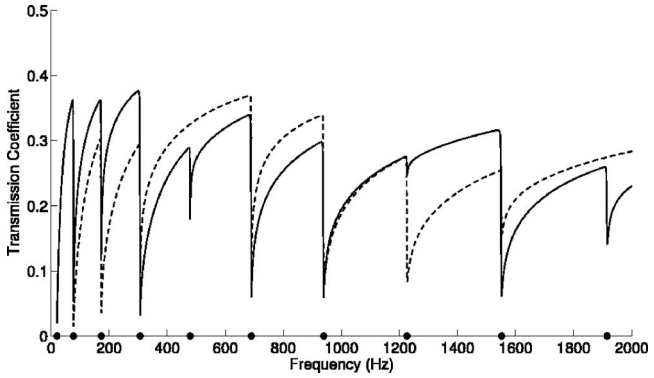


FIG. 6. Comparison of the total transmission coefficients for two semi-infinite systems excited at $(x_1,y)=(742,309)$ (—) and $(x_1,y)=(100,100)$ (---). Also shown are the modal cut-on frequencies (●).

both the finite and semi-infinite cases, are nearly identical. Above the second modal cut-on frequency, the transmission coefficient for the semi-infinite system appears to occur centrally within the bounds of the finite results. The transmission coefficient predicted by Cremer *et al.*[5] for an infinite L-shaped plate is notably higher than the transmission coefficients for the semi-infinite and finite systems predicted by the methods presented in this paper.

Figure 6 compares the transmission coefficients for the semi-infinite system excited at two different force locations corresponding to $(x_1,y)=(742,309)$ and (100,100). Figure 6 shows that the amplitude of the transmission coefficient varies with input location. This result is expected as Eq. (24) predicts that the transmission coefficient depends on the modal composition of the energy incident on the plate, which in turn is highly dependent on the input force location. For the semi-infinite L-shaped plate of width 500 mm between its simply supported edges, a force location of $(x_1,y)=(100,100)$ will not excite the fifth mode. The absence of this mode increases the total transmission coefficient in the frequency region between the fifth and sixth cut-on frequencies.

To further investigate the effect of the excitation location, Fig. 7 shows the variation of the transmission coefficient for a semi-infinite system with changes in the force location in the y direction when the input is fixed in the x direction at $x_1=0.618L_x$ ($x_1=742$ mm). Similarly, Fig. 8 presents the variation of the transmission coefficient with changes in the force location in the x direction when fixed in the y direction at $y=0.618L_y$ ($y=309$ mm). Figure 8 shows little to no variation in the transmission coefficient along the x direction except at locations close to the junction. Hence, the results presented in Fig. 7 are indicative of the transmission coefficient at all excitation locations on the plate except in regions close to the junction. Figure 7 distinctly shows that the transmission coefficient becomes a minimum value at each cut-on frequency, as shown in Fig. 3.

In order to examine the effect of incident modal energy not dominated by a single mode, on the transmission coefficient of coupled structures, two other excitation types are also investigated. Figure 9 shows the transmission coefficient for a plate subject to random point force excitation, where the transmission coefficient is obtained by exciting the plate at 100 different locations and then taking the total transmission coefficient to be the average value calculated over all the excitation locations considered.[8] Examining the transmission coefficients for the semi-infinite structure shows that for multiple force excitation, all of the modes are active and the dominance of each of these modes appears to be very similar, that is, the peak values of the transmission coefficient are nearly constant. The transmission coefficient for the finite coupled plates still varies significantly with frequency.

Figure 10 presents the transmission coefficients for the semi-infinite and finite systems subject to an incident wave excitation. In this case the plate is not excited by a point force, but instead the incident wave displacement amplitudes are set to unity for each mode and the waveguide model is then solved to determine the remaining wave displacement amplitudes.[8] Figure 10 shows that under incident wave excitation, the transmission coefficients for the semi-infinite and finite systems are nearly identical. This result is not surprising since the modal composition of waves incident on the junction have been constrained such that they are identical for the finite and semi-infinite cases. The transmission coefficients approach a constant value of approximately 0.34, which is well below the transmission coefficient of 0.5 predicted by Cremer *et al.*[5] for an infinite L-shaped plate.

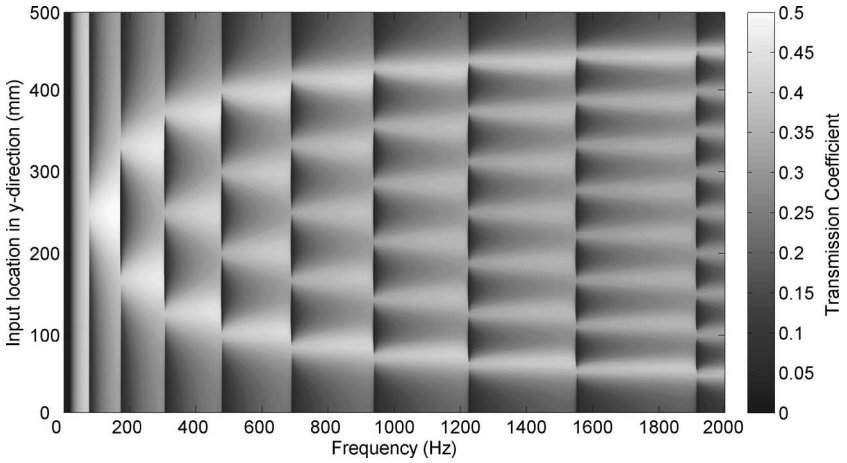


FIG. 7. The variation of total transmission coefficient with input location in the y direction for a semi-infinite system. The input location is fixed in the x direction at $x_1=742$ mm.

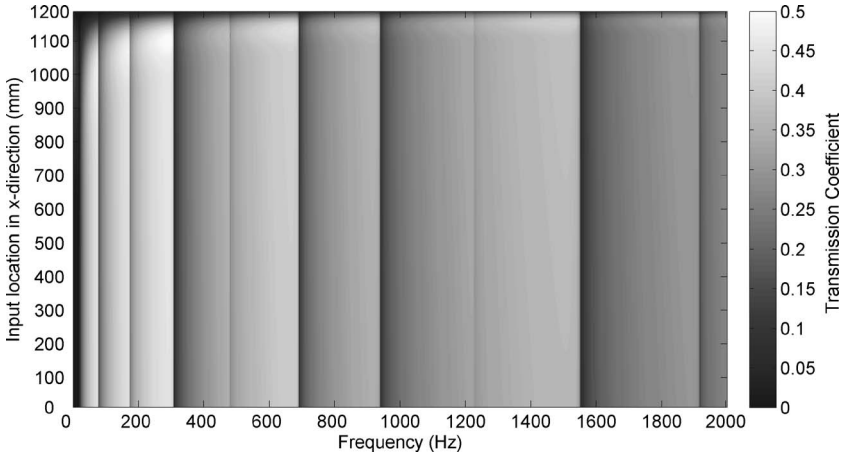


FIG. 8. The variation of total transmission coefficient input location in the x direction for a semi-infinite system. The input location is fixed in the y direction at $y=309$ mm.

B. Modal transmission coefficient relationship

In the previous section, it was briefly discussed that the total transmission coefficient is dependent on the modal composition of the incident energy. Equation (24) predicts that the total transmission efficiency is made up of modal transmission coefficients, with the dominance of each modal transmission coefficient determined by the proportion of modal energy incident on the junction. Figure 11 shows the total and individual modal transmission coefficients for the semi-infinite system, for an excitation location at $(x_1,y)=(100,100)$. Each modal transmission coefficient commences at its cut-on frequency at zero, and asymptotically approaches a constant value which never exceeds the value of the preceding modal transmission coefficient. In the low frequency region between the first and second modal cut-on frequencies, the first modal and total transmission coefficients are equal. Beyond the second cut-on frequency, the total transmission coefficient never exceeds the value of the transmission coefficient for the first mode.

The total and individual modal transmission coefficients for the finite system are shown in Fig. 12, for a force located at $(x_1,y)=(100,100)$. Similar to the results for the semi-infinite L plate, the total transmission coefficient for the finite plate is the same as the modal transmission coefficient for the first mode until the second cut-on frequency is reached, and the maximum value of the total transmission coefficient never exceeds the value of the modal transmission coefficient for the first mode. For the finite plate, whilst the maximum value of the total transmission coefficient is often close to the value of the modal transmission coefficient for the first mode, the minimum value of the total transmission coefficient roughly follows the values of the modal transmission coefficients as each subsequent mode becomes active. In Fig. 12 it is particularly evident that at a force location of $(x_1,y)=(100,100)$ (for a plate of width 500 mm between its simply supported boundaries), the fifth mode has not been excited which affects the general trend of the minimum value of the total transmission coefficient between the fifth and sixth cut-on frequencies.

It is of interest to note that the modal transmission coefficients are the same for the finite and semi-infinite structures, and are hence independent of the semi-infinite and finite boundaries. Figures 11 and 12 also compare the total transmission coefficients calculated using Eq. (1), and using the modal transmission coefficients and proportion of corresponding modal incident energies [Eq. (24)]. As the modal transmission coefficients for both finite and semi-infinite structures are the same, the total transmission coefficient for the finite structure (Fig. 12) was calculated using the modal transmission coefficient for the semi-infinite system, and hence use of the scattering matrix method was not required.

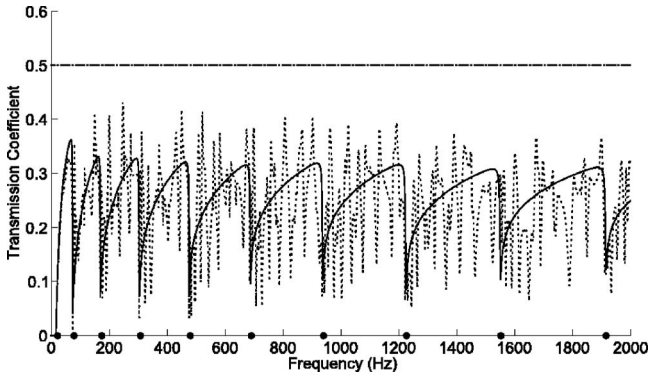


FIG. 9. Comparison of the total transmission coefficient for semi-infinite (—) and finite (···) systems excited at 100 random point locations on plate 1. Also shown is the transmission coefficient predicted by Cremer *et al.*—Ref. 5 (-·-) and the modal cut-on frequencies (●).

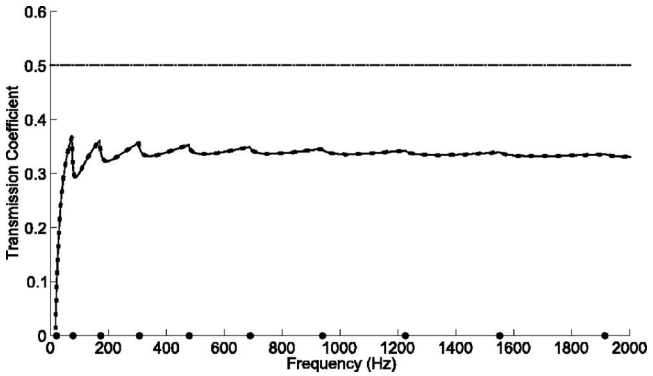


FIG. 10. Comparison of the total transmission coefficient for semi-infinite (—) and finite (···) systems excited by an incident wave. Also shown is the transmission coefficient predicted by Cremer *et al.*—Ref. 5 (-·-) and the modal cut-on frequencies (●).

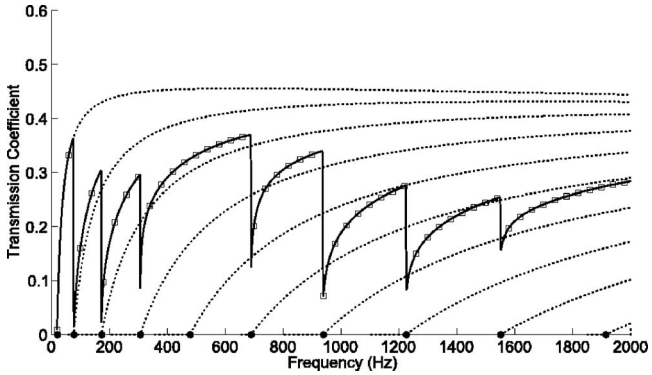


FIG. 11. Total transmission coefficient for a semi-infinite (—) system excited at $(x_1, y)=(100, 100)$ and corresponding modal transmission coefficients $(\cdots)$. Also shown are the modal cut-on frequencies (●) and total transmission coefficient (□) derived using the modal transmission coefficient relation.

It is of interest to note that for both the semi-infinite and finite systems, the total transmission coefficients obtained by both methods are an exact match.

In Fig. 13, the proportion of incident energy associated with each mode for the semi-infinite L-shaped plate is presented. It is evident that an active mode is always dominant at its cut-on frequency. The sudden reduction in the total transmission coefficient at each modal frequency, as observed in Figs. 3 and 9, is due to a combination of the mode under consideration being dominant, and a nearly zero value in its corresponding modal transmission coefficient at its cut-on frequency. In regions away from the cut-on frequencies, the incident energy is made up of well-defined proportions of the incident modal energies. Similar results are observed for the contribution of the incident modal energy to the total incident energy for the finite system, although the resonant behavior results in the dominant mode changing rapidly with frequency. When a mode becomes dominant, its proportion of incident energy is equal or close to unity and hence accounts for almost all of the incident energy on the junction. As the frequency increases and more modes are active, the increased modal overlap makes it difficult for a single mode to dominate the incident energy.

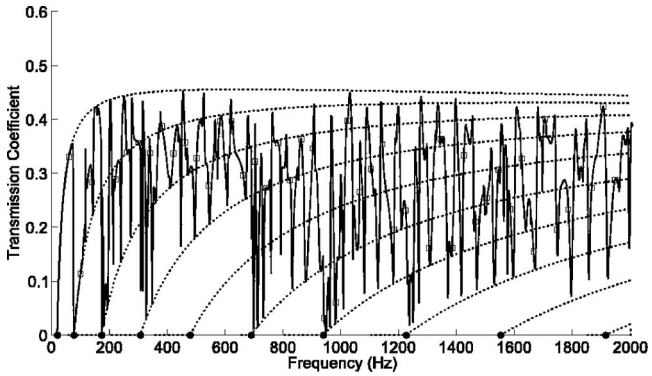


FIG. 12. Transmission coefficient for a finite (—) system excited at $(x_1, y)=(100, 100)$ and corresponding modal transmission coefficients $(\cdots)$. Also shown are the modal cut-on frequencies (●) and transmission coefficient (□) derived using the modal transmission coefficient relation.

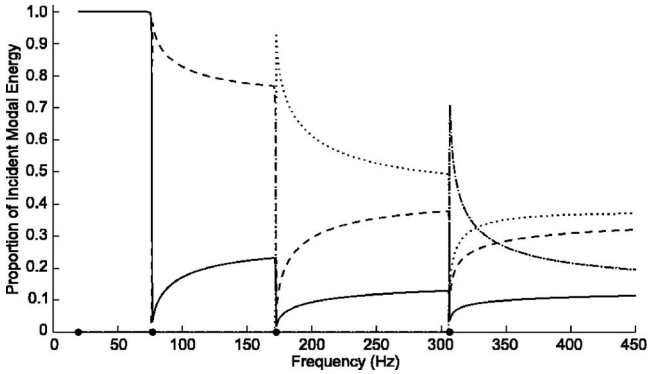


FIG. 13. Proportion of incident modal power attributed to mode 1 (—), 2 (---), 3 $(\cdots)$, and 4 (--) for a semi-infinite L-shaped plate system excited at $(x_1, y)=(100, 100)$. Also shown are the modal cut-on frequencies (●).

C. Frequency averaged transmission coefficients

The frequency averaged transmission coefficient is commonly used in energy methods such as statistical energy analysis.[1] Figures 14 and 15 compare the one-third octave band frequency averaged transmission coefficients for the semi-infinite and finite L-shaped plates, for a single force located at $(x_1, y)=(100, 100)$ and multiple point force excitation, respectively. Figure 14 shows that while significant discrepancies exist between the frequency averaged transmission coefficients for the semi-infinite and finite systems for single force excitation, the semi-infinite transmission coefficient is a very good approximation for the finite system for multiple force excitation. Comparison of Fig. 15 with Figs. 5 and 9 indicates that although there are significant differences between the transmission coefficients for finite and semi-infinite coupled plates, these differences become less pronounced with both frequency averaging and multiple force excitation.

IV. CONCLUSION

Transmission coefficients are widely used as a parameter in energy based methods such as SEA for the vibroacoustic analyses of large scale built-up structures, particularly for buildings, aircraft, and ship hulls. This paper presents a thorough investigation on transmission coefficients for finite systems, developed from a wave approach combined with an

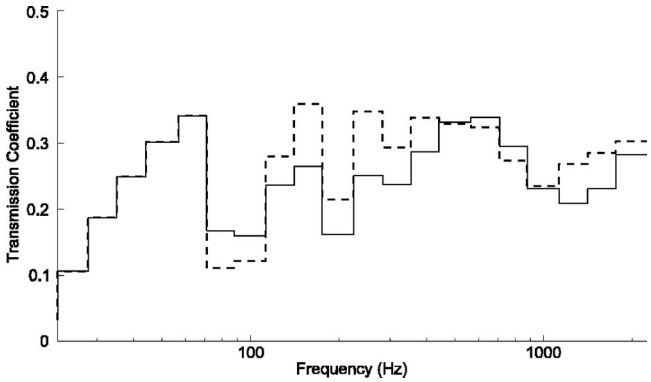


FIG. 14. One-third octave band frequency averaged transmission coefficient for finite (---) and semi-infinite (—) coupled L-shaped plates excited at $(x_1, y)=(100, 100)$.

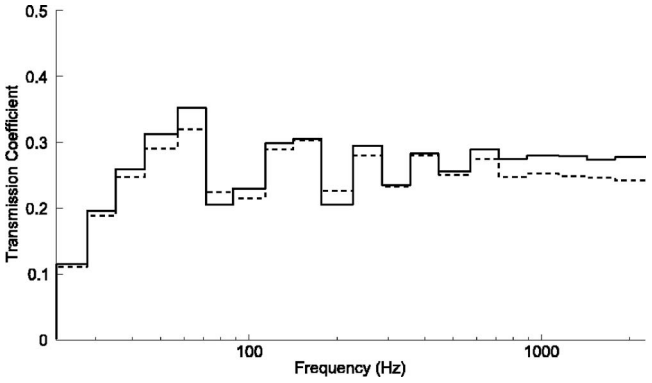


FIG. 15. One-third octave frequency averaged transmission coefficients for finite (---) and semi-infinite (—) coupled L-shaped plates with multiple force excitation on plate 1.

energy flow equation and impedance method, and using a scattering matrix to separate the transmitted and reflected waves leaving the junction in a finite coupled structure. A separate analysis has revealed that the transmission coefficient can be described in terms of the product of its individual modal transmission coefficients and proportion of corresponding incident modal energy. Furthermore, it was shown that the individual modal transmission coefficients predicted for finite and semi-infinite L-shaped plates are identical, which allows the transmission coefficient for a finite system to be calculated without using the scattering matrix method. Results show that the total transmission coefficients for both semi-infinite and finite structures never exceed the value of the modal transmission coefficient for the first mode. Whilst there are significant differences in the total transmission coefficients for semi-infinite and finite coupled plates, these differences become less obvious using frequency averaging, and under multiple force excitation (resulting in equipartition of energy in all modes). In this analysis, only bending waves in plates are considered, but the use of the scattering matrix method in conjunction with the analytical waveguide method could be extended to include contributions of other wave types; for example, in order to account for the in-plane motion which becomes more significant at higher frequencies.

ACKNOWLEDGMENT

This work was carried out by funding provided by the Australian Research Council under ARC Discovery Grant No. DP0451313.

[1]R. H. Lyon and R. G. Dejong, *Theory and Application of Statistical Energy Analysis* (Butterworth-Heinemann, Boston, 1995).
[2]P. E. Cho and R. J. Bernhard, "Energy flow analysis in coupled beams," J. Sound Vib. **211**, 593–605 (1998).
[3]A. J. Romano, P. B. Abraham, and E. G. Williams, "A Poynting vector formulation for thin shells and plates, and its application to structural intensity analysis and source localization. I. Theory," J. Acoust. Soc. Am. **87**, 1166–1175 (1990).
[4]E. C. N. Wester and B. R. Mace, "Wave component analysis of energy flow in complex structures. I. A deterministic model," J. Sound Vib. **285**, 209–227 (2005).
[5]L. Cremer, M. Heckl, and E. E. Ungar, *Structure-borne Sound*, 2nd ed. (Springer, Berlin, 1988).
[6]R. S. Langley and K. H. Heron, "Elastic wave transmission through plate/beam junctions," J. Sound Vib. **143**, 241–253 (1990).
[7]R. S. Langley and P. J. Shorter, "The wave transmission coefficients and coupling loss factors of point connected structures," J. Acoust. Soc. Am. **113**, 1947–1964 (2003).
[8]W. H. Park, D. J. Thompson, and N. S. Ferguson, "The influence of modal behaviour on the energy transmission between two coupled plates," J. Sound Vib. **276**, 1019–1041 (2004).
[9]R. H. Lyon and T. D. Scharton, "Power flow and energy sharing in random vibration," J. Acoust. Soc. Am. **43**, 1332–1343 (1967).
[10]R. J. M. Craik, *Sound Transmission through Buildings using Statistical Energy Analysis* (Gower, Aldershot, 1996).
[11]P. J. Shorter and R. S. Langley, "On the reciprocity relationship between direct field radiation and diffuse reverberant loading," J. Acoust. Soc. Am. **117**, 85–95 (2005).
[12]S. Timoshenko and S. Woinowsky-Krieger, *Theory of Plates and Shells* (McGraw-Hill, Singapore, 1959).
[13]X. Pan and C. H. Hansen, "Active control of vibratory power transmission along a semi infinite plate," J. Sound Vib. **184**, 585–610 (1995).
[14]N. Kessissoglou, "Power transmission in L-shaped plates including flexural and in-plane vibration," J. Acoust. Soc. Am. **115**, 1157–1169 (2004).

Wild African elephants (*Loxodonta africana*) discriminate between familiar and unfamiliar conspecific seismic alarm calls

Caitlin E. O'Connell-Rodwell[a] and Jason D. Wood
Department of Otolaryngology, Head and Neck Surgery, Stanford University School of Medicine, Stanford, California 94305-5739

Colleen Kinzley
Oakland Zoo, Oakland, California 94605

Timothy C. Rodwell
University of California, San Diego, Department of Family and Preventative Medicine, La Jolla, California 92093

Joyce H. Poole
Amboseli Research Project, Buskhellinga 3, 3236 Sandefjord, Norway

Sunil Puria
Department of Otolaryngology, Head and Neck Surgery, and Department of Mechanical Engineering, Stanford University, Stanford, California 94305-5739

(Received 5 December 2006; revised 23 April 2007; accepted 10 May 2007)

The ability to discriminate between call types and callers as well as more subtle information about the importance of a call has been documented in a range of species. This type of discrimination is also important in the vibrotactile environment for species that communicate via vibrations. It has recently been shown that African elephants (*Loxodonta africana*) can detect seismic cues, but it is not known whether they discriminate seismic information from noise. In a series of experiments, familiar and unfamiliar alarm calls were transmitted seismically to wild African elephant family groups. Elephants respond significantly to the alarm calls of familiar herds ($p=0.004$) but not to the unfamiliar calls and two different controls, thus demonstrating the ability of elephants to discriminate subtle differences between seismic calls given in the same context. If elephants use the seismic environment to detect and discriminate between conspecific calls, based on the familiarity of the caller or some other physical property, they may be using the ground as a very sophisticated sounding board. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2747161]

PACS number(s): 43.40.Ng, 43.66.Wv, 43.66.Gf, 43.38.Md [JAS] Pages: 823–830

I. INTRODUCTION

Animals assess their acoustic environment based on frequency, amplitude and temporal properties of sounds. These parameters have a different level of importance for different species, depending on the limitations of the ear, the importance of the acoustic environment to survival, the social organization of the species and the context of the sound. A variety of species are able to use the acoustic properties of their calls to detect differences within conspecific vocalizations that distinguish call types as illustrated in Barbary macaques (Fischer, 1998) and elephants (Langbauer *et al.*, 1991; Poole, 1999), as well as familiar versus unfamiliar callers in the spear-nosed bat (Boughman and Wilkinson, 1998), sheep (Ligout *et al.*, 2004), lions (McComb *et al.*, 1993), bottlenose dophins (Sayigh *et al.*, 1999) and elephants (McComb *et al.*, 2001), or even body size (Cheney and Seyfarth, 1991).

Research on suricate vocalizations has also shown that information about the level of danger presented by the proximity of a predator can also be discerned, based on the severity of the call (Manser, 2001). Information about the type of predator is even encoded in vervet monkey vocalizations (Seyfarth *et al.*, 1980). In addition, subtle frequency differences that distinguish the individual caller have been found in marmots (Blumstein *et al.*, 2004). The ability to detect subtle changes in frequency has been demonstrated in such species as the squirrel monkey (Weinicke *et al.*, 2001), where this species is able to detect frequency differences as small as 20–40 Hz in the range of 4–8 kHz, and an especially keen discrimination ability above 10 kHz, followed by the bottlenose dolphin (Thompson and Herman, 1975) and the lesser spear-nosed bat (Esser and Keifer, 1996). But in the lower frequency range, in the hundreds of hertz, the squirrel monkey has poor frequency discrimination ability. Frogs that vocalize in the range of 350–400 Hz assess the size of the caller based on frequency (Bee *et al.*, 2000), where frequency discrimination is on the order of 12–14%; some frogs discriminating as little as a 5.7% difference (Wagner, 1992).

Frequency sensitive touch receptors have been described in humans in the ranges of 5–15, 10–65 and 65–400 Hz (Makous *et al.*, 1995). Such vibrotactile sensory structures have been found in primates and other large mammals, in-

[a]Author to whom correspondence should be addressed. Electronic mail: ceoconnell@stanford.edu

TABLE I. Elephant group sizes and composition for each trial (where subadults are three quarters the size of a full size adult at the shoulder and back, half being half the size of an adult, and one quarter is a range between greater than one year old to less than half the size of an adult. A baby fits under the stomach of a full size elephant and ranges up to one year.

Adult	Sub-adult	Half	Quarter	Baby	Total No.	Playback
						Unfamiliar alarm
7	5	7	5	1	25	1
8	4	4	2	0	18	2
7	4	8	2	2	23	3
4	4	5	6	2	21	4
9	0	8	3	2	22	5
3	2	4	2	1	12	6
3	1	3	3	0	10	7
8	2	2	5	4	21	Control 1
5	4	4	2	1	16	2
4	2	4	3	0	13	3
6	1	5	4	3	19	4
3	4	4	4	0	15	5
8	4	8	3	2	25	6
4	4	6	3	1	18	7
7	5	4	2	2	20	8
11	8	9	2	5	35	9
1	1	4	3	1	10	Familiar alarm 1
5	4	3	6	2	20	2
7	3	4	5	3	22	3
9	5	6	4	3	27	4
3	1	3	1	2	10	5
5	2	3	1	1	12	6
8	2	7	4	1	22	7
6	3	6	3	2	20	8

cluding elephants (Rasmussen and Munger, 1996). The ability of these touch receptors to discriminate very small changes in frequency (2 Hz) has been demonstrated in humans and other primates (Recanzone *et al.*, 1992). For those species that communicate seismically (see O'Connell-Rodwell *et al.*, 2000a for review), the ability to distinguish call type and individual callers has been demonstrated in the kangaroo rat (Randall, 2001).

We have previously demonstrated that elephant vocalizations propagate in the ground (O'Connell-Rodwell *et al.*, 2000a; Gunther *et al.*, 2004) and that elephants are capable of detecting seismic cues (O'Connell-Rodwell *et al.*, 2006), but it has not yet been established whether they have the ability to discriminate between various seismic signals. In this study, we test the ability of African elephant family groups to discriminate subtle differences between familiar and unfamiliar callers within the same call type.

II. METHODS

A. Experimental design

A series of seismic playback experiments were conducted by transmitting previously recorded acoustic vocalizations of known context into the ground to elephant family groups at a remote waterhole in Etosha National Park, Namibia between the hours of 4:00 P.M. and 2:00 A.M. Experiments were videotaped, and night vision used for experiments occurring after sunset. Total numbers of individuals were counted in real time, or from the videotape and a herd composition breakdown compiled for each group (Table I) to confirm that groups were not being treated more than once within any one playback stimulus. Two different seismic stimuli were delivered to determine if elephants can distinguish subtle differences between meaningful biological signals made in the same context (alarm) but by familiar and unfamiliar callers. As controls, we played back a generated warble tone or no stimulus at all. Each trial began 2 min after the arrival of the elephants, to allow them to drink and settle down. Following this, 5 min of base line observations were made. Three minutes of playback stimuli was delivered, where 15 s of signals were played seismically at the beginning of each minute. A subsequent 5 min period was used to monitor any changes in behavior.

The playback stimuli were as follows. **Familiar alarm calls** consisted of three alarm calls emitted by the individuals of one family group while lions were hunting near them at this study site in Etosha National Park. These calls have been shown to elicit a vigilant response (O'Connell-Rodwell *et al.*, 2006). **Unfamiliar alarm calls** consisted of three alarm calls emitted by two different family groups in Amboseli National Park, Kenya while lions were hunting near them. Since these calls are rarely recorded by researchers, the familiar and unfamiliar calls used were the only ones available

for this playback study. All alarm signals (familiar and unfamiliar) were filtered with a Butterworth bandpass filter (low cut at 10 Hz and high cut between 50 and 60 Hz) such that only the fundamental and second harmonic were still present in the calls.

Controls consisted of either no seismic stimulus at all, or a series of three simulated warble tones. The warble tones were designed with frequency content and duration similar to an elephant rumble. Its base frequency of 30 Hz was modulated by 3 Hz at a rate of 1 Hz for 3 s, with 2 s of silence between the three signals.

All signals were played back seismically through two Guitammer Buttkicker LFE shakers (frequency range 5–200 Hz with 9 Hz resonant frequency), buried in the ground 20 m from the water hole. A TASCAM digital two-channel recorder provided the signal source for the transmitters. A 1000 W amplifier was used to raise the amplitude of the signals to a level resembling the power of an elephant vocalization at a distance of 20 m. See Figs. 1(a)–1(c) for spectrograms of playback signals.

Playback signals were recorded during the trials on a Geometrics Geode 24 channel seismic recorder through two, 4.5 Hz Mark Products vertical geophones 10 m from the source, one placed 10 m from the shaker toward the water hole and the other 10 m from the shaker in the opposite direction to measure the signal strength at the noisy waterhole versus a more quiet area away from the waterhole. These sensors were used to monitor the integrity of the playback signal. A Neumann KM131 low frequency microphone was used to record the trials in the acoustic environment, to record vocal responses to the different stimuli, as well as to ensure that no presentation signal coupled with the air. This microphone was placed directly above the geophone that was placed in the direction away from the water hole.

The order of trial type (familiar alarm, unfamiliar alarm, control) was randomized. Each trial was presented during separate waterhole visits by single family groups. Family groups were distinguished by herd size and composition and the data presented are representative of distinct groups within each playback treatment type.

Elephant behaviors were monitored for adult members of each of the family group tested. Individual behaviors were scored, including freezing/leaning, scanning, lifting one foot, smelling, head shakes and vocalizing, each as a measure of vigilance, or heightened wariness in the context of a potentially threatening situation. Herd spacing was scored separately (similar to O'Connell-Rodwell *et al.*, 2006; McComb *et al.*, 2001; Poole, 1999; Langbauer *et al.*, 1991), where individuals were noted as being within a body length, at one body length, or greater than one body length apart. One experienced elephant behavior observer (naïve to the trial types and timing of the trials) recorded individual behaviors while another documented herd spacing. The occurrence of the vigilant behaviors listed above was noted every 15 s during the trials, then summed (equal weighting) for the pre and postplayback periods, then divided by the duration of that period and the number of elephants present to give us a measure of vigilant behavior. Herd spacing was also noted every 15 s during trials and then averaged for the pre and postplayback periods. All playback sessions were video recorded by a third observer who also documented herd size and composition. All observations were made from a tower and two platforms, 100 m from the water hole.

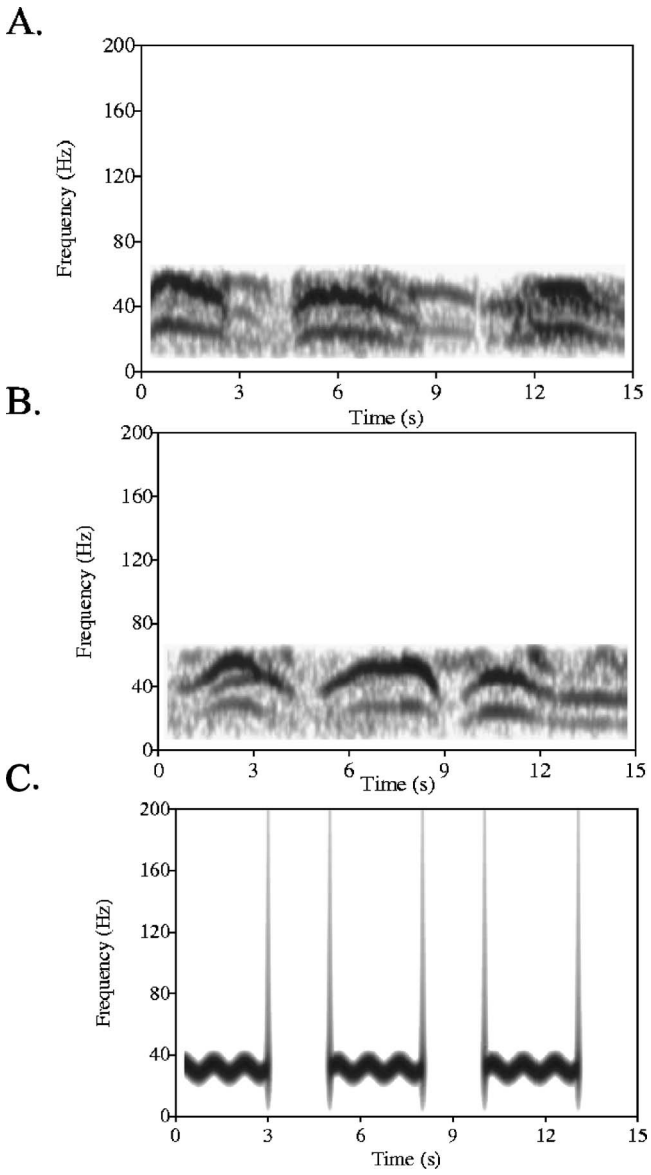


FIG. 1. Spectrograms of the playbacks that were used. A: Unfamiliar Alarm Call Rumbles B: Familiar Alarm Call Rumbles C: Warble tones. Each spectrogram was band pass filtered to remove higher frequencies that tend to couple with the air when played seismically as well as very low frequency noise (filter used was a 20th order Butterworth band pass with low cut at 10 Hz and high cut between 50 and 60 Hz). Spectrograms generated in Praat software V4.1 (Institute of Phonetic Sciences, University of Amsterdam) with the following settings: sampling rate 16 000 Hz, window length 0.3 s, max freq 200 Hz, Gaussian window (equivalent to an FFT size of 4800 and resultant frequency resolution of 3.33 Hz).

B. Signal calibration and acoustic coupling

We used a matched filter technique to confirm that there was no evidence of the seismic signal in the acoustic environment (acoustic coupling) during our seismic playback trials. We modified the following matched filter from http://cnx.rice.edu/content/m10757/latest/

$$Matched\ Filter = \frac{|\langle f, g_i \rangle|}{\|g_i\|}$$

to

$$Matched\ Filter = \frac{|\langle f, g_i \rangle|}{\|f\|^* \|g_i\|}$$

by adding the norm of f so that the matched filter would vary between 0 and 1; f is the signal while g_i is the recording at time i. The numerator is the absolute value of the inner products of f and g_i. The denominator in the modified equation is the norm of f multiplied by the norm of g_i. As one moves the matched filter along the recording the output varies between 0 and 1, with 1 being a perfect match. To determine a threshold above which acoustic coupling would occur, we calculated the relationship between the matched filter output and a biological criterion. For this study, we calculated the relationship between the matched filter output and the signal to noise ratio (SNR) in order to throw out any trials with a matched filter equivalent to −2 dB SNR in our microphone recordings during the playbacks. We chose this as our cutoff point because, to date, the most comparable hearing pattern to that of the elephant, in terms of frequency range where data are available to serve as a reference point, is that of the human which has a signal detection threshold at −2 dB SNR (Zwicker and Feldtkeller, 1999).

Because the matched filter output varies depending on the noise in which the signal is embedded, we felt it was most appropriate to utilize noise from each of the trials. Therefore a 15 s segment of noise was extracted from the microphone recording for each trial. The playback signal was then added to this noise so as to achieve a SNR of −2 dB for each trial. We focused on the second harmonic of the alarm call playback as the second harmonics were of a larger amplitude and higher frequency and therefore more likely to couple with the air. For the control (warble) playbacks there were no harmonics, and so we focused on the fundamental frequency.

SNR was measured before adding the signal to the noise by calculating the spectrum of the highest amplitude 1 s segment of the signal and the spectrum of the same 1 s of noise in the microphone recording. We then integrated the energy across the signal width of the rumble and integrated the noise energy across the estimated critical bandwidth for elephants at the center frequency of the signal. To estimate critical bandwidth, we followed Günther *et al.*, 2004 and Greenwood, 1961. The bandwidth for each call is different due to the difference in frequencies of each call. We used the following estimates in Hz: (frequency: critical bandwidth) familiar alarm: 52:19, unfamiliar alarm: 49:18, control warble tone: 33:15.

Once we had inserted the signal into the microphone noise of each trial at a −2 dB SNR, we ran the matched filter on these files. The output of the matched filter for each trial was used as the cutoff point for our seismic playback trials (i.e., the matched filter equivalent of −2 dB SNR). We then ran the matched filter on the microphone recording of each seismic playback trial. Any trial that had a matched filter output greater than the −2 dB SNR equivalent that was within 100 ms of our playbacks was not included in our further analyses. See Figs. 2(a)–2(c) for matched filter output in the microphone/geophone pairs for each playback type showing the signal present in the ground but not in the air at a level greater than −2 dB SNR.

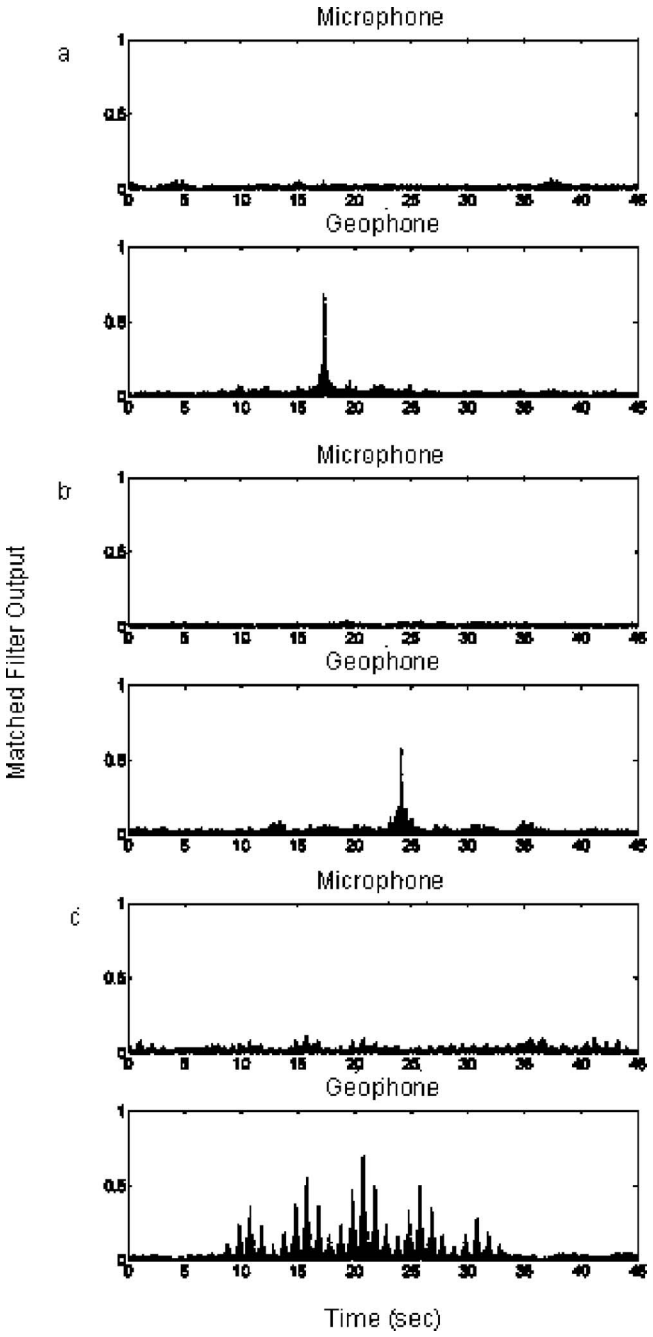


FIG. 2. Typical examples of the matched filter output in the microphone recording (top graph of each pair) and the geophone recordings (bottom graph of each pair). The playback peak is clearly visible in the geophone graph, but not in the microphone graph. (a) Result of a familiar alarm call seismic playback. Playback occurred at ~17 s. (b) Result of an unfamiliar alarm call seismic playback. Playback occurred at ~24 s. (c) Result of warble seismic playback. Playback occurred at ~20 s. Multiple peaks are evident because the same signal was repeated three times and because there is a repetitive pattern within each signal. The highest peak though occurs at the playback time.

C. Statistical analysis

To test if our various seismic stimuli had an effect on vigilant behavior and herd spacing, we ran a series of repeated measures multivariate analysis of variance (MANOVA) tests. This allowed us to test vigilant behavior and herd spacing at the same time, while controlling for any correlation between these two dependent variables. All statistical tests were conducted in MINTAB (v 13) (MINITAB Inc., State College, PA).

After excluding trials that had evidence of acoustic coupling, we had the following sample sizes of individual family groups treated for each of our three stimulus types: Unfamiliar Alarm Calls: $N=7$, Control: $N=9$, Familiar Alarm Calls: $N=8$. As a test to ensure that combining our control stimuli (warble and no stimulus) was appropriate, we ran two sample t tests comparing vigilant behavior or herd spacing in the postplayback periods for these two control stimuli. Since we found no significant difference we felt it justified to combine these into a single control stimulus in order to simplify our experimental design and increase our sample size [Warble $N=4$, No stimulus $N=5$, Spacing ($t=1.56$, $P=0.162$, $DF=7$), Behavior ($t=-0.78$, $P=0.459$, $DF=7$)]. Our original MANOVA tested if there was a difference in vigilant behavior and herd spacing before versus after our seismic playbacks. The Wilks' lambda criterion found a significant difference ($F_{2,20}=18.649$, $P<0.001$). Given this result we tested each stimulus type separately using a repeated measures MANOVA to determine which stimulus type resulted in significant differences before and after the seismic playback.

D. Call type differences

In order to assess whether there were any quantitative differences between the playback signals used in these experiments, a script was written for MATLAB (Mathworks Inc., Natick, MA) to extract the rumble frequency contour of all six alarm calls used in the seismic playbacks as well as the generated warble noise tone. This script was similar to the one used by McCowan (1995) and Wood *et al.* (2005) in that it extracted the rumble frequency at 40 evenly spaced points along the duration of each rumble. This was done by calculating the spectrum at each of these 40 points and recording the peak frequency. The sampling rate of the signals was 1000 Hz while the fast Fourier transform (FFT) length was set at 2048 making the frequency resolution 0.5 Hz.

Since the second harmonic was of a higher amplitude than the fundamental in our playbacks, we concentrated on it by filtering the calls and sampling the second harmonic at 40 points. Wood *et al.* (2005) found ten acoustic parameters that differed significantly between the elephant rumble types they analyzed. We calculated these same ten parameters from our rumble contours to see if there were noticeable differences between our playback calls. We did not, however, run any statistical tests on these parameters, as the sample size was too small.

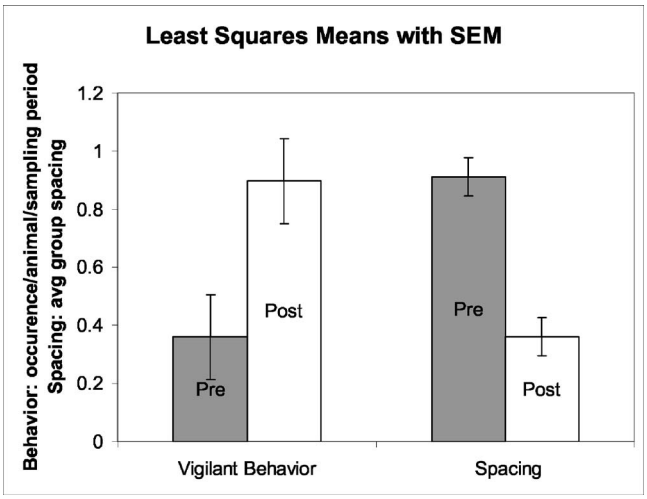


FIG. 3. Least squares means and standard error of the means from the MANOVA test comparing pre and postperiod vigilant behavior and spacing for the familiar alarm call playbacks. Behavior is measured as number of occurrences per animal per sampling period. Spacing is measured in body lengths.

III. RESULTS

We found no significant change in vigilant behavior and herd spacing when comparing the pre and postseismic playbacks periods for the control, or unfamiliar alarm calls using the Wilk's lambda criterion for the MANOVA; Control: $F_{2,7}=4.328$, $P=0.060$, Unfamiliar Alarm Call: $F_{2,5}=3.572$, $P=0.109$. We did, however, find a significant change in vigilant behavior and herd spacing when comparing the pre and postseismic playback periods for the familiar alarm calls (MANOVA: $F_{2,6}=15.720$, $P=0.004$). Vigilant behavior increased after the playbacks while spacing decreased (Fig. 3). See Table II for the least squares means and standard error of the means.

A. Call type differences

Figure 4 depicts the frequency contours extracted from the six alarm call and warble tone playbacks, while Table III

TABLE II. Least squares means and the standard errors of the means from the MANOVA tests on vigilant behavior and spacing.

Stimulus	Period	LSM behavior	SEM behavior	LSM spacing	SEM spacing
Control	Pre	0.36	0.13	0.91	0.09
Control	Post	0.69	0.13	0.53	0.09
Unfamiliar alarm	Pre	0.45	0.08	1.00	0.11
Unfamiliar alarm	Post	0.54	0.08	0.55	0.11
Familiar alarm	Pre	0.36	0.15	0.91	0.07
Familiar alarm	Post	0.90	0.15	0.36	0.07

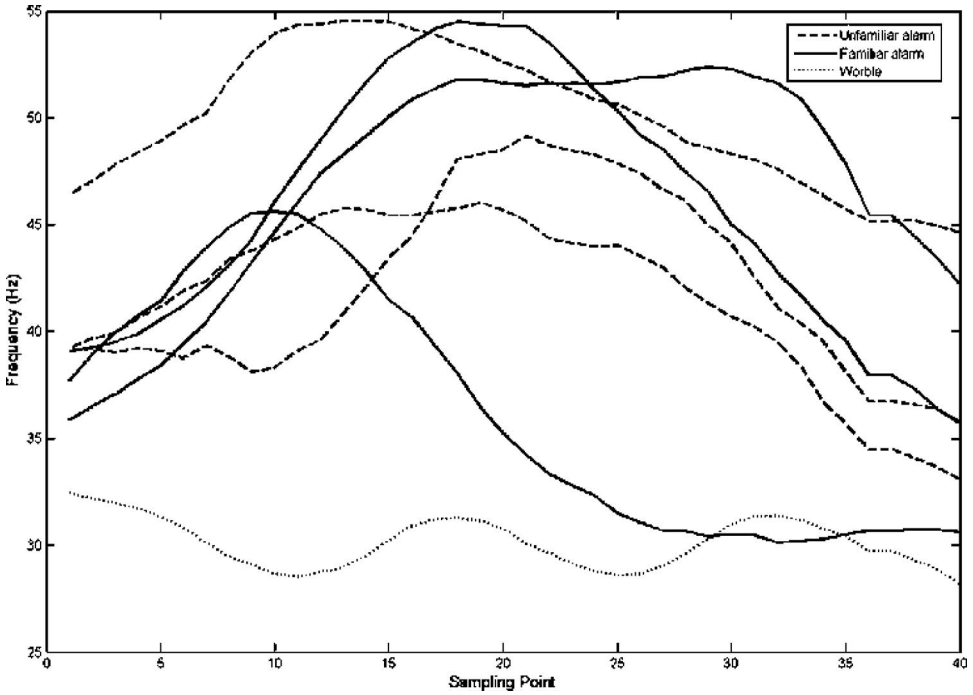


FIG. 4. Graphic of rumble contour differences between the unfamiliar alarm call rumbles recorded in Amboseli (dashed line), the familiar alarm call rumbles recorded in Etosha (solid line), and the generated warble tone (dotted line). All contours depict the second harmonic, other than for the warble tone, which is the fundamental.

lists the ten acoustic parameters extracted from these frequency contours. The warble tone is easily differentiated in the figure and table by its lower frequency (min, max, mean), smaller frequency modulation (FR), and the way in which the frequency is modulated (CV). The variables with the most consistent differences between unfamiliar and familiar alarm calls are frequency variability (CV), which is measuring the magnitude of the frequency modulation across the rumbles, and frequency range (FR). Familiar alarm call rumbles have a larger amount of frequency modulation and a larger frequency range (Table III).

IV. DISCUSSION

In this study, we played familiar and unfamiliar seismic alarm call signals to elephant family groups while they visited a water hole, as well as a warble tone which served as a seismic control stimulus. Only one of these signals (familiar alarm calls) was found to cause a significant change in behavior. Vigilant behaviors increased while spacing decreased. This leads us to draw two main conclusions.

First, because there was a significant change in behavior after the familiar alarm calls, but not during control trials, we are able corroborate our earlier findings that elephants are able to detect a seismic signal of biological importance in the absence of its acoustic counterpart (O'Connell-Rodwell *et al.*, 2006). Second, there was a significant change in behavior after the familiar alarm calls, but not after the unfamiliar alarm calls. This suggests a very fine ability to discriminate between biological seismic signals given in the same context. Either the familiar alarm was a more intense call (in terms of frequency modulation), thus inducing a more dramatic response, or it is also possible that alarms made from unfamiliar callers may not be perceived as being a reliable source of information.

A. Call recognition

The familiar alarm calls had a slightly higher frequency modulation, which could code for a more severe threat as the relative change in frequency is somewhat analogous to motivation-structural rules (Morton, 1977). If the herd ex-

TABLE III. Parameters used to quantify differences between the anti-predator alarm calls recorded at two different sites ($n=3$ for each site) and the generated warble tone. CV is the frequency variability index calculated as the variance in frequency divided by the square of the mean frequency, and then multiplying by 10. IF is the inflection factor calculated as the percentage of points showing a reversal in slope. PAF is the peak amplitude frequency. FF is the final frequency. FR is the frequency range. Duration is in seconds.

Recording	CV	IF	PAF	FF	MIN	MAX	MEAN	FR	MAX/MEAN	DUR
Unfamiliar alarm 1	0.042	0.077	48.01	44.60	44.60	54.53	49.94	9.93	1.09	2.49
Unfamiliar alarm 2	0.089	0.128	41.99	33.04	33.04	46.01	41.58	12.97	1.11	3.61
Unfamiliar alarm 3	0.106	0.154	46.60	35.65	35.65	49.10	42.10	13.45	1.17	4.44
Familiar alarm 1	0.176	0.026	49.15	35.73	35.73	54.47	45.72	18.74	1.19	3.73
Familiar alarm 2	0.128	0.128	45.41	42.16	35.84	52.36	47.10	16.52	1.11	3.97
Familiar alarm 3	0.251	0.128	45.57	30.60	30.11	45.57	36.28	15.46	1.26	5.16
Warble	0.015	0.103	[a]	28.16	28.16	32.42	30.10	4.26	1.08	3.00

[a]There was no amplitude modulation during warble tone

posed to the familiar alarm calls interpreted the higher frequency modulation as a warning of more imminent threat, their reactions of increased vigilance and decreased herd spacing would be entirely appropriate. On the other hand, if threatening calls contain enough geographic variation such that the Kenyan elephant alarm calls were unrecognizable to Namibian herds, this could explain the lack of response to the unfamiliar call. This seems unlikely however, as the physical properties of the alarm calls are very similar and it seems likely that this type of call would be fairly universal. And in a previous study, we showed that elephants responded with different levels of intensity to the same alarm call played back in different areas of the same park (O'Connell-Rodwell *et al.*, 2000b).

A more likely explanation for the difference in reaction to these alarm calls is that the elephant herds differentiated the alarm calls as being familiar and unfamiliar. Given that the local herds would have interacted with the herd that originally made the familiar alarm calls, they would be in a much better position to evaluate the reliability of those calls. They would know from prior experience whether or not the signaling herd was likely to be correct in its assessment of the level of danger. Given that McComb *et al.* (2001, 2003) found that elephant herds could distinguish the contact calls of other herds as being part of their bond group, or outside their bond group, it seems plausible that the elephants in our playback studies are capable of doing the same.

B. Call structure, detection pathways and frequency discrimination

The analysis of the physical parameters of our seismic playback calls provides some insight into how these elephants might be distinguishing between seismic signals. The control warble is easily distinguishable from the other signals by a number of variables (mean frequency, frequency variability (CV) and frequency range (FR); see Table III), and also by the fact that only the fundamental is present in the signal, while both the fundamental and second harmonic are present in the alarm call presentations. The alarm calls, however, are more similar to each other, but most distinguishable from each other by their frequency variability and range (CV and FR). Since these elephants discriminated between these two sets of calls, they may be relying on the differences in frequency modulation for this discrimination.

The frequency range of the second harmonic of the alarm signals varied from about 10–19 Hz and should be within the range of vibrotactile frequency discrimination ability of elephants. No one has measured this directly in the African elephant, but we can make estimates based on work in other species, using similar sensory structures. The frequency discrimination ability of seismic signals in these elephants would depend on which pathway of detection is used. Two pathways have been proposed, bone conduction from the feet to the ear (Reuter *et al.*, 1998), or somatosensory (O'Connell *et al.*, 1999) via vibrotactile corpuscles in the feet (Weissengruber *et al.*, 2006; Bouley *et al.*, *in press*).

If the pathway of detection of seismic signals is via bone conduction to the ear, then the frequency discrimination ability will be reliant on the acoustic frequency discrimination ability of this species. As noted above, we estimated the acoustic critical bandwidth in the frequency range of our playback calls to be around 15–19 Hz. Fletcher (1940) found that the minimum perceptible frequency change (Δf) was related to the critical bandwidth (CBW) in the following way: $\mathrm{CBW} = \Delta f^{*}20$. Therefore, if this equation holds true for elephants as well, we would estimate a Δf of 0.75–0.95 Hz, which would allow them the ability to detect very small changes in frequency modulation across these calls.

If the seismic detection pathway is via vibrotactile corpuscles, then elephants should still be able to discriminate fine frequency differences. Recanzone *et al.* (1992) tested the tactile frequency discrimination ability of adult owl monkeys, using 20 Hz as the reference tone. They found that the monkeys' ability to discriminate frequency differences improved from an initial 6 Hz down to 2 Hz. They report that their final threshold was similar to those found in humans and macaques (Goff, 1967; LaMotte *et al.*, 1975; Mountcastle *et al.*, 1969, 1990). Given that primates have not been shown to use seismic signaling, while our data support the idea of elephants using this modality, it is likely that elephants have at least the same vibrotactile frequency discrimination abilities as primates, if not better. Elephants could be using the Pacinian corpuscles found in their trunks (Rasmussen and Munger, 1996) and possibly in their feet to distinguish the frequency modulation between familiar and unfamiliar alarm calls.

The ability to tap into the seismic channel to discriminate biologically relevant information from background noise and to discriminate subtle differences between calls of familiar versus unfamiliar groups indicates that elephants may be using the ground as a sounding board for much more subtle cues than previously thought. Given the ability to detect subtle frequency differences, they most probably could also distinguish larger events such as an approaching vehicle, helicopters, airplanes, weather (thunderstorms) or earthquakes, providing the elephant with a sophisticated ability to exploit the seismic modality for many different purposes. Having previously shown that elephants produce and detect seismic cues and now demonstrating that elephants respond to and discriminate between seismic cues, we present the full complement of signal and receiver assessment components necessary from signal detection theory to state that elephants may indeed be communicating seismically.

ACKNOWLEDGMENTS

Simon Klemperer for his technical support, Robert Sapolsky research support, David Shriver as a naive observer, Katie Ekhart for videography, herd size and composition data collection and Bob Dickerson of Jim Walters Sound Co. for his technical advice. We would like to thank Jo Tagg for the loan of field and camping supplies, Johannes Kapner, Wilfred Versfeld and Werner Kilian at Etosha Ecological Institute, Ministry of Environment & Tourism (MET) for field support; Pauline Lindeque, Director of Scientific Services, MET; Namibia Nature Foundation for logistical support. Funding support came from a Stanford University Bio-X Inter-disciplinary Research award, the U.S. Fish & Wildlife

Service, Oakland Zoo Conservation Fund, UC Davis School of Veterinary Medicine Student Travel Grant, and a generous grant from the Seaver Institute.

Bee, M. A., Perrill, S. A., and Owen, P. C. (**2000**). "Male green frogs lower the pitch of acoustic signals in defense of territories: A possible dishonest signal of size?," Behav. Ecol. Sociobiol. **11**(2), 169–177.

Blumstein, D. T., Verneyre, L., and Daniel, J. C. (**2004**). "Reliability and the adaptive utility of discrimination among alarm callers," Proc. R. Soc. London, Ser. B **271**, 1851–1857.

Boughman, J. W., and Wilkinson, G. S. (**1998**). "Greater spear-nosed bats discriminate group mates by vocalizations," Anim. Behav. **55**, 1717–1732.

Bouley, D. M., Alarcon, C., Hildebrandt, T. and O'Connell-Rodwell, C. E. (**in press**). "The distribution, density and three dimensional histomorphology of Pacinian Corpuscles in the Asian elephant (*Elephas maximus*) foot and their potential role in detecting seismic information," J. Anant.

Cheney, D. L., and Seyfarth, R. M. (**1991**). *Cognitive Ethology: The Minds of Other Animals*, edited by C. A. Ristau (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 127–151.

Esser, K. H., and Keifer, R. (**1996**). "Detection of frequency modulation in the FM-bat *Phyllostomus discolor*," J. Comp. Physiol., A **178**, 787–796.

Fischer, J. (**1998**). "Barbary macaques categorize shrill barks into two call types," Anim. Behav. **55**, 799–807.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Goff, G. D. (**1967**). "Differential discrimination of frequency of cutaneous mechanical vibration," J. Exp. Psychol. **74**, 294–299.

Greenwood, D. (**1961**). "Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Am. **33**(484), 1344–1356.

Gunther, R., O'Connell-Rodwell, C. E., and Klemperer, S. (**2004**). "Seismic waves from elephant vocalizations: A possible communication mode?," Geophys. Res. Lett. **31**(L11602), 1–4.

LaMotte, R. H., and Mountcastle, V. B. (**1975**). "Capacities of humans and monkeys to discriminate between vibratory stimuli of different frequency and amplitude: A correlation between neural events and psychophysical measurements," J. Neurophysiol. **38**, 539–559.

Langbauer, W. R., Jr., Payne, K. B., Charif, R. A., Rapaport, L., and Osborn, F. (**1991**). "African elephants respond to distant playbacks of low-frequency conspecific calls," J. Exp. Biol. **157**, 35–46.

Ligout, S., Sebe, F., and Porter, R. (**2004**). "Vocal discrimination of kin and non-kin age mates among lambs," Behaviour **141**, 355–369.

Makous, J. C., Friedman, R. M., and Vierck, C. J., Jr. (**1995**). "A critical band filter in touch," J. Neurosci. **15**(4), 2808–2818.

Manser, M. (**2001**). "The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency," Proc. R. Soc. London, Ser. B **268**, 2315–2324.

McComb, K., Moss, C., Durant, S. M., Baker, L., Sayialel, S., *et al.* (**2001**). "Matriarchs as repositories of social knowledge in African elephants," Science **292**(5516), 491–494.

McComb, K., Pusey, A., Packer, C., and Grinnell, J. (**1993**). "Female lions can identify potentially infanticidal males from their roars," Proc. R. Soc. London, Ser. B **252**(1333), 59–64.

McComb, K., Reby, D., Baker, L., Moss, C., and Sayialel, S. (**2003**). "Long-distance communication of acoustic cues to social identity in African elephants," Anim. Behav. **65**, 317–329.

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.

Morton, E. S. (**1977**). "On the occurrence and significance of motivation-structural rules in some bird and mammal sounds," Am. Nat. **111**, 855–869.

Mountcastle, V. B., Talbot, W. H., Sakata, H., and Hyvärinen, J. (**1969**). "Cortical neuronal mechanisms in flutter-vibration studied in unanesthetized monkeys. Neuronal periodicity and frequency discrimination." J. Neurophysiol. **32**, 452–484.

Mountcastle, V. B., Steinmetz, M. A., and Romo, R. (**1990**). "Frequency discrimination in the sense of flutter: Psychophysical measurements correlated with post central events in behaving monkeys," J. Neurosci. **10**, 3032–3044.

O'Connell-Rodwell, C. E., Wood, J. D., Rodwell, T. C., Puria, S., Shriver, D., Partan, S. R., Keefe, R., Arnason, B. T., and Hart, L. A. (**2006**). "Wild elephant (*Loxodonta africana*) breeding herds respond to artificially transmitted seismic stimuli," Behav. Ecol. Sociobiol. **59**(6), 842–850.

O'Connell-Rodwell, C. E., Arnason, B., and Hart, L. A. (**2000a**). "Seismic properties of Asian elephant (*Elephas maximus*) vocalizations and locomotion," J. Acoust. Soc. Am. **108**(6), 3066–3072.

O'Connell-Rodwell, C. E., Rodwell, T. C., Rice, M., and Hart, L. A. (**2000b**). "The modern conservation paradigm: Can agricultural communities co-exist with elephants? (Five-year case study in East Caprivi, Namibia)," Biol. Conserv. **93**, 381–391.

O'Connell, C. E., Hart, L. A., and Arnason, B. (**1999**). "Response to "Elephant hearing" [see comments] J. Acoust. Soc. Am. **104**, 1122–1123 (1998)," J. Acoust. Soc. Am. **105**, 2051–2052.

Poole, J. H. (**1999**). "Signals and assessment in African elephants: Evidence from playback experiments," Anim. Behav. **58**(1), 185–193.

Randall, J. (**2001**). "Evolution and function of drumming as communication in mammals," Am. Zool. **41**, 1143–1156.

Rasmussen, L. E. L., and Munger, B. L. (**1996**). "The sensorineural specializations of the trunk tip (finger) of the asian elephant, Elephas maximus," Anat. Rec. **246**, 127–134.

Recanzone, G. H., Jenkins, W. M., Hradek, G. T., and Merzenich, M. M. (**1992**). "Progressive improvement in discriminative abilities in adult Owl Monkeys performing a tactile frequency discrimination task," J. Neurophysiol. **67**(5), 1015–1030.

Reuter, T., Nummela, S., and Hemilea, S. (**1998**). "Elephant hearing [letter]," J. Acoust. Soc. Am. **104**, 1122–1123.

Sayigh, L. S., Tyack, P. L., Wells, R. S., Solow, A. R., Scott, M. D., and Irvine, A. B. (**1999**). "Individual recognition in wild bottlenose dolphins: A field test using Playback experiments," Anim. Behav. **57**, 41–50.

Seyfarth, R. M., Cheney, D. L., and Marler, P. (**1980**). "Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication," Science **210**(4471), 801–803.

Thompson, R. K., and Herman, L. M. (**1975**). "Underwater frequency discrimination in the bottlenose dolphin (1–40 kHz) and the human (1–8 kHz)," J. Acoust. Soc. Am. **57**, 943–948.

Wagner, W. E., Jr. (**1992**). "Deception or honest signaling of fighting ability? A test of alternative hypotheses for the function of changes in call dominance frequency by male cricket frogs." Anim. Behav. **44**, 449–462.

Weinicke, A., Häusler, U., and Jürgens, U. (**2001**). "Auditory frequency discrimination in the squirrel monkey," J. Comp. Physiol., A **187**, 189–195.

Weissengruber, G. E., Egger, G. F., Hutchinson, J. R., Groenewald, H. B., Elsasser, L., Famini, D., Forstenpointner, G. (**2006**) "The structure of the cushion in the feet of African elephants (*Loxodonta Africana*)," J. Anat. **209**(6), 781–792.

Wood, J. D., McCowan, B., Langbauer, W. R., Jr., Viljoen, J. J., and Hart, L. A. (**2005**). "Classification of African elephant (*Loxodonta africana*) rumbles using acoustic parameters and cluster analysis," Bioacoustics **15**, 143–161.

Zwicker, E., and Feldtkeller, R. (**1999**). *The Ear as a Communication Receiver*, translated by H. Müsch *et al.* (Acoustical Society of America, Woodbury, NY).

Poroelastic modeling of seismic boundary conditions across a fracture[a)]

Seiji Nakagawa[b)]
Earth Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720

Michael A. Schoenberg
5 Mountain Road, West Redding, Connecticut 06896

(Received 19 December 2006; revised 15 May 2007; accepted 16 May 2007)

Permeability of a fracture can affect how the fracture interacts with seismic waves. To examine this effect, a simple mathematical model that describes the poroelastic nature of wave-fracture interaction is useful. In this paper, a set of boundary conditions is presented which relate wave-induced particle velocity (or displacement) and stress including fluid pressure across a compliant, fluid-bearing fracture. These conditions are derived by modeling a fracture as a thin porous layer with increased compliance and finite permeability. Assuming a small layer thickness, the boundary conditions can be derived by integrating the governing equations of poroelastic wave propagation. A finite jump in the stress and velocity across a fracture is expressed as a function of the stress and velocity at the boundaries. Further simplification for a thin fracture yields a set of characteristic parameters that control the seismic response of single fractures with a wide range of mechanical and hydraulic properties. These boundary conditions have potential applications in simplifying numerical models such as finite-difference and finite-element methods to compute seismic wave scattering off nonplanar (e.g., curved and intersecting) fractures.
[DOI: 10.1121/1.2747206]

PACS number(s): 43.40.Ph, 43.20.Gp, 43.20.Bi [LLT] Pages: 831–847

I. INTRODUCTION

Rock is often permeated by compliant plane discontinuities (such as fractures and faults) that, depending on their permeability relative to the background, serve as either conduits or barriers to subsurface fluid flow (e.g., Aydin, 1978; Adams and Dart, 1998). In the following, we shall collectively call these discontinuities "fractures." The fluid permeability of a fracture is often a key parameter, yet the quantitative relationship between permeability and its effect on seismic wave scattering is not fully understood. Strong scattering of seismic waves by a fracture is usually related to large permeability, because an open fracture with partial surface contacts has increased mechanical compliance (deformability) (e.g., Pyrak-Nolte and Morris, 2001). However, if a fluid-containing fracture is filled with debris, or a single fracture consists of a large number of microcracks, complex interactions between rock and pore fluid in the fracture result. In this paper, we will develop a simple mathematical model that captures the essential nature of solid-fluid interaction within a fracture, to predict the effect of hydraulic permeability and other fracture properties on seismic wave scattering.

One logical tool for probing the hydrological properties of rocks using seismic waves is Biot's theory of poroelasticity (Biot, 1956a, b), which describes the dynamic interactions of rock and fluid within the pore space. It has been widely recognized, however, that the dispersion and attenuation of seismic waves predicted by the original Biot's theory—which applies to macroscopically homogeneous porous media saturated by a single fluid phase—is often too small to explain the measured velocity dispersion and attenuation of seismic waves. In recent decades, many researchers realized that significant velocity dispersion and velocity attenuation can result at field-relevant frequencies if a rock contains heterogeneity at mesoscale (smaller than seismic wavelength but larger than pore and grain size). One such effect is due to a local fluid-pressure gradient induced at scales comparable to the pressure diffusion length (or, wavelength of Biot's slow compressional waves). These heterogeneities can be, for example, a "patchy" distribution of fluid and gas within rocks (e.g., White, 1975; Dutta and Odé, 1979a, b; and Johnson, 2001) and stratified sedimentary units with different mechanical and hydrological properties (Norris, 1993; Gurevich *et al.*, 1994, 1997; Gelinsky and Shapiro, 1997; Shapiro and Müller, 1999; Pride *et al.*, 2002). A more general theory for heterogeneous poroelastic media, with arbitrary distributions of mechanical and hydraulic properties for both solid and fluid phases, was recently developed by Pride and Berryman (2003a, b).

In general, within porous fluid-bearing rocks, the stronger the fluid permeability and mechanical-property heterogeneity, the more the velocity and attenuation of seismic waves are affected. Fractures are a special case of such heterogeneity, exhibiting an extremely wide range of mechanical com-

[a)]Portions of this work were presented in "Poroelastic modeling of seismic boundary conditions across a fracture," Expanded abstract for the annual meeting for the Society of Exploration Geophysicists, New Orleans, 1–5 Oct. 2006.

[b)]Electronic mail: snakagawa@lbl.gov

pliance and hydraulic permeability (for example, open, air-filled joints to near-rigid, mineral-filled veins), even though they typically occupy only a small volume. Berryman and Wang (1995) examined the mechanical consolidation of media containing a system of compliant high-permeability fractures within a porous background medium, and then used the derived elastic moduli to examine the velocity dispersion and attenuation of low-frequency seismic waves (Berryman and Wang, 2000). The results indicated that the mechanical and hydraulic properties of fractures in a porous host rock affect the behavior of seismic waves. For one-dimensional *P*-wave propagation within a medium containing parallel periodic fractures, Brajanovski *et al.* (2005) derived an analytical model for the dispersion and attenuation of waves. This model was derived by using a wave propagation model for alternating poroelastic layers developed by Norris (1993) and taking the zero-thickness limit of one of the constituting layers to model fractures. Through numerical experiments, attenuation and dispersion of *P* wave propagation was found to be strongly dependent upon the fracture properties (fracture stiffness and density) and the background porosity.

In contrast to previous research, which focused on the velocity and attenuation of waves propagating through materials containing many fractures, in this paper we will develop a simple mathematical model for single poroelastic fractures that can be used to study discrete scattering of seismic waves. The model consists of a set of boundary conditions that relate the stress and displacement (or particle velocity) induced on the fracture surface by passing seismic waves. These boundary conditions are derived using plane-wave theory, by treating a fracture as a thin poroelastic layer with an infinite extent and a small finite thickness. Alternatively, scattering of the plane waves can be examined by using a propagator-matrix method (e.g., Haskell, 1953) and Kennett's reflectivity method (Kennett, 1983) to find an exact relationship between the amplitude of incident and scattered waves. However, the propagator-matrix method suffers an instability when the Biot's slow *P* wave decays too quickly; and Kennett's method results in very complex expressions of the boundary conditions that are not amenable to simple parametrization and interpretation of the consequences for physical acoustics. Further, because both of these methods require knowledge of incident plane waves on both sides of a fracture, they are not well suited to use in other numerical models, such as finite-difference and finite-element methods.

The model developed in this paper provides "jump conditions" that directly relate a wave's particle motions and stress across a fracture without the knowledge of the wave field in the background. Such boundary conditions were initially developed for elastic and viscoelastic fractures, and called the "linear-slip interface model" (Schoenberg, 1980), which led to a plethora of theories and models describing the complex interaction between seismic waves and fractures—e.g., plane-wave scattering theories by Schoenberg (1980), Nakagawa *et al.* (2000), laboratory experiments by Pyrak-Nolte *et al.* (1990), Hsu and Schoenberg (1993), fracture-based anisotropic effective medium theories of Schoenberg and Sayers (1999), Bakulin *et al.* (2000), fracture guided wave studies by Pyrak-Nolte and Cook (1987), and Nihei *et al.* (1999). More recently, Bakulin and Molotkov (1997) developed a similar model for poroelastic fractures, but without including the effect of fracture permeability. These models can be very simple, because when the relative thickness of a fracture is much smaller than the seismic wavelengths (Biot's fast *P* waves and *S* waves), and inertia-related quantities (given as a product of density and fracture thickness) can be ignored, only quasistatic behavior needs to be described (Rokhlin and Wang, 1991). Gurevich *et al.* (1994) also used this fact to derive simple, computationally stable expressions describing the transmission and reflection coefficients of normally incident fast *P* waves for a thin poroelastic layer. Because of its simplicity, the linear-slip model can be used in finite-difference codes to determine the proper effective anisotropic-elastic-moduli values of the numerical grids on a fracture, when the thickness of the fracture is much smaller than the modeling grid spacing (Coates and Schoenberg, 1995).

One important aspect of the linear-slip interface model is that it helps to identify important characteristic parameters of a fracture that control the scattering of seismic waves. An example of such parameters is the fracture compliance. If a fracture is modeled as a mechanically equivalent, thin, compliant layer with a finite thickness, the fracture compliance can be defined as an inverse of the elastic moduli times fracture thickness (e.g., Rokhlin and Wang (1990) defined a fracture stiffness parameter [inverse of the fracture compliance] in this way). Coates and Schoenberg (1995) developed a finite-difference model for fractures and faults, based upon the finite-thickness approximation of fractures and faults. Conversely, when physical properties of a fracture are to be determined using seismic waves, what we can at best determine are these "phenomenological" model parameters (instead of the original material properties and fracture thickness). For a fracture viewed as a thin poroelastic layer, we will show that characteristic parameters similar to the original linear-slip interface model can be defined for a poroelastic fracture, along with other dimensionless parameters that describe its poroelastic properties.

In the following, first we will derive poroelastic seismic boundary conditions (linear-slip interface model) based upon the governing equations of linear, poroelastic wave propagation (Secs. II A and II B). This will result in two sets of independent matrix equations relating displacement and stress across a fracture, which are the primary results of this work. The critical step in this derivation is an approximation of the wave-induced pressure field within a fracture—this is necessary because the exact pressure distribution cannot be determined only from the boundary values. Subsequently, assuming a fracture thickness much smaller than the wavelength of propagating body waves, the derived boundary conditions will be simplified to obtain the characteristic fracture parameters (Sec. II C). The original and simplified boundary conditions will be used to derive explicit expressions for plane-wave transmission and reflection coefficients (Sec. II D). Sections III A and III B will examine the accuracy of the derived boundary conditions (both original and simplified) by comparing the predicted transmission and reflection coefficients to the exact results obtained via Ken-

nett's reflectivity method. Finally, the sensitivity of the transmission and reflection coefficients to the permeability of a fracture will be examined using a characteristic fracture parameter (Sec. III C).

II. THEORY

In this section, we will derive a set of boundary conditions for a thin, isotropic, homogeneous, poroelastic layer embedded within a background medium. (A derivation of boundary conditions assuming a transversely isotropic poroelastic layer for a fracture is also presented in Appendix A.) Subsequently, these boundary conditions are used to derive expressions for transmission and reflection coefficients of incident plane waves within a poroelastic background medium.

A. Governing equations

The governing equations of seismic wave propagation within an isotropic, homogeneous, poroelastic medium can be stated as (e.g., Pride *et al.*, 2002)

$$\boldsymbol{\tau} = G(\nabla \mathbf{u} + \mathbf{u}\nabla) + [(K_U - 2G/3)\nabla \cdot \mathbf{u} + C\nabla \cdot \mathbf{w}]\mathbf{I}, \quad (1)$$

$$-p_f = C\nabla \cdot \mathbf{u} + M\nabla \cdot \mathbf{w}, \quad (2)$$

$$\nabla \cdot \boldsymbol{\tau} = -\omega^2(\rho \mathbf{u} + \rho_f \mathbf{w}), \quad (3)$$

$$-\nabla p_f = -\omega^2(\rho_f \mathbf{u} + \tilde{\rho}\mathbf{w}), \quad \tilde{\rho} \equiv i\eta_f/\omega k(\omega), \quad (4)$$

where $\mathbf{u}$ is the locally averaged, solid-frame displacement vector and $\mathbf{w} \equiv \phi(\mathbf{U}\text{-}\mathbf{u})$ is the fluid-volume displacement vector relative to the solid frame. In this definition of $\mathbf{w}$, $\mathbf{U}$ is the locally averaged (in the pore space) fluid displacement vector and ϕ is the porosity. Equations (1)–(4) assume that the displacement and stress variables depend on $\exp(-i\omega t)$, where ω is the circular frequency. $\mathbf{I}$ indicates an identity tensor, $\boldsymbol{\tau}$ is the total stress tensor, and p_f is the fluid pressure (positive for compression). G is the solid-frame shear modulus, K_U is the undrained bulk modulus, ρ is the bulk density, ρ_f is the fluid modulus, and the parameter $\tilde{\rho}$ is defined in Eq. (4) via fluid viscosity η_f and the frequency-dependent permeability $k(\omega)$ (Johnson *et al.*, 1987). C and M are the Biot's coupling and fluid-storage moduli, respectively. When a plane harmonic wave field is assumed, these equations result in the four plane-wave modes of a Biot medium (fast and slow P waves and two S waves).

Consider an interface across which certain stress and displacement (velocity) components are conserved. We assume this interface to be normal to the 3 direction of Cartesian coordinates, and wave propagation parallel to the 1, 3 plane (Fig. 1). For a homogeneous medium, we can assume a plane harmonic wave field proportional to $\exp i\omega(\xi_1 x_1 - t)$, where ξ_1 is the slowness in the 1 direction. The plane-wave displacement and stress are introduced into Eqs. (1)–(4). Substituting $\partial/\partial x_1 \rightarrow i\omega\xi_1$ and $\partial/\partial x_2 \rightarrow 0$ and eliminating components of the vector and tensor variables w_1, w_2, τ_{11}, τ_{22}, and τ_{12} (which can be discontinuous across the interface), the following two independent sets of coupled first-order differential equations are derived:

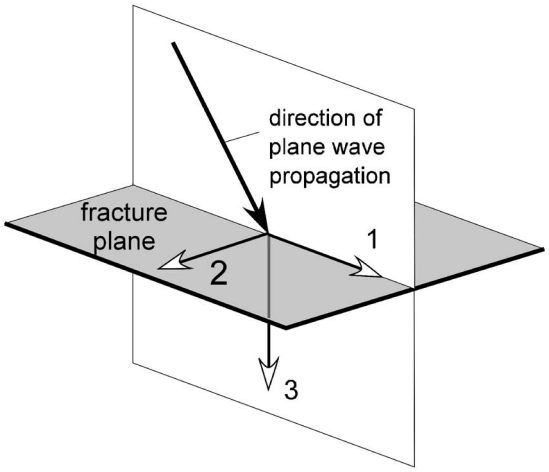


FIG. 1. Cartesian coordinate system used in this paper.

$$\frac{\partial}{\partial x_3}\begin{bmatrix} \dot{u}_2 \\ \tau_{23} \end{bmatrix} = -i\omega \begin{bmatrix} 0 & 1/G \\ -G\xi_1^2 + (\rho\tilde{\rho} - \rho_f^2)/\tilde{\rho} & 0 \end{bmatrix}\begin{bmatrix} \dot{u}_2 \\ \tau_{23} \end{bmatrix} \equiv -i\omega \mathbf{R}\begin{bmatrix} \dot{u}_2 \\ \tau_{23} \end{bmatrix}, \quad (5)$$

$$\frac{\partial}{\partial x_3}\begin{bmatrix} \dot{u}_1 \\ \tau_{33} \\ -p_f \\ \tau_{13} \\ \dot{u}_3 \\ \dot{w}_3 \end{bmatrix} = -i\omega \begin{bmatrix} \mathbf{0} & \mathbf{Q}_{XY} \\ \mathbf{Q}_{YX} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \dot{u}_1 \\ \tau_{33} \\ -p_f \\ \tau_{13} \\ \dot{u}_3 \\ \dot{w}_3 \end{bmatrix}, \quad (6)$$

where

$$\mathbf{Q}_{XY} \equiv \begin{bmatrix} 1/G & \xi_1 & 0 \\ \xi_1 & \rho & \rho_f \\ 0 & \rho_f & \tilde{\rho} \end{bmatrix}, \quad (7)$$

$$\mathbf{Q}_{YX} \equiv \begin{bmatrix} -4G\xi_1^2\left(1 - \frac{G}{H_D}\right) - \frac{\rho_f^2 - \rho\tilde{\rho}}{\tilde{\rho}} & \xi_1\left(1 - \frac{2G}{H_D}\right) & \xi_1\left(-\frac{\rho_f}{\tilde{\rho}} + \alpha\frac{2G}{H_D}\right) \\ \xi_1\left(1 - \frac{2G}{H_D}\right) & \frac{1}{H_D} & -\frac{\alpha}{H_D} \\ \xi_1\left(-\frac{\rho_f}{\tilde{\rho}} + \alpha\frac{2G}{H_D}\right) & -\frac{\alpha}{H_D} & \frac{\alpha^2}{H_D} + \frac{1}{M} - \frac{\xi_1^2}{\tilde{\rho}} \end{bmatrix}. \quad (8)$$

Equation (5) is for wave propagation of S waves with particle motions in the 2 direction, and Eq. (6) is for coupled P (both fast and slow)-S wave propagation with particle motions within the 1,3 plane. Note that both $|\mathbf{R}(\xi_1,\omega)|=0$ and $|\mathbf{Q}_{XY}(\xi_1,\omega)|=0$ yield the dispersion equation for S waves, and $|\mathbf{Q}_{YX}(\xi_1,\omega)|=0$ results in the dispersion equation for fast and slow P waves, where $|\cdot|$ indicates the matrix determinant. The dots over the displacement vector components in Eqs. (5) and (6) indicate that the related quantity is velocity. In Eqs. (7) and (8), $H_D \equiv K_D + 4G/3$ is the dry P-wave modulus and $\alpha=(1-K_D/K_U)/B$ is the Biot-Willis effective stress coefficient (with B as the Skempton coefficient). Using these coefficients, C and M in the governing equations can be expressed as $C=BK_U$ and $M=BK_U/\alpha$. Further, if grains in the porous rock are both isotropic and homogeneous,

$$\alpha = 1 - K_D/K_s, \tag{9}$$

$$B = \frac{1/K_D - 1/K_s}{(1/K_D - 1/K_s) + \phi(1/K_f - 1/K_s)}, \tag{10}$$

where K_D is the dry bulk modulus, K_s is the solid (grain) bulk modulus, K_f is the fluid bulk modulus, and ϕ is the porosity of the medium.

B. Derivation of poroelastic boundary conditions for a fracture

The boundary conditions for a poroelastic fracture are obtained by integrating the governing equations in Eqs. (5) and (6) over a small layer or fracture thickness h as

$$\begin{bmatrix} \dot{u}_2^+ - \dot{u}_2^- \\ \tau_{23}^+ - \tau_{23}^- \end{bmatrix} = -i\omega h \begin{bmatrix} 0 & 1/G \\ -G\xi_1^2 + (\rho\tilde{\rho} - \rho_f^2)/\tilde{\rho} & 0 \end{bmatrix} \begin{bmatrix} \bar{\dot{u}}_2 \\ \bar{\tau}_{23} \end{bmatrix}, \tag{11}$$

$$\begin{bmatrix} \dot{u}_1^+ - \dot{u}_1^- \\ \tau_{33}^+ - \tau_{33}^- \\ -p_f^+ - (-p_f^-) \\ \tau_{13}^+ - \tau_{13}^- \\ \dot{u}_3^+ - \dot{u}_3^- \\ \dot{w}_3^+ - \dot{w}_3^- \end{bmatrix} = -i\omega h \begin{bmatrix} \mathbf{0} & \mathbf{Q}_{XY} \\ \mathbf{Q}_{YX} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \bar{\dot{u}}_1 \\ \bar{\tau}_{33} \\ -\bar{p}_f \\ \bar{\tau}_{13} \\ \bar{\dot{u}}_3 \\ \bar{\dot{w}}_3 \end{bmatrix}, \tag{12}$$

where the superscripts + and − indicate quantities on the boundaries, and the bars above the variables indicate averaged quantities over the thickness of the fracture. At this point, Eqs. (11) and (12) are without approximations, except that we assumed homogeneity of the medium and plane-wave propagation. To derive boundary conditions, the averaged quantities on the right-hand side of the equations have to be expressed exclusively using quantities on the boundaries.

Since the thickness of a fracture h is usually much smaller than seismic wavelengths, the inertial effect and complex multiple scattering of the waves within the fracture can be ignored. This allows us to assume that the solid-frame

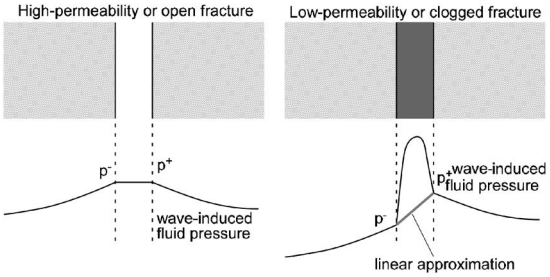


FIG. 2. Cartoon representation of pressure induced by seismic waves for high-permeability (left) and low-permeability (right) fractures. For a high-permeability fracture, fluid pressure on both sides of the fracture can equilibrate during a period of oscillation. In contrast, for a low-permeability fracture, the pressure induced within the fracture may not be able to dissipate. If the pressure field is approximated using a linear function using the boundary values, this can result in a significant error in evaluating the average fluid pressure across the fracture.

velocity and the total stress within the fracture vary smoothly, which can be approximated by a linear function. Also, since the field distribution must be defined by two boundary values on the fracture surfaces, and since there is no knowledge of the field's functional form, a linear function provides the best guess. For Eq. (11), therefore, the boundary condition becomes

$$\begin{bmatrix} \dot{u}_2^+ - \dot{u}_2^- \\ \tau_{23}^+ - \tau_{23}^- \end{bmatrix} = -\frac{i\omega h}{2} \begin{bmatrix} 0 & 1/G \\ -G\xi_1^2 + (\rho\tilde{\rho} - \rho_f^2)/\tilde{\rho} & 0 \end{bmatrix} \times \begin{bmatrix} \dot{u}_2^+ + \dot{u}_2^- \\ \tau_{23}^+ + \tau_{23}^- \end{bmatrix}. \tag{13}$$

In contrast, for Eq. (12), if the permeability of the fracture is low and the fluid within the fracture is not allowed to move freely, excess pore pressure can be induced, which can be very different from the pressure at the interfaces (as illustrated in Fig. 2). This excess pore pressure also induces rapidly changing fluid velocity. Therefore, to provide a better approximation of the spatially averaged fluid pressure and velocity on the right-hand side of Eq. (12), we must examine the behavior of a diffusing fluid pressure field within a low-permeability fracture.

Unfortunately, the quantities on the boundaries alone cannot provide enough information to determine the non-monotoic profile of the field within the fracture. To overcome this difficulty, we first assume that the fluid velocity relative to the frame at the boundaries can be attributed exclusively to slow P waves. This attribution can be justified if $\omega k_0 \rho_f/\eta_f = \rho_f/|\tilde{\rho}(0)| \ll 1$, where $k_0 \equiv k(0)$, because this factor essentially provides the amplitude ratio between the fluid velocity and the solid velocity for fast P waves and S waves. [For example, if a water-filled fracture has a permeability of 10 mD (10^{-14} m^2), $\omega k_0 \rho_f/\eta_f < 0.01$ for frequencies less than 100 kHz.] Therefore, if we define $-p_f^*$ and $\dot{w}_3^*$ as the pressure and fluid flow response excluding the contribution of slow P waves, these are given by an "undrained" fracture (sealed at the boundaries) (Pride, 2003). Note that $-p_f^*$ can be considered uniform across the fracture, due to the long wavelengths of the fast P wave and S wave. Further, we assume that the

spatial average of $\dot{u}_1$ and τ_{33} for this field can be approximated by the average of the total field $\overline{\dot{u}_1}$ and $\bar{\tau}_{33}$. This assumption can be justified if the incident wave is not a slow P wave and the frequency is low, which results in amplitudes of scattered slow P waves much smaller than the sum of the other waves.

Under these assumptions, the jump condition for fluid velocity in the matrix equation [the bottom row of Eq. (12)] can be used to obtain the following relationship:

$$\dot{w}_3^{+*} - \dot{w}_3^{-*} = 0 = -i\omega h\left[\xi_1\left(-\frac{\rho_f}{\tilde{\rho}} + 2\alpha\frac{G}{H_D}\right)\overline{\dot{u}_1} - \frac{\alpha}{H_D}\bar{\tau}_{33} + \left(\frac{\alpha^2}{H_D} + \frac{1}{M} - \frac{\xi_1^2}{\tilde{\rho}}\right)(-p_f^*)\right]. \tag{14}$$

The fluid velocity within the undrained fracture is $\dot{w}_3^*=0$. From the above equation,

$$-p_f^* = \xi_1\frac{\dfrac{\rho_f}{\tilde{\rho}} - 2\alpha\dfrac{G}{H_D}}{\dfrac{\alpha^2}{H_D} + \dfrac{1}{M} - \dfrac{\xi_1^2}{\tilde{\rho}}}\overline{\dot{u}_1} + \frac{\dfrac{\alpha}{H_D}}{\dfrac{\alpha^2}{H_D} + \dfrac{1}{M} - \dfrac{\xi_1^2}{\tilde{\rho}}}\bar{\tau}_{33} \equiv -2G\tilde{B}\tilde{\beta}\xi_1\overline{\dot{u}_1} + \tilde{B}\bar{\tau}_{33}, \tag{15}$$

where we introduced the following coefficients $\tilde{B}$ and $\tilde{\beta}$ for convenience:

$$\frac{1}{\tilde{B}} \equiv \alpha + \frac{H_D}{\alpha M} - \frac{\xi_1^2}{\tilde{\rho}}\frac{H_D}{\alpha} = \frac{H_U}{\alpha M} - \frac{\xi_1^2}{\tilde{\rho}}\frac{H_D}{\alpha}, \tag{16}$$

$$\tilde{\beta} \equiv 1 - \frac{H_D}{2\alpha G}\frac{\rho_f}{\tilde{\rho}}. \tag{17}$$

The first term on the right-hand side of Eq. (15) indicates a contribution of the strain induced in the fracture-parallel direction ($-\xi_1\overline{\dot{u}_1} = \partial\overline{u_1}/\partial x_1$).

Next, we derive an expression for the diffusing pressure and flow field within a fracture using the pressure and velocity at the boundaries and the pressure and velocity for the undrained condition. The solution of diffusing field for slow waves with a slowness ξ_{Ps} is expressed as

$$f(x_3) = A_1 e^{i\omega\xi_{Ps}x_3} + A_2 e^{-i\omega\xi_{Ps}x_3}. \tag{18}$$

We assume that the direction of the diffusion is in the plane-normal direction, which is a reasonable assumption if the velocity of the incoming wave is much faster than the slow P-wave velocity within the fracture. For a set of boundary conditions $f(0)=0$ and $f(h)=1$, the two unknown coefficients are determined, resulting in

$$f(x_3) = \frac{e^{i\omega\xi_{Ps}x_3} - e^{-i\omega\xi_{Ps}x_3}}{e^{i\omega\xi_{Ps}h} - e^{-i\omega\xi_{Ps}h}}. \tag{19}$$

When integrated over an interval $[0, h]$, Eq. (19) becomes

$$\int_0^h f(x_3)dx_3 = \frac{h}{2}\cdot\frac{\tan\omega\xi_{Ps}h/2}{\omega\xi_{Ps}h/2} \equiv \frac{h}{2}\Pi(\varepsilon), \tag{20}$$

where we defined the following dimensionless function:

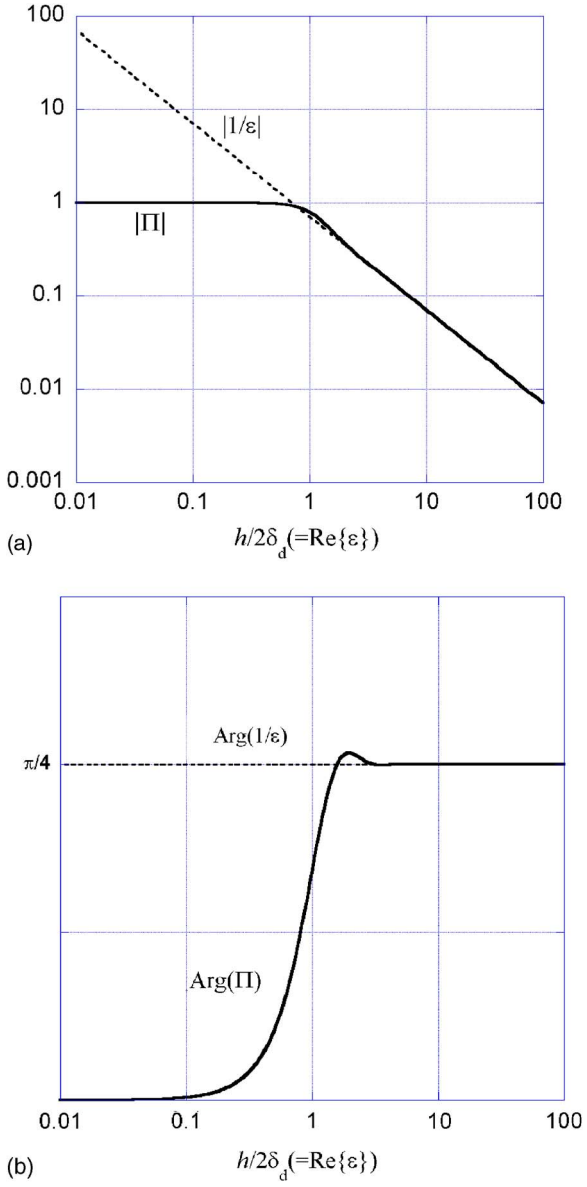


FIG. 3. (Color online) Fluid pressure dissipation factor Π as a function of the dimensionless length parameter $h/2\delta_d$. The behavior of the function changes when the diffusion length is half of the layer thickness, separating the low-frequency drained response (left-hand side of the $h/2\delta_d \sim 1$) and the high-frequency undrained response (right-hand side). (a) Amplitude; (b) Phase (Asymptotes to $\pi/4$ for the undrained response).

$$\Pi(\varepsilon) \equiv \frac{\tanh\varepsilon}{\varepsilon}, \quad \varepsilon \equiv -\frac{i\omega\xi_{Ps}h}{2} = \frac{h}{2\delta_d^*}. \tag{21}$$

The complex fluid-pressure diffusion length δ_d^* is defined through $i\omega\xi_{Ps} \equiv -1/\delta_d^*$. We shall call the dimensionless function Π (Fig. 3) a "fluid-pressure dissipation factor," which approaches unity for the low-frequency limit (drained response) and approaches zero for the high-frequency limit (undrained response). For the aforementioned low frequencies and low-permeability conditions satisfying $\omega\rho_f k_0/\eta_f \ll 1$, the following simple relationship can be used to compute δ_d^* (e.g., Pride, 2003):

$$\frac{1}{\delta_d^*} = \frac{1-i}{\delta_d}, \tag{22}$$

$$\delta_d = \sqrt{\frac{2D}{\omega}}, \tag{23}$$

$$D = \frac{k_0 M}{\eta_f}\left(1 - \frac{C^2}{H_U M}\right) = \frac{k_0 M}{\eta_f}\left(1 - \frac{\alpha^2 M}{H_U}\right), \tag{24}$$

where δ_d is the fluid-pressure diffusion length and D is the fluid-pressure diffusion coefficient. Using the solution for $f(x_3)$, the pressure and fluid velocity within a fracture is given by a superposition having the boundary conditions $-p_f(x_3=0)=-p_f^-$, $-p_f(x_3=h)=-p_f^+$, and $\dot{w}_3(x_3=0)=\dot{w}_3^-$, $\dot{w}_3(x_3=h)=\dot{w}_3^+$. These are

$$p_f(x_3) = p_f^* - (p_f^* - p_f^+)f(x_3) - (p_f^* - p_f^-)f(h - x_3), \tag{25}$$

$$\begin{aligned}\dot{w}_3(x_3) &= \dot{w}_3^* - (\dot{w}_3^* - \dot{w}_3^+)f(x_3) - (\dot{w}_3^* - \dot{w}_3^-)f(h - x_3) \\ &= \dot{w}_3^+ f(x_3) + \dot{w}_3^- f(h - x_3).\end{aligned} \tag{26}$$

As an example, pressure amplitude profiles are shown below in Fig. 4 for assumed boundary values of $-p_f^-=0.25$, $-p_f^+=0.75$, and $-p_f^*=1$. As seen from the plot, the transition between the drained response (linear pressure profile) and the undrained response (constant pressure within the fracture) occurs approximately when $h/2\delta_d=1$, i.e., the sum of the wavelength for the two diffusing pressure waves equals the thickness of the fracture.

Using the result in Eq. (20), pressure and fluid velocity averaged across a fracture is

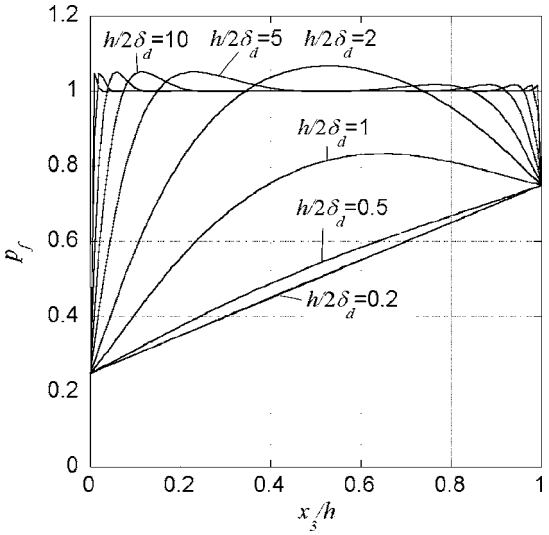


FIG. 4. Amplitude profiles of the pressure field within a fracture for a range of diffusion lengths δ_d. For very large δ_d's (i.e., high frequency, low permeability), the pressure within the fracture can take a value independent from the pressure on the fracture surfaces.

$$\begin{aligned}\bar{p}_f &= \frac{1}{h}\int_0^h p_f(x_3)dx_3 = p_f^* - (2p_f^* - p_f^- - p_f^+)\frac{1}{2}\cdot\Pi(\varepsilon) \\ &= \frac{p_f^- + p_f^+}{2}\cdot\Pi(\varepsilon) + p_f^*\cdot[1 - \Pi(\varepsilon)],\end{aligned} \tag{27}$$

$$\bar{\dot{w}} = \frac{1}{h}\int_0^h \dot{w}_3(x_3)dx_3 = \frac{\dot{w}_3^- + \dot{w}_3^+}{2}\Pi(\varepsilon). \tag{28}$$

Equations (15), (27), and (28) are introduced within the matrix boundary conditions in Eq. (12) to yield

$$\begin{bmatrix} \dot{u}_1^+ - \dot{u}_1^- \\ \tau_{33}^+ - \tau_{33}^- \\ -p_f^+ - (-p_f^-) \\ \tau_{13}^+ - \tau_{13}^- \\ \dot{u}_3^+ - \dot{u}_3^- \\ \dot{w}_3^+ - \dot{w}_3^- \end{bmatrix} = -i\omega h \begin{bmatrix} \mathbf{0} & \mathbf{Q}_{XY} \\ \mathbf{Q}_{YX} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \frac{\dot{u}_1^+ + \dot{u}_1^-}{2} \\ \frac{\tau_{33}^+ + \tau_{33}^-}{2} \\ \frac{-p_f^- + (-p_f^+)}{2}\cdot\Pi + \tilde{B}\left(\frac{\tau_{33}^+ + \tau_{33}^-}{2} - 2G\tilde{\beta}\tilde{\xi}_1\frac{\dot{u}_1^+ + \dot{u}_1^-}{2}\right)\cdot(1-\Pi) \\ \frac{\tau_{13}^+ + \tau_{13}^-}{2} \\ \frac{\dot{u}_3^+ + \dot{u}_3^-}{2} \\ \frac{\dot{w}_3^- + \dot{w}_3^+}{2}\cdot\Pi \end{bmatrix}. \tag{29}$$

The solid displacement (or velocity) u_1, u_3 and total stress τ_{13}, τ_{33} are assumed to vary linearly, because the field changes slowly within the fracture. The above equation is recast in the following form:

$$\begin{bmatrix} \dot{u}_1^+ - \dot{u}_1^- \\ \tau_{33}^+ - \tau_{33}^- \\ -p_f^+ - (-p_f^-) \\ \tau_{13}^+ - \tau_{13}^- \\ \dot{u}_3^+ - \dot{u}_3^- \\ \dot{w}_3^+ - \dot{w}_3^- \end{bmatrix} = -\frac{i\omega h}{2}\begin{bmatrix} \mathbf{0} & \tilde{\mathbf{Q}}_{XY} \\ \tilde{\mathbf{Q}}_{YX} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \dot{u}_1^+ + \dot{u}_1^- \\ \tau_{33}^+ + \tau_{33}^- \\ -p_f^+ + (-p_f^-) \\ \tau_{13}^+ + \tau_{13}^- \\ \dot{u}_3^+ + \dot{u}_3^- \\ \dot{w}_3^+ + \dot{w}_3^- \end{bmatrix}, \tag{30}$$

where

$$\tilde{\mathbf{Q}}_{XY} = \mathbf{Q}_{XY}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \Pi \end{bmatrix} = \begin{bmatrix} 1/G & \xi_1 & 0 \\ \xi_1 & \rho & \rho_f \cdot \Pi \\ 0 & \rho_f & \tilde{\rho} \cdot \Pi \end{bmatrix}, \tag{31}$$

$$\tilde{\mathbf{Q}}_{YX} = \mathbf{Q}_{YX}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2G\tilde{B}\tilde{\beta}\xi_1 \cdot (1-\Pi) & \tilde{B} \cdot (1-\Pi) & \Pi \end{bmatrix}. \tag{32}$$

The components of the matrix $\tilde{\mathbf{Q}}_{YX}$ are given explicitly as

$$\tilde{Q}_{YX}(1,1) = -4G\xi_1^2\left(1 - \frac{G}{H_D}\right) - \frac{\rho_f^2 - \rho\tilde{\rho}}{\tilde{\rho}} - 2G\tilde{B}\tilde{\beta}\xi_1^2\left(-\frac{\rho_f}{\tilde{\rho}} + \alpha\frac{2G}{H_D}\right) \cdot (1-\Pi), \tag{33}$$

$$\tilde{Q}_{YX}(1,2) = \xi_1\left[\left(1 - \frac{2G}{H_D}\right) + \left(-\frac{\rho_f}{\tilde{\rho}} + \alpha\frac{2G}{H_D}\right)\tilde{B} \cdot (1-\Pi)\right], \tag{34}$$

$$\tilde{Q}_{YX}(1,3) = \xi_1\left(-\frac{\rho_f}{\tilde{\rho}} + \alpha\frac{2G}{H_D}\right) \cdot \Pi, \tag{35}$$

$$\tilde{Q}_{YX}(2,1) = \xi_1\left(1 - \frac{2G}{H_D} + 2\tilde{B}\tilde{\beta}\alpha\frac{G}{H_D} \cdot (1-\Pi)\right), \tag{36}$$

$$\tilde{Q}_{YX}(2,2) = \frac{1}{H_D} - \alpha\tilde{B}\frac{1}{H_D} \cdot (1-\Pi), \tag{37}$$

$$\tilde{Q}_{YX}(2,3) = -\alpha\frac{1}{H_D} \cdot \Pi, \tag{38}$$

$$\tilde{Q}_{YX}(3,1) = \xi_1\left[-\frac{\rho_f}{\tilde{\rho}} + \alpha\frac{2G}{H_D} - 2\tilde{B}\tilde{\beta}\left(\alpha^2\frac{G}{H_D} + \frac{G}{M} - \frac{\xi_1^2 G}{\tilde{\rho}}\right) \cdot (1-\Pi)\right], \tag{39}$$

$$\tilde{Q}_{YX}(3,2) = -\alpha\frac{1}{H_D} + \left(\alpha^2\frac{1}{H_D} + \frac{1}{M} - \frac{\xi_1^2}{\tilde{\rho}}\right)\tilde{B} \cdot (1-\Pi), \tag{40}$$

$$\tilde{Q}_{YX}(3,3) = \left(\alpha^2\frac{1}{H_D} + \frac{1}{M} - \frac{\xi_1^2}{\tilde{\rho}}\right) \cdot \Pi. \tag{41}$$

Together, Eqs. (13) and (30) are the seismic boundary conditions for a poroelastic fracture.

C. Simplified boundary conditions and characteristic parameters of a fracture

Mathematically, if the components of the matrices in Eqs. (13) and (30) remain finite when the fracture thickness is reduced to zero, the right-hand side of the equations vanishes, and all the variables are continuous across the fracture. However, in reality, a very thin fracture can produce a large discontinuity in displacement and pressure field if viewed as a boundary. For our model to properly capture this behavior, the material properties of a fracture contained in the matrix boundary conditions in Eqs. (13) and (30) have to take values that result in significantly large matrix components, even when multiplied by the small fracture thickness h. To deal with this situation, we can define composite characteristic parameters of a fracture as a combination of the material properties and the fracture thickness, which control the dynamic behavior of the fracture. Conversely, when physical properties of a fracture are to be determined using seismic waves without the knowledge of the fracture thickness, at best we can determine these composite or "phenomenological" parameters instead of the original material properties, such as bulk permeability and elastic moduli.

From Eqs. (13) and (30), following parameters involving fracture thickness h may be defined:

$$\eta_T \equiv \frac{h}{G} \quad \text{(shear compliance)}, \tag{42}$$

$$\eta_{N_D} \equiv \frac{h}{H_D} \quad \text{(dry or drained normal compliance)}, \tag{43}$$

$$\hat{\kappa}(\omega) \equiv \frac{k(\omega)}{h} \quad \text{(membrane permeability)}. \tag{44}$$

If we assume that these parameters are finite for small fracture thicknesses h's, approximate boundary conditions can be obtained by replacing the moduli and permeability in the equations by the parameters and eliminating $O(h)$ terms. For *Sh* waves, this reduces the coefficient matrix in Eq. (13) to

$$h \times \begin{bmatrix} 0 & 1/G \\ -G\xi_1^2 + (\rho\tilde{\rho} - \rho_f^2)/\tilde{\rho} & 0 \end{bmatrix} \to \eta_T\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}. \tag{45}$$

Therefore, the boundary conditions are

$$\begin{cases} \dot{u}_2^+ - \dot{u}_2^- = (-i\omega)\,\eta_T\tau_{23}^- \\ \tau_{23}^+ = \tau_{23}^- \end{cases}, \tag{46}$$

which are exactly the same as the original linear-slip interface model (Schoenberg, 1980).

For the two coupled matrix boundary conditions in Eq. (30) for fast and slow P waves and an S wave, the coefficient matrices $\tilde{\mathbf{Q}}_{XY}$ and $\tilde{\mathbf{Q}}_{XY}$ in Eqs. (31) and (32), multiplied by h, respectively, reduce to

$$h \times \tilde{\mathbf{Q}}_{XY} \rightarrow \begin{bmatrix} \eta_T & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & i\eta_f/\omega\hat{\kappa}(\omega)\cdot\Pi \end{bmatrix}, \tag{47}$$

$$h \times \tilde{\mathbf{Q}}_{YX} \rightarrow \eta_{N_D} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1-\alpha\tilde{B}(1-\Pi) & -\alpha\cdot\Pi \\ 0 & -\alpha\cdot\Pi & \alpha/\tilde{B}\cdot\Pi \end{bmatrix}, \tag{48}$$

where we used $\tilde{\beta}\approx 1$, which resulted from Eq. (17) through $O(h)\rightarrow 0$. The Skempton coefficient-like parameter in Eq. (16) also reduces to

$$\tilde{B} \approx \alpha\frac{M}{H_U}. \tag{49}$$

Compared to the original Skempton coefficient B, this new coefficient is defined with the undrained P-wave modulus H_U rather than the bulk modulus K_U. Furthermore, the fluid pressure dissipation factor $\Pi(\varepsilon)$ is simplified by approximating the complex diffusion coefficient in Eq. (24) as

$$D \approx \frac{k_0 M}{\eta_f}\left(1-\frac{C^2}{H_U M}\right) = \frac{k_0}{\eta_f}\frac{MH_D}{H_U} = \frac{\tilde{B}\hat{\kappa}_0}{\alpha\eta_f\eta_{N_D}}h^2, \tag{50}$$

which results in

$$\varepsilon = \frac{h}{2\delta_d^*} = \frac{h}{2}(1-i)\sqrt{\frac{\omega}{2D}} = \frac{1-i}{2}\sqrt{\omega\frac{\alpha\eta_f\eta_{N_D}}{2\tilde{B}\hat{\kappa}_0}}. \tag{51}$$

Therefore, for a set of characteristic fracture parameters, $\Pi(\varepsilon)$ also does not depend on the fracture thickness. Note that in deriving Eq. (48), the matrix components in Eqs. (33)–(41) containing $h/\tilde{\rho}=\omega k(\omega)h/i\eta_f$ were ignored, even for fractures with very high (static) permeability. This is because the dynamic permeability of a fracture is finite even when the static permeability of the material within the fracture approaches infinity, which results in an explicit bound $|h/\tilde{\rho}| \le h/\rho_f$ (Appendix B).

Using the simplified relationships in Eqs. (47)–(51), the boundary conditions are written explicitly as

$$\begin{cases} \dot{u}_1^+ - \dot{u}_1^- = (-i\omega)\eta_T\tau_{13}^- \\ \dot{u}_3^+ - \dot{u}_3^- = (-i\omega)\eta_{N_D}\left[\left(1-\alpha\tilde{B}(1-\Pi)\right)\tau_{33}^- - \alpha\dfrac{-p_f^+ + (-p_f^-)}{2}\cdot\Pi\right] \\ \dot{w}_3^+ - \dot{w}_3^- = (-i\omega)\alpha\eta_{N_D}\left[-\tau_{33}^- + \dfrac{1}{\tilde{B}}\dfrac{-p_f^+ + (-p_f^-)}{2}\right]\cdot\Pi \\ \tau_{13}^+ = \tau_{13}^- \\ \tau_{33}^+ = \tau_{33}^- \\ -p_f^+ - (-p_f^-) = \dfrac{\eta_f}{\hat{\kappa}(\omega)}\dfrac{\dot{w}_3^+ + \dot{w}_3^-}{2}\cdot\Pi \end{cases}. \tag{52}$$

An important feature of these boundary conditions is that they do not explicitly contain the plane-parallel slowness ξ_1. This allows us to use Eqs. (46) and (52) for plane waves at any angle of incidence or in the spatial domain of numerical models, such as finite-difference and finite-element models. The last equation for pressure discontinuity in Eq. (52) can be viewed as a generalization of the results from Gurevich and Schoenberg (1999) for Darcy's law extended to a single finite permeability interface. Therefore, from Eqs. (46) and (52), the five fundamental characteristic parameters of a poroelastic fracture are the dry shear and normal fracture compliances η_T, η_{Nd}, membrane permeability $\hat{\kappa}(\omega)$, the fracture Biot-Willis effective stress coefficient α, and the fracture Skempton coefficient $\tilde{B}$. From Eqs. (42) and (43), the dry normal fracture compliance cannot exceed the shear fracture compliance because $H_D=K_D+4G/3\ge G$. This restriction arises because we have assumed that the fracture-filling medium is isotropic. If the layer modeling a fracture is allowed to be transversely isotropic, however, the two compliance parameters can be independent, whereas the same five characteristic parameters of a fracture can be used to describe the boundary conditions in Eqs. (46) and (52) (Appendix A).

The high-permeability limit (open fracture) of Eq. (52) is obtained by taking the limit $\hat{\kappa}_0(\equiv\hat{\kappa}(0))\rightarrow\infty$ ($k_0\rightarrow\infty$ for any h). Using the result $\tilde{\rho}\rightarrow\rho_f$ (Appendix B), $|\eta_f/\hat{\kappa}(\omega)| = |\tilde{\rho}(\omega)|\omega h \rightarrow \rho_f\omega h$, which vanishes for small h's. Because $\Pi\rightarrow 1$, the equations reduce to

$$\begin{cases} \dot{u}_1^+ - \dot{u}_1^- = (-i\omega)\eta_T\tau_{13}^- \\ \dot{u}_3^+ - \dot{u}_3^- = (-i\omega)\eta_{N_D}[\tau_{33}^- - \alpha(-p_f^-)] \\ \dot{w}_3^+ - \dot{w}_3^- = (-i\omega)\alpha\eta_{N_D}[-\tau_{33}^- + (1/\tilde{B})(-p_f^-)] \\ \tau_{13}^+ = \tau_{13}^- \\ \tau_{33}^+ = \tau_{33}^- \\ -p_f^+ = -p_f^- \end{cases}. \tag{53}$$

This is essentially the same result as the boundary conditions derived by Bakulin and Molotkov (1997).

In contrast, the low-permeability limit (impermeable fracture) is obtained by $\hat{\kappa}_0 \to 0$ ($k_0 \to 0$ for any h). Because $\Pi \to O(1/\varepsilon) = O(\sqrt{\hat{\kappa}_0})$ and $1/\hat{\kappa}(\omega) \to 1/\hat{\kappa}_0$, from the third and sixth equations in Eq. (52), $\dot{w}_3^+ - \dot{w}_3^- \to O(\sqrt{\hat{\kappa}_0})$ and $\dot{w}_3^+ + \dot{w}_3^- \to O(\sqrt{\hat{\kappa}_0})$, $\dot{w}_3^+ = \dot{w}_3^- = 0$.As a result, we obtain

$$\begin{cases} \dot{u}_1^+ - \dot{u}_1^- = (-i\omega)\eta_T \tau_{13}^- \\ \dot{u}_3^+ - \dot{u}_3^- = (-i\omega)\eta_{N_U} \tau_{33}^- \\ \dot{w}_3^+ = \dot{w}_3^- = 0 \\ \tau_{13}^+ = \tau_{13}^- \\ \tau_{33}^+ = \tau_{33}^- \end{cases}. \tag{54}$$

In Eq. (54), the undrained normal fracture compliance is defined as a derived new fracture parameter by

$$\eta_{N_U} \equiv \frac{h}{H_U} = \eta_{N_D}(1 - \alpha\tilde{B}). \tag{55}$$

For a compliant, fluid-saturated fracture, $1/\tilde{B} \approx 1/B \approx \alpha$, and Eqs. (53) and (54) can be simplified even further.

Although assuming a vanishingly small fracture thickness h results in simple boundary conditions, in reality a finite h may result in non-negligible effects because of the neglected $O(h)$ terms in the matrices. This error will be examined briefly in the examples given later in Sec. III B.

D. Plane-wave transmission and reflection coefficients

In applying the obtained boundary conditions, we will derive explicit expressions for the transmission and reflection coefficients of plane waves scattered by a poroelastic fracture. From the velocity and stress components used in the equation, Eq. (13) can be used for the scattering of S waves with fracture-parallel particle motions (Sh waves), and Eq. (30) can be used for the scattering of fast and slow P waves, as well as for S waves with particle motions within the plane of wave propagation (Sv waves). In the following, we will first examine the P-Sv case.

First, we split the second matrix boundary conditions in Eq. (30) into the following two coupled equations:

$$\mathbf{b}_X(0_+) - \mathbf{b}_X(0_-) = -i\omega\frac{h}{2}\tilde{\mathbf{Q}}_{XY}(\mathbf{b}_Y(0_+) + \mathbf{b}_Y(0_-)), \tag{56}$$

$$\mathbf{b}_Y(0_+) - \mathbf{b}_Y(0_-) = -i\omega\frac{h}{2}\tilde{\mathbf{Q}}_{YX}(\mathbf{b}_X(0_+) + \mathbf{b}_X(0_-)), \tag{57}$$

$$\mathbf{b}_X \equiv \begin{bmatrix} \dot{u}_1 \\ \tau_{33} \\ -p_f \end{bmatrix}, \tag{58}$$

$$\mathbf{b}_Y \equiv \begin{bmatrix} \tau_{13} \\ \dot{u}_3 \\ \dot{w}_3 \end{bmatrix}, \tag{59}$$

where 0_- indicates the incident side of the fracture and 0_+ is the transmitted side of the fracture. For the matrices $h \times \tilde{\mathbf{Q}}_{XY}$ and $h \times \tilde{\mathbf{Q}}_{YX}$, either the original boundary conditions in Eqs. (31) and (32) or simplified conditions in Eqs. (47) and (48) can be used. The vector variables are decomposed into incident (I), transmitted (T), and reflected (R) fields as

$$\mathbf{b}_X(0_+) = \mathbf{b}_X^T(0_+) = -i\omega\mathbf{X}^+\mathbf{a}^T, \tag{60}$$

$$\mathbf{b}_X(0_-) = \mathbf{b}_X^I(0_-) + \mathbf{b}_X^R(0_-) = -i\omega(\mathbf{X}^+\mathbf{a}^I + \mathbf{X}^-\mathbf{a}^R), \tag{61}$$

$$\mathbf{b}_Y(0_+) = \mathbf{b}_Y^T(0_+) = -i\omega\mathbf{Y}^+\mathbf{a}^T, \tag{62}$$

$$\mathbf{b}_Y(0_-) = \mathbf{b}_Y^I(0_-) + \mathbf{b}_Y^R(0_-) = -i\omega(\mathbf{Y}^+\mathbf{a}^I + \mathbf{Y}^-\mathbf{a}^R). \tag{63}$$

The vectors $\mathbf{b}_{X,Y}^{I,T,R}$ are expressed via coefficient vectors $\mathbf{a}^I$, $\mathbf{a}^T$, and $\mathbf{a}^R$ containing complex amplitudes of solid frame displacement for fast P wave (P_f), slow P wave (P_s), and S wave (S) as their three components (for example, $\mathbf{a}^I = [a_{Pf}^I, a_{Ps}^I, a_S^I]^T$, where the superscript T here indicates vector transpose). The coefficient matrices containing normalized displacement and stress components of these waves in each column are given by

$$\mathbf{X}^{\pm} \equiv \begin{bmatrix} \xi_1/\xi_{Pf} & \xi_1/\xi_{Ps} & \xi_3^S/\xi_S \\ -\xi_{Pf}(H_U^B + f_{Pf}C^B) + 2\xi_1^2 G^B/\xi_{Pf} & -\xi_{Ps}(H_U^B + f_{Ps}C^B) + 2\xi_1^2 G^B/\xi_{Ps} & 2\xi_1\xi_3^S G^B/\xi_S \\ -\xi_{Pf}(C^B + f_{Pf}M^B) & -\xi_{Ps}(C^B + f_{Ps}M^B) & 0 \end{bmatrix} \equiv \mathbf{X} \tag{64}$$

$$\mathbf{Y}^{\pm} \equiv \pm \begin{bmatrix} -2\xi_1\xi_3^{Pf}G^B/\xi_{Pf} & -2\xi_1\xi_3^{Ps}G^B/\xi_{Ps} & -(\xi_S^2 - 2\xi_1^2)G^B/\xi_S \\ \xi_3^{Pf}/\xi_{Pf} & \xi_3^{Ps}/\xi_{Ps} & -\xi_1/\xi_S \\ f_{Pf}\xi_3^{Pf}/\xi_{Pf} & f_{Ps}\xi_3^{Ps}/\xi_{Ps} & -f_S\xi_1/\xi_S \end{bmatrix} \equiv \pm\mathbf{Y} \tag{65}$$

The expressions for the displacement and stress components can be found in, for example, Pride *et al.* (2002). The superscripts + and − indicate waves propagating in the $+x_3$ and $-x_3$ directions, respectively. Arranging displacement and stress components of plane waves as in these matrices has been shown to result in particularly simple expressions for plane-wave transmission and reflection coefficients for materials with "up-down symmetry" across a plane scattering interface (Schoenberg and Protazio, 1992). In the matrices $\mathbf{X}$ and $\mathbf{Y}$, all the slowness components are for the background

TABLE I. Baseline material properties used for the numerical examples are shown. Although some of the values in the table may seem unrealistic for natural fractures, these values are assumed to reduce the number of free parameters in the study.

Matrix properties	Values	Fracture properties	Values
Porosity	0.15	Porosity	0.5
Permeability	1.0×10^{-13} m^2 or 100 mD		
Solid bulk modulus	36.0×10^{9} Pa		
Fluid bulk modulus	2.25×10^{9} Pa		
Frame bulk modulus	9.0×10^{9} Pa	Dry normal compliance	1.0×10^{-11} m/Pa
Frame shear modulus	7.0×10^{9} Pa	Shear compliance	3.0×10^{11} m/Pa
Solid density	2700 kg/m^3	Solid density	2700 kg/m^3
Fluid density	1000 kg/m^3	Fluid density	1000 kg/m^3
Fluid viscosity	1.0×10^{-3} Pa s	Fluid viscosity	1.0×10^{-3} Pa s
Tortuosity	3	Tortuosity	1
Saturation ratio	1		

medium, and coefficients f_{Pf} and f_{Ps} and f_S are the complex-valued ratios of the relative fluid displacement to the solid-frame displacement for fast and slow P waves and the Sv waves, respectively (e.g., Pride *et al.*, 2002). To avoid confusion, moduli for the background medium H_U^B, M^B, and C^B are indicated by a superscript B. Also, all the slowness components are associated with the background medium.

Introducing Eqs. (60)–(65) into Eqs. (56) and (57) results in

$$\mathbf{X}(\mathbf{a}^T-\mathbf{a}^I-\mathbf{a}^R)=-i\omega\frac{h}{2}\tilde{\mathbf{Q}}_{XY}\mathbf{Y}(\mathbf{a}^T+\mathbf{a}^I-\mathbf{a}^R), \tag{66}$$

$$\mathbf{Y}(\mathbf{a}^T-\mathbf{a}^I+\mathbf{a}^R)=-i\omega\frac{h}{2}\tilde{\mathbf{Q}}_{YX}\mathbf{X}(\mathbf{a}^T+\mathbf{a}^I+\mathbf{a}^R). \tag{67}$$

By solving these equations for the unknown coefficient vectors $\mathbf{a}^T$ and $\mathbf{a}^R$, the transmission and reflection coefficient matrices $\mathbf{T}$, $\mathbf{R}$ are determined, respectively, as

$$\mathbf{a}^T=\left[\left(\mathbf{I}+\frac{i\omega h}{2}\mathbf{Y}^{-1}\tilde{\mathbf{Q}}_{YX}\mathbf{X}\right)^{-1}+\left(\mathbf{I}+\frac{i\omega h}{2}\mathbf{X}^{-1}\tilde{\mathbf{Q}}_{XY}\mathbf{Y}\right)^{-1}-\mathbf{I}\right]\mathbf{a}^I\equiv\mathbf{T}\mathbf{a}^I, \tag{68}$$

$$\mathbf{a}^R=\left[\left(\mathbf{I}+\frac{i\omega h}{2}\mathbf{Y}^{-1}\tilde{\mathbf{Q}}_{YX}\mathbf{X}\right)^{-1}-\left(\mathbf{I}+\frac{i\omega h}{2}\mathbf{X}^{-1}\tilde{\mathbf{Q}}_{XY}\mathbf{Y}\right)^{-1}\right]\mathbf{a}^I\equiv\mathbf{R}\mathbf{a}^I. \tag{69}$$

By recognizing the same structure in Eqs. (13) and (30), the same procedure can be followed to determine the scattering coefficients for Sh waves. This can be done by simple substitutions $\mathbf{X}\to1$, $\mathbf{Y}\to-\xi_3^SG^B$, $\tilde{\mathbf{Q}}_{XY}\to1/G$, $\tilde{\mathbf{Q}}_{YX}\to-G\xi_1^2+(\rho\tilde{\rho}-\rho_f^2)/\tilde{\rho}$, $\mathbf{I}\to1$ in Eqs. (68) and (69), resulting, respectively, in

$$a^T=\left[\left(1+\frac{i\omega h}{2}\frac{G\xi_1^2-\rho+\rho_f^2/\tilde{\rho}}{\xi_3^SG^B}\right)^{-1}+\left(1-\frac{i\omega h}{2}\frac{\xi_3^SG^B}{G}\right)^{-1}-1\right]a^I\equiv Ta^I, \tag{70}$$

$$a^R=\left[\left(1+\frac{i\omega h}{2}\frac{G\xi_1^2-\rho+\rho_f^2/\tilde{\rho}}{\xi_3^SG^B}\right)^{-1}-\left(1-\frac{i\omega h}{2}\frac{\xi_3^SG^B}{G}\right)^{-1}\right]a^I\equiv Ra^I. \tag{71}$$

T and R are the transmission and reflection coefficients for Sh waves, respectively.

The general expressions for the transmission and reflection coefficients for poroelastic fractures will be used in the following examples, to examine the accuracy of both the original and simplified boundary conditions.

III. EXAMPLES AND DISCUSSIONS

In this section, we will examine the effects of some of the fracture parameters on seismic wave scattering. The models for the examples share a set of baseline material properties shown in Table I, which are intended to be for a "typical" sandstone (e.g., Berea) containing a compliant fracture. Also, we will focus on the amplitudes of transmitted and reflected fast and slow P waves and an S wave generated by an incident fast P wave. No phase responses are examined.

The accuracy of the derived boundary conditions is assessed by computing the transmission and reflection coefficients using Eqs. (68) and (69) and comparing the results to the prediction of the Kennett's reflectivity algorithm (Kennett, 1983; for poroelastic wave propagation, see Pride *et al.*, 2002). Since the Kennett algorithm computes the scattering coefficients without approximation (Appendix C), the results are considered to be the correct solution.

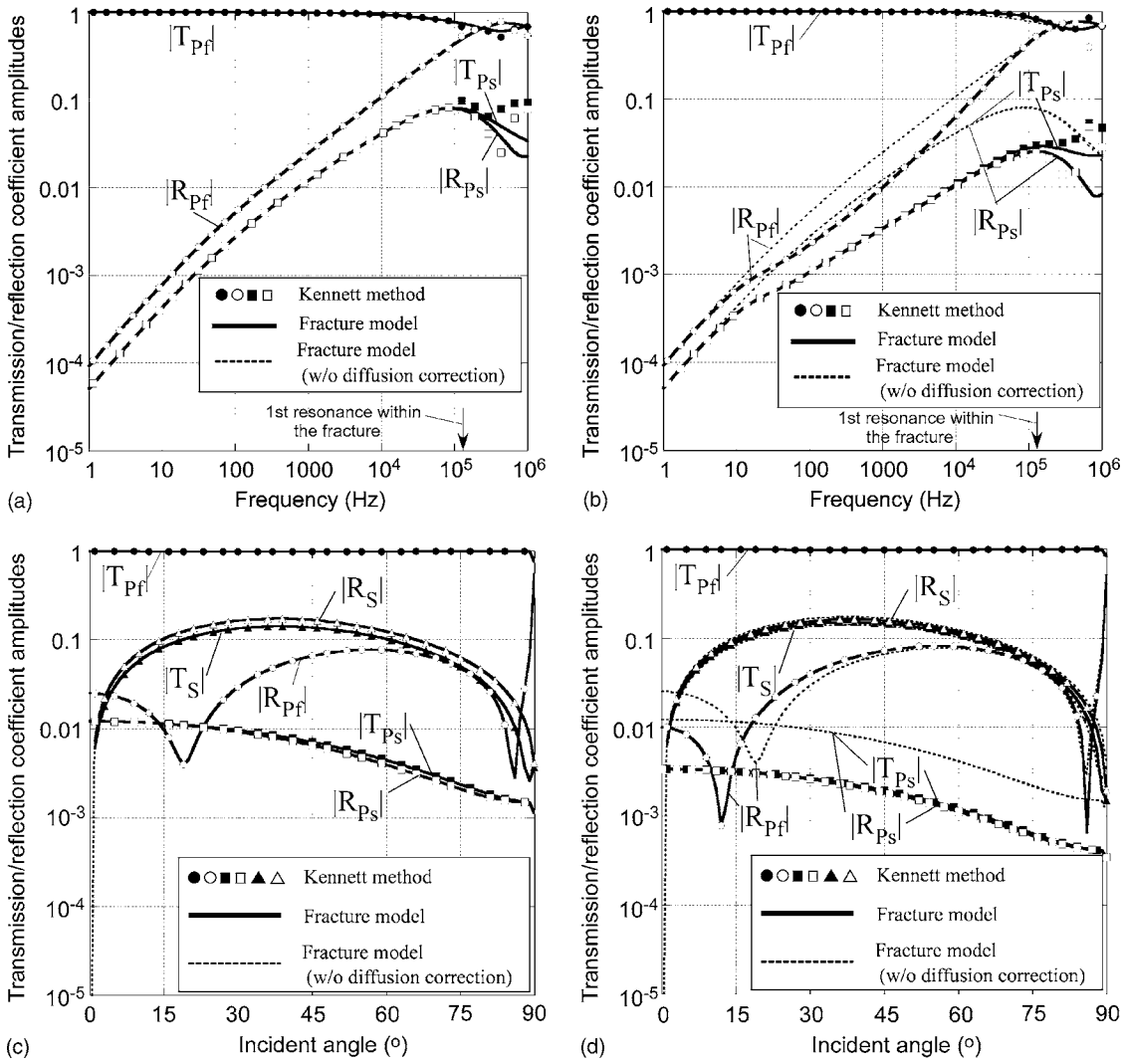


FIG. 5. Amplitudes of the displacement transmission and reflection coefficients for incident fast P waves computed for both low-permeability (1/1000× background) and high-permeability (1000× background) fractures with a thickness $h=1$ mm. The labels indicate T=transmission coefficient and R=reflection coefficient, with subscript Pf=fast P wave, Ps=slow P wave, and $S=S$ wave. (Discrete symbols were computed by the Kennett method.) Solid curves and dotted curves were computed using the full fracture model in Eq. (30) with and without the correction of the pressure diffusion by the fluid pressure dissipation factor Π, respectively. For the high-permeability fracture, the correction is negligible, and both models agree very well with the "correct solution" computed by Kennett method. However, for the low-permeability fracture, the model without the correction shows significant errors. (a) Normal incidence frequency response. $k_0=10^{-10}$ m^2; (b) Normal incidence frequency response. $k_0=10^{-16}$ m^2; (c) 1-kHz angle-of-incidence response. $k_0=10^{-10}$ m^2; (d) 1-kHz angle-of-incidence response. $k_0=10^{-16}$ m^2.

A. Impact of pressure diffusion within a fracture

In the first example, we examine the accuracy of the poroelastic fracture boundary conditions in Eq. (30) and the impact of the fluid pressure dissipation factor Π in Eq. (21) on wave scattering. For this example, in addition to the properties shown in Table I, we assume both high fracture permeability $k_0=10^{-10}$ m^2 (100 D; 1000 times the background permeability) and low fracture permeability $k_0=10^{-16}$ m^2 (0.1 mD; 0.001 times the background permeability), with a fracture thickness $h=1$ mm. The fracture is fully saturated with the same fluid as the background, and the bulk modulus of the solid (grains) is also the same as the background.

Both normal-incidence frequency responses for a frequency range of 10 Hz to 1 MHz—Figs. 5(a) and 5(b)—and angle-of-incidence responses at 1 kHz—Figs. 5(c) and 5(d)—show very good agreement between the Kennett algorithm (shown in discrete symbols) and the full fracture model (shown in thick solid lines). The errors seen above 100 kHz for the normal incidence case result primarily from the multiple scattering of waves within the fracture (layer), which is not accounted for by the fracture model. An approximate frequency corresponding to the lowest-frequency resonance (reverberation) of the fast P wave within the fracture is indicated in the plots by an arrow.

When the effect of pressure diffusion within a fracture is ignored by enforcing the fluid-pressure dissipation factor $\Pi=1$ in Eqs. (30)–(41), the fracture model (shown by dotted lines in Fig. 5) significantly overestimates the reflected fast P wave [Figs. 5(b) and 5(d)] and scattered slow P waves for a low-permeability fracture above a frequency near 10 Hz. This frequency is a transient (critical) frequency that separates the drained and undrained response of a fracture. (More detailed discussion will be given later in Sec. III C.) In con-

trast, for a high-permeability fracture, ignoring the effect of pressure diffusion does not result in noticeable errors [Figs. 5(a) and 5(c)].

B. Fracture thickness and the accuracy of fracture models

We derived the simplified fracture model in Eq. (67) by assuming that the $O(h)$ terms in the original boundary conditions can be ignored except for the characteristic parameters of a fracture. For a finite fracture thickness, however, this assumption has to be scrutinized.

In the following example, we assume a set of characteristic fracture parameters $\eta_T = 3 \times 10^{-11}$ m/Pa, $\eta_{N_D} = 1 \times 10^{-11}$ m/Pa, $\alpha = 0.85$, $\tilde{B} = 0.29$, and $\hat{\kappa}_0 = k_0/h = 1 \times 10^{-13}$ m, and examine the effect of fracture thickness on the wave scattering for three thickness values: $h = 1$ mm, 1 cm, and 10 cm. These characteristic parameters were chosen for the typical physical parameters in Table I, assuming that the 10-cm-thick fracture was 100% saturated by the same fluid as the background. (The bulk modulus of the fluid in the thinner fractures has to be reduced to maintain the same α and $\tilde{B}$ values, which can be realized physically by introducing a small amount of gas in the fluid.) Elastic properties of the material within the fracture are determined from these parameters as a function of fracture thickness and used in the full fracture model in Eqs. (30)–(41) as well as the layer model.

In this example, we also examine the accuracy of the simplified fracture model in Eq. (52) compared to the full fracture model and the layer solution of Kennett's reflectivity algorithm. Reflection and transmission coefficient amplitudes for fast and slow P waves generated from normally incident fast P waves are shown in Fig. 6. For thin fracture thickness h below 1 cm, both full and simplified fracture models agree very well with the layer model, with the upper limit of applicable range of frequency reducing with increasing h. The error becomes large near and above the first resonance frequency of the fast P wave within the fracture, as indicated by a gray vertical line in the plots. The resonance frequencies for this example are lower than the previous example, because the combination of material properties used here yields large undrained fracture compliance, which results in slower fast P-wave velocity within the fracture. Further, for the $h = 10$-cm case, the reflected fast P wave is inaccurately predicted by the simplified model, even well below the resonant frequency. This is probably caused by the effect of mass within the layer. This effect is neglected in the simplified model, since the scattering of slow P waves, which is predominantly governed by the diffusion of fluid pressure within the pore space, is still accurately predicted.

Because the characteristic fracture parameters are not dependent on fracture thickness, the simplified model shows identical responses in Figs. 6(a)–6(c) (shown by thin solid lines). The slight differences in the transmitted slow P-wave response predicted by the model in Fig. 6(c) compared to Figs. 6(a) and 6(b) result from the complex dependence of

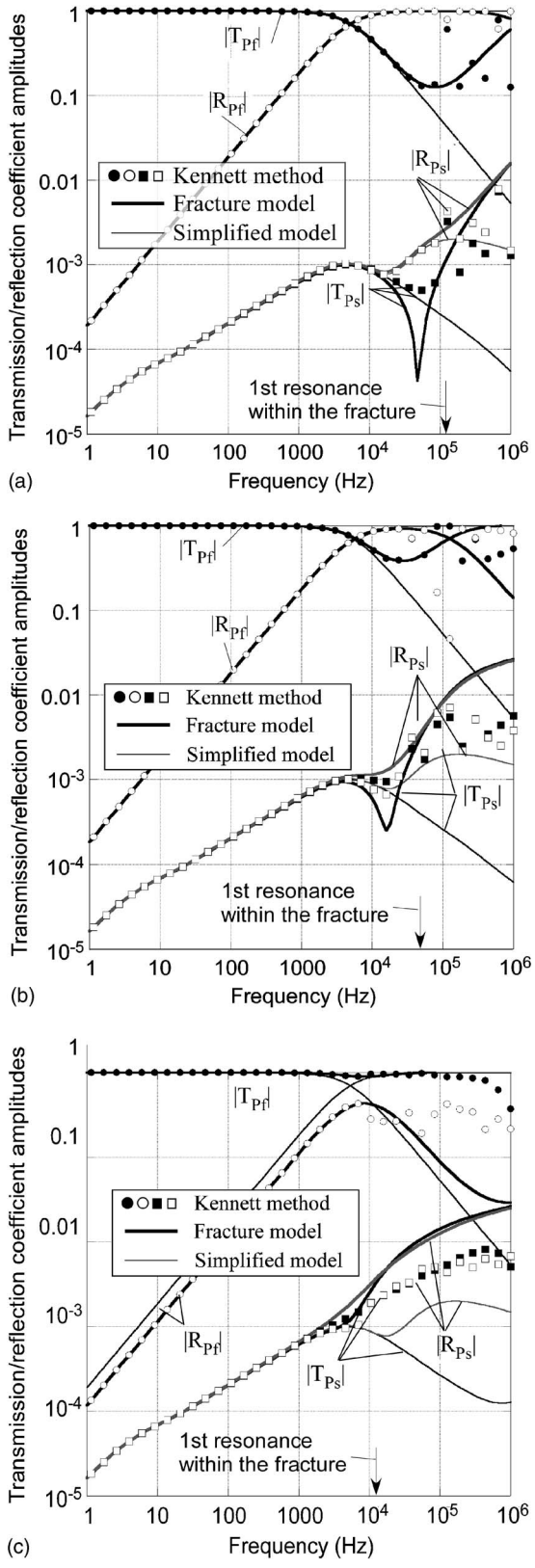


FIG. 6. Normal-incidence reflection and transmission coefficient amplitudes for fast and slow P waves for the same characteristic fracture parameters and three values of fracture thickness. For thin fracture thickness (below 1 cm), both full and simplified fracture models agree very well with the layer model, with the upper limit of the applicable range of frequency reducing with increasing h. For the $h = 10$ cm case, the reflected fast P wave is inaccurately predicted by the simplified model, possibly because of the effect of mass within the layer, which is not considered. (a) $h = 1$ mm; (b) $h = 1$ cm; (c) $h = 10$ cm.

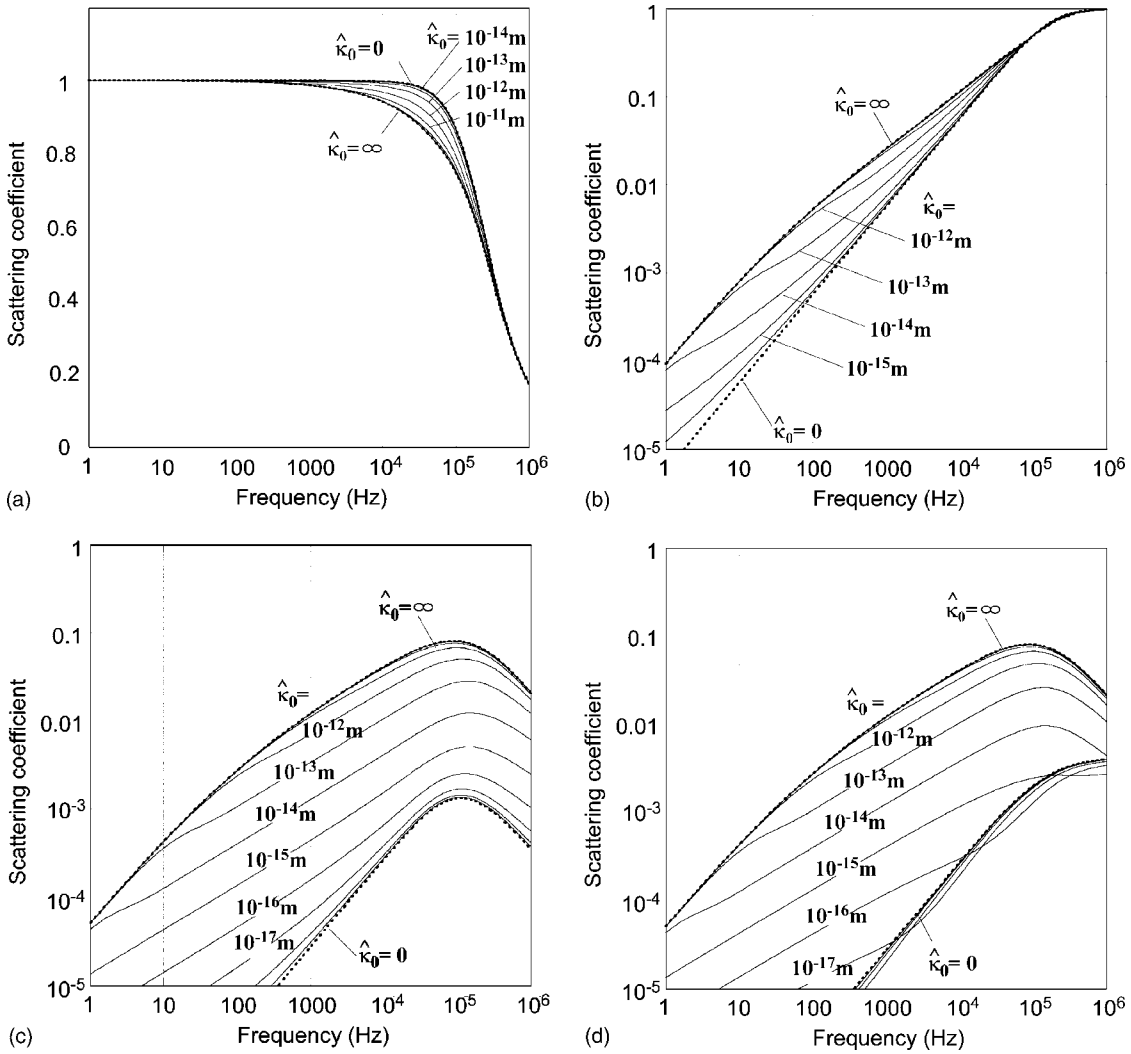


FIG. 7. Scattering amplitude responses (scattering coefficients) for a range of membrane permeability of a fracture. Fast *P*-wave transmission coefficients are shown with semi-log scales to clearly show changes at large amplitudes. Dotted lines are both for a fracture with infinite permeability and an impermeable fracture. Different membrane permeability values correspond to different type curves that "saturate" at both low and high permeability values. (a) Fast *P*-wave transmission; (b) Fast *P*-wave reflection; (c) Slow *P*-wave transmission; (d) Slow *P*-wave reflection.

dynamic permeability on h at high frequencies for a given static membrane permeability [defined by Eq. (44), with a frequency $\omega=0$].

C. Effect of fracture permeability on seismic wave scattering

In the third example, we repeat our experiment of the first example to further examine the effect of fracture hydraulic permeability (or membrane permeability). The material properties and fracture thicknesses used here are the same as in the first example ($h=1$ mm), which results in characteristic fracture parameters $\eta_T=3\times10^{-11}$ m/Pa, $\eta_{N_D}=1\times10^{-11}$ m/Pa, $\alpha=0.998$, and $\tilde{B}=0.98$.

The scattering amplitudes are computed using the simplified fracture model in Eq. (52) for a wide range of membrane permeability values (Fig. 7). To examine the behavior of the waves more closely, each wave and scattering mode is shown separately. We also compute the high-and low-permeability limits of the scattering responses using Eq. (53) and Eq. (54), respectively, which are shown in thick dotted lines bounding finite-permeability responses (except for the zero-permeability bound in the slow *P*-wave reflection in Fig. 7(d), for which the scattering response of a low-permeability fracture is more complicated and undershoots the low-permeability limit).

A distinct characteristic of the reflected fast *P* wave and both transmitted and reflected slow *P* waves is that the slope changes when each frequency response curve departs from that of the high-permeability limit (labeled as $\hat{\kappa}_0=\infty$). For example, for $\hat{\kappa}_0=10^{-13}$ m [corresponding to the case previously shown in Fig. 5(b)], this occurs near 10 Hz. For a saturated fracture with high, dry compliance, this transition (critical) frequency can be evaluated as follows:

From Eq. (51),

$$\varepsilon \approx \frac{1-i}{2}\sqrt{\omega\frac{\alpha^2\eta_f\eta_{N_D}}{2\hat{\kappa}_0\tilde{B}}}. \tag{72}$$

From the behavior of $\Pi(\varepsilon)$ shown in Fig. 3, the transition frequency (critical frequency) ω_d between the drained and undrained responses of a low-permeability fracture is evalu-

ated by a frequency corresponding to $\mathrm{Re}\{\varepsilon\}=1$, resulting in

$$\omega_d = \frac{\hat{\kappa}_0}{\eta_f}\frac{8\tilde{B}}{\alpha^2 \eta_{N_D}}, \tag{73}$$

which is proportional to the membrane permeability of a fracture and inversely proportional to the dry normal compliance of the fracture. When the parameter values used in this example are introduced with $\hat{\kappa}_0 = 10^{-13}$ m, the characteristic frequency is $f_d = \omega_d / 2\pi = 12.7$ Hz, which is close to the 10 Hz observed in both Fig. 5(b) and Figs. 7(b)–7(d). Also, ω_d can be viewed as a critical frequency below which the reflection of fast P waves and the scattering of slow P waves become insensitive to the changes in fracture permeability. Conversely, if the permeability of the fracture is low, for a given frequency, membrane permeability higher than the following critical permeability cannot be determined using seismic waves from the scattering of plane waves,

$$\hat{\kappa}_{0c} = \frac{\omega \eta_f \alpha^2 \eta_{N_D}}{8}. \tag{74}$$

At 1 kHz, the critical membrane permeability is $\hat{\kappa}_{0c} = 7.9 \times 10^{-12}$ m.

IV. CONCLUSIONS

A fluid-filled, flat fracture is a special case of heterogeneous poroelastic media, for which the effect of poroelastic material properties on discrete scattering of seismic waves can be examined analytically, owing to its simple geometry. We hypothesize that a compliant fracture can be viewed as a flat, thin, soft inclusion within a matrix. This simplification results in sets of boundary conditions relating a finite jump in the stress and velocity across a fracture to the stress and velocity at the boundaries (fracture surfaces).

The key step in the derivation of the boundary conditions is the approximation of the pressure field within a fracture: although the thickness of a fracture can usually be considered much smaller than the wavelength of an incoming wave (fast P wave and S wave), the pressure diffusion length (or the wavelength of the generated slow wave) within the fracture can be comparable to the fracture thickness, resulting in a rapid change in the pressure distribution. In turn, this complex pressure distribution due to diffusion affects how the wave is scattered, as a function of permeability and fluid properties within the fracture.

For a thin fracture, however, the permeability parallel to the fracture cannot be resolved from the wave scattering, as indicated by the results in Appendix A. In this case, the permeability needs to be inferred indirectly from the dry and wet fracture compliances—parameters which depend on a fracture's internal structure (such as porosity, asperity, contact spacing), which also affects the permeability. In contrast, fracture-normal permeability can affect wave scattering if the permeability is below a threshold value and the wave frequency is above a critical frequency.

Typically, the scattering behavior of a fracture changes at a frequency where the fluid-solid interaction within the fracture changes between drained (low frequency and high permeability) and undrained (high frequency and low permeability) regimes. In general, for a normally incident fast P wave, a fracture with higher fracture-normal permeability exhibits larger reflection of fast P waves and generates more slow P waves. However, amplitudes of slow P waves generated by a single fracture are generally small. For the effect to be clearly measurable, high-frequency seismic waves and/or multiple fractures may be necessary.

The scattering of waves by a fracture is controlled by a set of characteristic (phenomenological) parameters similar to the fracture compliance used in the linear slip interface model (Schoenberg, 1980). These parameters are shear compliance, drained normal compliance, Biot-Willis effective stress coefficient, fracture Skempton coefficient, and membrane permeability. For a sufficiently small fracture thickness, fractures having identical parameter values result in the same observed seismic response.

Finally, throughout the modeling presented in this paper, a fracture is assumed to be an isotropic and homogeneous layer. (An extension of the model to anisotropic elastic moduli and permeability is presented in Appendix A.) The question remains, can an open fracture with partial surface contacts and a fault with complex internal geometry be modeled with such a simple model? For example, scattering of waves may be strongly affected by the local fluid motion around contacting asperities and within the complex internal structure of a well-developed fault originating from shearing (e.g., Sibson, 1977). For such cases, more complex boundary conditions, considering the effect of internal heterogeneity of a fracture, are necessary.

ACKNOWLEDGMENTS

This research has been supported by the Office of Science, Office of Basic Energy Sciences, Division of Chemical Sciences of the U.S. Department of Energy under Contract No. DE-AC76SF00098. The authors would like to thank Dr. Steven Pride at Lawrence Berkeley National Laboratory for many useful suggestions and discussions during the development of the models presented in this article.

APPENDIX A: DERIVATION OF SEISMIC BOUNDARY CONDITIONS FOR A TRANSVERSELY ISOTROPIC POROELASTIC FRACTURE

Constitutive relationships for a general anisotropic poroelastic medium can be written using index notations as (Cheng, 1997)

$$\tau_{ij} = C^D_{ijkl} u_{k,l} + \alpha_{ij}(-p_f) = C^U_{ijkl} u_{k,l} + M \alpha_{ij} w_{k,k}, \tag{A1}$$

$$-p_f = M(w_{k,k} + \alpha_{ij} u_{i,j}), \tag{A2}$$

where α_{ij} is the symmetric Biot-Willis effective stress coefficient tensor and C^D_{ijkl} and $C^U_{ijkl} = C^D_{ijkl} + M\alpha_{ij}\alpha_{kl}$ are the dry (drained) and undrained stiffness tensors for the solid frame, respectively. M is the fluid storage modulus. The momentum balance equations are

$$\tau_{ij,j} = -\omega^2(\rho u_i + \rho_f w_i), \tag{A3}$$

$$-p_{f,i} = -\omega^2(\rho_f u_i + \tilde{\rho}_{ij} w_j), \tag{A4}$$

where $\tilde{\rho}_{ij}$ is defined via an anisotropic dynamic permeability tensor $\mathbf{k}(\omega)$ through $\tilde{\boldsymbol{\rho}} \equiv (i\eta_f/\omega)\mathbf{k}^{-1}(\omega)$.

For the following, we will focus on the transversely isotropic case with the axis of symmetry aligned in the 3 direction (fracture-normal direction). In the reduced matrix notation, the above constitutive relationship becomes

$$\begin{bmatrix} \tau_{11} \\ \tau_{22} \\ \tau_{33} \\ \tau_{23} \\ \tau_{31} \\ \tau_{12} \end{bmatrix} = \begin{bmatrix} C_{11}^D & C_{12}^D & C_{13}^D & & & \\ C_{12}^D & C_{11}^D & C_{13}^D & & & \\ C_{13}^D & C_{13}^D & C_{33}^D & & & \\ & & & G & & \\ & & & & G & \\ & & & & & G' \end{bmatrix} \begin{bmatrix} u_{1,1} \\ u_{2,2} \\ u_{3,3} \\ u_{2,3} + u_{3,2} \\ u_{3,1} + u_{1,3} \\ u_{1,2} + u_{2,1} \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_1 \\ \alpha_3 \\ 0 \\ 0 \\ 0 \end{bmatrix} (-p_f), \tag{A5}$$

$$-p_f = M(\alpha_1 u_{1,1} + \alpha_1 u_{2,2} + \alpha_3 u_{3,3}) + M(w_{1,1} + w_{2,2} + w_{3,3}), \tag{A6}$$

where $G' = (C_{11}^D - C_{12}^D)/2$, and α_i $(i=1, 3)$ are the diagonal entries of the effective stress coefficient tensor α. The momentum balance equation is the same as the general anisotropic case, except that the permeability tensor also becomes diagonal: $\mathbf{k}(\omega) = \mathrm{diag}[k_1(\omega), k_1(\omega), k_3(\omega)]$, where each diagonal component can be computed using the dynamic permeability model proposed by Johnson *et al.* (1987). Following the same procedure as in the isotropic case, we obtain a counterpart to the governing equations Eqs. (5) and (6) with coefficient matrices,

$$\mathbf{R} \equiv \begin{bmatrix} & 1/G \\ -G'\xi_1^2 + \rho - \dfrac{\rho_f^2}{\tilde{\rho}_1} & \end{bmatrix}, \tag{A7}$$

$$\mathbf{Q}_{XY} \equiv \begin{bmatrix} 1/G & \xi_1 & 0 \\ \xi_1 & \rho & \rho_f \\ 0 & \rho_f & \tilde{\rho}_3 \end{bmatrix}, \tag{A8}$$

$$\mathbf{Q}_{YX} \equiv \begin{bmatrix} \rho - \dfrac{\rho_f^2}{\tilde{\rho}_1} - \left(C_{11}^D - \dfrac{C_{13}^{D2}}{C_{33}^D}\right)\xi_1^2 & \xi_1 \dfrac{C_{13}^D}{C_{33}^D} & \xi_1\left(\alpha_1 - \alpha_3 \dfrac{C_{13}^D}{C_{33}^D} - \dfrac{\rho_f}{\tilde{\rho}_1}\right) \\ \xi_1 \dfrac{C_{13}^D}{C_{33}^D} & \dfrac{1}{C_{33}^D} & -\dfrac{\alpha_3}{C_{33}^D} \\ \xi_1\left(\alpha_1 - \alpha_3 \dfrac{C_{13}^D}{C_{33}^D} - \dfrac{\rho_f}{\tilde{\rho}_1}\right) & -\dfrac{\alpha_3}{C_{33}^D} & \dfrac{1}{M} + \dfrac{\alpha_3^2}{C_{33}^D} - \dfrac{1}{\tilde{\rho}_1}\xi_1^2 \end{bmatrix}. \tag{A9}$$

Using Eqs. (A7)–(A9), for a small fracture thickness h, formally identical simplified boundary conditions as in the isotropic case [Eqs. (46) and (52)] are obtained if the definitions of the characteristic fracture parameters are modified as follows:

$$\eta_T \equiv \frac{h}{G}, \tag{A10}$$

$$\eta_{N_D} \equiv \frac{h}{C_{33}^D}, \tag{A11}$$

$$\alpha \equiv \alpha_3, \tag{A12}$$

$$\hat{\kappa}(\omega) \equiv \frac{k_3(\omega)}{h}, \tag{A13}$$

$$\tilde{B} \equiv \frac{\alpha_3}{C_{33}^D} \Big/ \left(\frac{1}{M} + \frac{\alpha_3^2}{C_{33}^D}\right) = \alpha_3 \frac{M}{C_{33}^U}. \tag{A14}$$

Note that only the anisotropic material properties related to the 3 direction (fracture-normal direction) appear in these definitions, which indicates that the scattering of waves is not affected by the quantities related to the fracture-parallel directions (Specifically, permeability along the fracture). The fluid pressure dissipation factor Π is the same as the isotropic case if the direction of the slow P-wave propagation within the fracture is approximately in the fracture-normal direction. One important difference from the isotropic case, however, is that the normal and shear fracture compliances can take arbitrary values independent from each other.

APPENDIX B: THE DYNAMIC PERMEABILITY OF OPEN AND VERY PERMEABLE FRACTURES

For a highly permeable fracture or an open fracture in which fluid flow parallel to the fracture can be affected by the viscous friction along the fracture surfaces, the effective permeability of the layer representing a fracture in the frac-

ture parallel direction must be reduced. As shown in Appendix A, for a transversely isotropic fracture (a layer modeling the fracture is transversely isotropic), $\tilde{\mathbf{Q}}_{XY}$ contains only the fracture-normal permeability $k_3(\omega)$ and $\tilde{\mathbf{Q}}_{YX}$ contains only the fracture-parallel permeability $k_1(\omega)$. Therefore, for the isotropic fracture model discussed in the main body of this paper, the permeability needs to be allowed to be anisotropic.

In discussing the high-permeability case, we are concerned only with fracture-parallel permeability $k_1(\omega)$ in $\tilde{\mathbf{Q}}_{YX}$ because terms including $k_3(\omega)$ in the boundary conditions appear only as $h/k_3(\omega)$, which become negligibly small for small hs. If fracture-parallel fluid flow within the fracture is laminar and the flow on the fracture surfaces can be ignored because of the small permeability in the background, the maximum possible permeability for this fracture can be evaluated using Biot's results for the dynamic permeability of plane parallel flows (Biot, 1956b),

$$|k_1(\omega)| \leq |k_{\text{plane}(\omega)}| = \left| \frac{h^2}{4\theta^2}\left(1 - \frac{\tanh\,\theta}{\theta}\right)\right|, \tag{B1}$$

$$\theta \equiv \frac{h}{2}\sqrt{\frac{\rho_f \omega}{i\eta_f}}.$$

Therefore, the permeability for the flow in the fracture parallel direction is bounded by taking the limit of Eq. (B1) for $h \to \infty$

$$|k_1(\omega)| \leq \left| \frac{h^2}{4\theta^2}\right| = \frac{\eta_f}{\rho_f \omega}. \tag{B2}$$

Equation (B2) gives the maximum possible dynamic permeability of any fracture for a given fluid type. This limit can also be obtained directly from the momentum balance equation for an acoustic medium,

$$\nabla(-p_f) = -\omega^2 \rho_f \mathbf{U}. \tag{B3}$$

This equation can be rewritten as

$$\dot{\mathbf{U}} = \frac{i}{\omega\rho_f}\nabla(-p_f) \equiv \frac{k(\omega)}{\eta_f}\nabla(-p_f). \tag{B4}$$

Therefore, we identify the permeability as

$$k(\omega) = \frac{i\eta_f}{\omega\rho_f}, \tag{B5}$$

which is identical to Eq. (B2). The same expression can also be obtained by bringing the static permeability k_0 to infinity in the expression for in dynamic permeability given by Johnson *et al.* (1987),

$$k(\omega) = k_0 \Bigg/ \left(\sqrt{1 - i\frac{4}{n_J}\frac{\omega}{\omega_J}} - i\frac{\omega}{\omega_J}\right), \tag{B6}$$

where n_J is a finite parameter determined by the pore geometry (a value of 8 is recommended for common sandstones), and ω_J is the viscous-boundary characteristic frequency given by $\omega_J \equiv \eta_f/\rho_f F k_0 = \eta_f \phi/\rho_f \alpha_\infty k_0$, where F is the electrical formation factor and α_∞ is the high-frequency limit pore-space tortuosity, both of which approach unity for an open fracture.

Using Eq. (B2) and the definition in Eq. (4), we obtain

$$|\tilde{\rho}| \geq |\tilde{\rho}(k_0 \to \infty)| = \rho_f. \tag{B7}$$

This indicates that the magnitude of terms $h/\tilde{\rho}$ in $\tilde{\mathbf{Q}}_{YX}$ is bounded by a negligibly small value h/ρ_f for small h's. Therefore, permeability in the fracture parallel direction does not appear in the seismic boundary conditions for any static permeability values of the medium and does not affect the scattering of seismic waves. Conversely, permeability of a fracture in the fracture-parallel direction cannot be determined from measured seismic responses if the fracture thickness is much smaller than the wavelength of propagating seismic waves.

APPENDIX C: KENNETT'S REFLECTIVITY ALGORITHM APPLIED TO A SINGLE POROELASTIC LAYER

Pride *et al.* (2002) applied Kennett's reflectivity algorithm (Kennett, 1983) to piecewise-homogeneous layered poroelastic media. Exact expressions for the transmission and reflection coefficients of a single poroelastic layer representing a fracture can be obtained as a special case of the application.

Kennett method is based upon the following recursive relationships between the transmission and reflection coefficients for a group of n parallel interfaces and coefficients, for the remaining $n-1$ interfaces after the first interface in the series is removed:

$$\mathbf{T}^{(n)} = \mathbf{T}^{(n-1)}\mathbf{E}_n(\mathbf{I} - \mathbf{R}_n^- \mathbf{E}_n \mathbf{R}^{(n-1)}\mathbf{E}_n)^{-1}\mathbf{T}_n^+, \tag{C1}$$

$$\mathbf{R}^{(n)} = \mathbf{R}_n^+ + \mathbf{T}_n^- \mathbf{E}_n \mathbf{R}^{(n-1)}\mathbf{E}_n(\mathbf{I} - \mathbf{R}_n^- \mathbf{E}_n \mathbf{R}^{(n-1)}\mathbf{E}_n)^{-1}\mathbf{T}_n^+, \tag{C2}$$

where the transmission and reflection coefficient matrices for the removed interface are given as $\mathbf{T}_n$ and $\mathbf{R}_n$, respectively, with a sign in the superscript indicating the incident wave direction. $\mathbf{T}^{(n)}$, $\mathbf{T}^{(n-1)}$, $\mathbf{R}^{(n)}$, and $\mathbf{R}^{(n-1)}$ are for the n and $n-1$ interfaces, as indicated in the parentheses, and for incident waves propagating in the positive direction. $\mathbf{E}_n$ is the diagonal-phase advance matrix between the interfaces.

For the case of a single layer (two interfaces), no recursion is necessary to compute the transmission and reflection coefficients $\mathbf{T}$ and $\mathbf{R}$ for the whole system. By setting $\mathbf{T}^{(0)} = \mathbf{T}_0^+$, $\mathbf{R}^{(0)} = \mathbf{R}_0^+$, $\mathbf{T} = \mathbf{T}^{(1)}$, and $\mathbf{R} = \mathbf{R}^{(1)}$,

$$\mathbf{T} = \mathbf{T}_0^+ \mathbf{E}(\mathbf{I} - \mathbf{R}_1^- \mathbf{E}\mathbf{R}_0^+ \mathbf{E})^{-1}\mathbf{T}_1^+, \tag{C3}$$

$$\mathbf{R} = \mathbf{R}_1^+ + \mathbf{T}_1^- \mathbf{E}\mathbf{R}_0^+ \mathbf{E}(\mathbf{I} - \mathbf{R}_1^- \mathbf{E}\mathbf{R}_0^+ \mathbf{E})^{-1}\mathbf{T}_1^+. \tag{C4}$$

For in-plane wave propagation (fast and slow P waves and S waves with particle motions parallel to the plane of wave propagation), these matrices correspond to the transmission and reflection coefficient matrices in Eqs. (68) and (69). The phase advance matrix is $\mathbf{E} \equiv \text{diag}[e^{i\omega\xi_z^{Pf}h}\ e^{i\omega\xi_z^{Ps}h}\ e^{i\omega\xi_z^{S}h}]$. The transmission and reflection coefficient matrices for the individual interfaces are a function of material properties for both the background and the fracture layer, which results in very complex expressions for Eqs. (C3) and (C4) (albeit they are in closed form). These equations are evaluated numerically to obtain "correct solutions" in the example presented

in this paper. The interface scattering matrices can be computed, for example, using the equations presented by Pride *et al.* (2002).

Adams, J. T., and Dart, C. (**1998**). "The appearance of potential sealing faults on borehole images," in Faulting, Fault Sealing and Fluid Flow in Hydrocarbon Reservoirs, edited by G. Jones, Q. J. Fisher, and R. J. Knipe, Geol. Soc. Spec. Publ. **147**, 71–86.

Aydin, A. (**1978**). "Small faults formed as deformation bands in sandstone," Pure Appl. Geophys. **116**, 913–930.

Bakulin, A., and Molotkov, L. (**1997**). "Poroelastic medium with fractures as limiting case of stratified poroelastic medium with thin and soft Biot layers," *67th Annual International Meeting*, SEG, Expanded Abstracts, 1001–1004.

Bakulin, A., Grechka, V., and Tsvankin, I. (**2000**). "Estimation of fracture parameters from reflection seismic data. I. HTI model due to a single fracture set," Geophysics **65**, 1788–1802.

Berryman, J. G., and Wang, H. F. (**1995**). "The elastic coefficients of double-porosity models for fluid transport in jointed rock," J. Geophys. Res. **100**, 24611–24627.

Berryman, J. G., and Wang, H. F. (**2000**). "Elastic wave propagation and attenuation in a double-porosity dual-permeability medium," Int. J. Rock Mech. Min. Sci. **37**, 63–78.

Biot, M. A. (**1956a**). "Theory of elastic waves in a fluid-saturated porous solid. I. Low frequency range," J. Acoust. Soc. Am. **28**, 168–178.

Biot, M. A. (**1956b**). "Theory of elastic waves in a fluid-saturated porous solid. II. High frequency range," J. Acoust. Soc. Am. **28**, 179–191.

Brajanovski, M., Gurevich, B., and Schoenberg, M. (**2005**). "A model for P-wave attenuation and dispersion in a porous medium permeated by aligned fractures," Geophys. J. Int. **163**, 372–384.

Cheng, A. H.-D. (**1997**). "Material coefficients of anisotropic poroelasticity," Int. J. Rock Mech. Min. Sci. **34**(2), 199–205.

Coates, R. T., and Schoenberg, M. (**1995**). "Finite-difference modeling of faults and fractures," Geophysics **60**(5), 1514–1526.

Dutta, N. C., and Odé, H. (**1979a**). "Attenuation and dispersion of compressional waves in fluid-filled porous rocks with partial gas saturation (White Model)–I. Biot theory," Geophysics **44**, 1777–1788.

Dutta, N. C., and Odé, H. (**1979b**). "Attenuation and dispersion of compressional waves in fluid-filled porous rocks with partial gas saturation (White Model)–II. Results," Geophysics **44**, 1806–1812.

Gelinsky, S., and Shapiro, S. A. (**1997**). "Dynamic-equivalent medium approach for thinly layered saturated sediments," Geophys. J. Int. **128**, F1–F4.

Gurevich, B., Marschall, R., and Schapiro, S. A. (**1994**). "Effect of fluid flow on seismic reflections from a thin layer in a porous medium," J. Seism. Explor. **3**, 125–140.

Gurevich, B., Zyrianov, V. B., and Lopatnikov, S. L. (**1997**). "Seismic attenuation in finely layered porous rocks: Effects of fluid flow and scattering," Geophysics **62**(1), 319–324.

Gurevich, B., and Schoenberg, M. A. (**1999**). "Interface conditions for Biot's equations of poroelasticity," J. Acoust. Soc. Am. **105**(5), 2585–2589.

Haskell, N. A. (**1953**). "The dispersion of surface waves in multilayered media," Bull. Seismol. Soc. Am. **43**, 17–34.

Hsu, C.-J., and Schoenberg, M. A. (**1993**). "Acoustic waves through a simulated fractured medium," Geophysics **58**(7), 964–977.

Johnson, D. L. (**2001**). "Theory of frequency dependent acoustics in patchy-saturated porous media," J. Acoust. Soc. Am. **110**(2), 682–694.

Johnson, D. L., Koplik, J., and Dashen, R. (**1987**). "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402.

Kennett, B. L. N. (**1983**). *Seismic Wave Propagation in Stratified Media* (Cambridge University Press, Cambridge).

Nakagawa, S., Nihei, K. T., and Myer, L. R. (**2002**). "Elastic wave propagation along a set of parallel fractures," Geophys. Res. Lett. **29**, 31-1–31-4.

Nihei, K. T., Weidong, Y., Myer, L. R., Cook, N. G. W., and Schoenberg, M. A. (**1999**). "Fracture channel waves," J. Geophys. Res. **104**(B3), 4769–4781.

Norris, A. N. (**1993**). "Low-frequency dispersion and attenuation in partially saturated rocks," J. Acoust. Soc. Am. **94**, 359–370.

Pride, S. R., Tromeur, E., and Berryman, J. G. (**2002**). "Biot slow-wave effects in stratified rock," Geophysics **67**(1), 271–281.

Pride, S. R. (**2003**). "Relationships between seismic and hydrological properties," in *Hydrogeophysics*, edited by Y. Rubin and S. Hubbard (Kluwer Academic, New York), pp. 1–31.

Pride, S. R., and Berryman, J. G. (**2003a**). "Linear dynamics of double-porosity and dual-permeability materials, I. Governing equations and acoustic attenuation," Phys. Rev. E **68**, 036603.

Pride, S. R., and Berryman, J. G. (**2003b**). "Linear dynamics of double-porosity and dual-permeability materials. II. Fluid transport equation," Phys. Rev. E **68**, 036604.

Pyrak-Nolte, L. J., and Cook, N. G. W. (**1987**). "Elastic interface waves along a fracture," Geophys. Res. Lett. **14**(11), 1107–1110.

Pyrak-Nolte, L. J., and Morris, J. P. (**2000**). "Single fractures under normal stress: The relation between fracture specific stiffness and fluid flow," Int. J. Rock Mech. Min. Sci. **37**, 245–262.

Pyrak-Nolte, L., Cook, N. G. W., and Myer, L. R. (**1990**). "Transmission of seismic waves across single natural fractures," J. Geophys. Res. **95**, 8516–8538.

Rokhlin, S. I., and Wang, Y. J. (**1991**). "Analysis of boundary conditions for elastic wave interaction with an interface between two solids," J. Acoust. Soc. Am. **89**(2), 503–515.

Schoenberg, M. A. (**1980**). "Elastic wave behavior across linear slip interfaces," J. Acoust. Soc. Am. **68**, 1516–1521.

Schoenberg, M. A., and Protazio, J. (**1992**). "Zoeppritz rationalized and generalized to anisotropy," J. Seism. Explor. **1**(2), 125–144.

Schoenberg, M. A., and Sayers, C. M. (**1995**). "Seismic anisotropy of fractured rock," Geophysics **60**, 204–211.

Shapiro, S. A., and Müller, T. M. (**1999**). "Seismic signatures of permeability in heterogeneous porous media," Geophysics **64**(1), 99–103.

Sibson, R. H. (**1977**). "Fault rock and fault mechanisms," J. Geol. Soc. (London) **133**, 191–213.

White, J. E., Mikhahaylova, N. G., and Lyakhovistsky, F. M. (**1975**). "Low-frequency seismic waves in fluid-saturated layered rocks," Izv., Acad. Sci., USSR, Phys. Solid Earth **11**, 654–659.

Experimental investigation of an inversion technique for the determination of broadband duct mode amplitudes by the use of near-field sensor arrays

Fabrice O. Castres and Phillip F. Joseph[a)]
ISVR, Southampton University, University Road, Highfield, Southampton SO17 1BJ, United Kingdom

(Received 5 February 2007; revised 9 May 2007; accepted 10 May 2007)

This paper is an experimental investigation of an inverse technique for deducing the amplitudes of the modes radiated from a turbofan engine, including schemes for stablizing the solution. The detection of broadband modes generated by a laboratory-scaled fan inlet is performed using a near-field array of microphones arranged in a geodesic geometry. This array geometry is shown to allow a robust and accurate modal inversion. The sound power radiated from the fan inlet and the coherence function between different modal amplitudes are also presented. The knowledge of such modal content is useful in helping to characterize the source mechanisms of fan broadband noise generation, for determining the most appropriate mode distribution model for duct liner predictions, and for making sound power measurements of the radiated sound field. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2747166]

PACS number(s): 43.50.Gf, 43.20.Ye [NX] Pages: 848–859

I. INTRODUCTION

The noise radiated by a turbofan engine may be categorized as either tonal or broadband in nature. The tones, at subsonic rotor tip speeds, occur at multiples of the blade passage frequency (BPF) and are generated by the interaction between the rotor wakes and stator vanes or struts.[1] At supersonic rotor tip speeds, the tones occur at harmonics of the shaft rotation frequency that only radiate upstream and are collectively called buzz saw noise. The broadband noise is characterized by a smooth continuous frequency spectrum over a wide frequency band and its source mechanism is poorly understood. In recent times, improved liners in the engine inlets and bypass sections, modern fan blade configurations, and cut-off design of the rotor and stator (to prevent propagating interaction modes) have become standard noise control solutions. These have allowed significant reductions in the levels of the tones, while the broadband noise is relatively unaffected. The broadband sound field has therefore become more important and its understanding and reduction are now regarded as one of the most urgent challenges facing aeroengine noise control engineers. The sound field radiated by a fan inlet can be expressed as a weighted sum of its modes and therefore the knowledge of these modes is essential in determining the characteristics of the broadband sound field. While several experimental techniques aimed at deducing the mode amplitudes of the tonal noise radiated by a turbofan inlet have been proposed and implemented,[2–4] very few measurement techniques for the detection of mode amplitudes of the broadband sound field can be found in the literature. One such technique has been presented recently by Enghardt *et al.*,[5] which uses an in-duct microphone array to deduce the mode amplitudes and hence the in-duct transmitted sound power for tones and broadband noise. The technique requires cross-spectral measurements to be made between a number of microphones and a single reference microphone at the duct wall. The cross spectra of pressure inside the duct can be modeled in terms of the constituent duct modes via a transfer matrix which, in a narrow frequency band, can be inverted to deduce the mean square mode amplitudes. The technique can only be implemented if the assumption, that individual modes are mutually uncorrelated, is made. Furthermore, the matrix established for modal inversion may be poorly conditioned and extraneous noise contaminating the measurements will be greatly amplified in the modal solution. Enghardt *et al.*[5] regularized the solution and hence reduced this sensitivity to noise by discarding small singular values thereby introducing small errors in the solution. However, the system of equations considered is not positive definite, such that negative mean square mode amplitudes can be obtained. The implementation of an iterative least-squares solver is required to ensure positive mean square mode amplitudes. The technique was applied to a single-stage compressor comprising 24 fan blades and 17 stator vanes. An azimuthally traversable duct section is attached downstream of the rotor. The mode detection is performed using 8 microphones installed in the moving duct section and traversed over 36 azimuthal positions. A wall-flush mounted microphone was used as a reference sensor. The sound power level (SWL) deduced from the modal analysis was compared with that obtained from the ISO 5136 standard. The estimate obtained from the mode detection method was found to overpredict the SWL by 2 dB. However, this discrepancy could be equally due to inaccuracies in the ISO standard as from the mode detection technique.

In this paper, we experimentally investigate another technique described in detail by Castres and Joseph[6] for deducing the mean square mode amplitudes of a ducted broadband sound field based on pressure measurements made in

[a)]Author to whom correspondence should be addressed. Electronic mail: pfj@isvr.soton.ac.uk

the radiated near field. The motivation of such a technique is to make use of the turbulence control screen (TCS) to determine the broadband mode amplitudes of turbofan inlets. The TCS is an acoustically transparent surface used to remove the ground vortices and incoming turbulence at the inlet during ground testing on a test bed. A microphone array mounted on the TCS has several advantages over measurement techniques based on other microphone distribution, in that, it is generally nonintrusive and easy to implement in full-scale engine tests. Furthermore, the TCS offers a platform to mount microphones from which all modes can be deduced simultaneously. We first briefly review the principles of the mode inversion technique described in Castres and Joseph[6] and then give the guidelines used in the design of the near-field sensor array geometry used in the experimental investigation. The paper then describes the results from the laboratory-scale experiments aimed at inverting the mean square mode amplitudes of the sound field radiated from a low speed industrial fan rig based on the simultaneous pressure measurements at 91 microphones in the radiated near field. The modal inversion results are compared with that determined from 24 in-duct microphones. Measurements are also presented of the coherence function between the different mode amplitudes aimed at deducing for the first time, the degree of statistical interdependence between different modes in the broadband sound field radiated from a ducted fan.

II. THE MODAL INVERSE PROBLEM

As discussed in Castres and Joseph,[6] at a single frequency, the sound field radiated at a point $\mathbf{r}$ in space from a cylindrical duct may be written as the sum of modal components

$$p(\mathbf{r}) = \sum_{m=-\infty}^{+\infty} \sum_{n=1}^{+\infty} a_{mn} D_{mn}(\mathbf{r}), \tag{1}$$

where $D_{mn}(\mathbf{r})$ and a_{mn} are the radiation directivity factor and the complex pressure amplitude of a mode of spinning mode number m and radial mode number n, respectively. If $k = \omega/c$ denotes the free space wave number and c the speed of sound, an important parameter for this study is the modal cut-on ratio α_{mn} defined by

$$\alpha_{mn} = \frac{\kappa_{mn}}{k}, \tag{2}$$

where κ_{mn} is the transverse modal wave number. Note that $\alpha_{mn} \mapsto 0$ for modes excited well above their cut-on frequencies and $\alpha_{mn} \mapsto 1$ as the cut-on frequency is approached from above. From Castres and Joseph,[6] Eq. (1) is restricted to cut-on modes only and can now be expressed in matrix form

$$\mathbf{p} = \mathbf{D}\mathbf{a}, \tag{3}$$

where $\mathbf{p}$ is a vector of radiated acoustic complex pressures at K sensors, $\mathbf{a}$ is a vector of L complex modal amplitudes, and $\mathbf{D}$ is the modal directivity matrix and contains the radiation properties of L cut-on modes at K sensor positions on the near-field array.

A vector $\hat{\mathbf{p}}$ of K pressure measurements may be expressed as the sum of the modeled pressures given by Eq. (3), and a vector $\mathbf{n}$ whose elements represent the departure of the measurements from the model and may also include, for example, the effect of contaminating noise at the sensors. Thus,

$$\hat{\mathbf{p}} = \mathbf{D}\mathbf{a} + \mathbf{n}. \tag{4}$$

We seek the vector of modal amplitudes $\hat{\mathbf{a}}$ that ensures the least-squares fit of the modeled acoustic pressures to the measured pressure data, i.e., that which minimizes $\|\hat{\mathbf{p}} - \mathbf{D}\mathbf{a}\|_2^2$, the 2-norm of the squared errors. The least-squares solution $\hat{\mathbf{a}}$ is given by[7]

$$\hat{\mathbf{a}} = \mathbf{D}^{+}\hat{\mathbf{p}}, \tag{5}$$

where $\mathbf{D}^{+}$ is the pseudoinverse of the modal directivity matrix $\mathbf{D}$.

For broadband radiated noise, the least-squares estimate for the mode amplitudes is fully characterized by the cross-spectral density matrix of mode amplitudes given by

$$\mathbf{S}_{\hat{\mathbf{a}}\hat{\mathbf{a}}} = \lim_{T\to\infty} E\left[\frac{1}{T}\hat{\mathbf{a}}\hat{\mathbf{a}}^{\mathbf{H}}\right] = \mathbf{D}^{+}\mathbf{S}_{\hat{\mathbf{p}}\hat{\mathbf{p}}}\mathbf{D}^{+\mathbf{H}}, \tag{6}$$

where $\mathbf{S}_{\hat{\mathbf{p}}\hat{\mathbf{p}}}$ is the $K \times K$ cross-spectral matrix of pressure measurements defined by

$$\mathbf{S}_{\hat{\mathbf{p}}\hat{\mathbf{p}}} = \lim_{T\to\infty} E\left[\frac{1}{T}\hat{\mathbf{p}}\hat{\mathbf{p}}^{\mathbf{H}}\right]. \tag{7}$$

The diagonal of this matrix represents the mean-square pressure radiated at the sensors, while off-diagonal terms, appropriately normalized, provide an indication of the level of coherence between any two microphones. The $L \times L$ matrix of mode amplitude cross spectra of Eq. (6) is given by

$$\mathbf{S}_{\hat{\mathbf{a}}\hat{\mathbf{a}}} = \begin{bmatrix} S_{\hat{a}_1\hat{a}_1}(\omega) & & S_{\hat{a}_L\hat{a}_1}(\omega) \\ & \cdot & \\ & S_{\hat{a}_j\hat{a}_j}(\omega) & \\ & \cdot & \\ S_{\hat{a}_1\hat{a}_L}(\omega) & & S_{\hat{a}_L\hat{a}_L}(\omega) \end{bmatrix}. \tag{8}$$

The diagonal of this matrix represents the mean square mode amplitudes in a unit frequency band, also denoted by $\overline{a_{mn}^2}$. It was shown in Castres and Joseph[6] that the condition number of $\mathbf{D}$ is a useful parameter for evaluating the sensitivity of the modal solution to the presence of noise and modeling errors. If the noise and the pressure signals are uncorrelated, $\mathbf{S}_{\hat{\mathbf{p}}\hat{\mathbf{p}}}$ can be related to the modeled pressure cross-spectral matrix $\mathbf{S}_{\mathbf{pp}}$ and the matrix of noise cross-spectra $\mathbf{S}_{\mathbf{nn}}$ as

$$\mathbf{S}_{\hat{\mathbf{p}}\hat{\mathbf{p}}} = \mathbf{S}_{\mathbf{pp}} + \mathbf{S}_{\mathbf{nn}}. \tag{9}$$

If $\mathbf{e}$ denotes the error found in the modal solution $\hat{\mathbf{a}}$ (i.e., $\hat{\mathbf{a}} = \mathbf{a} + \mathbf{e}$) it follows from Eq. (5) that

$$\mathbf{a} + \mathbf{e} = \mathbf{D}^{+}\mathbf{p} + \mathbf{D}^{+}\mathbf{n}. \tag{10}$$

Since the first term of the right-hand side of Eq. (10) corresponds to the noise-free modal solution (i.e., $\mathbf{a}=\mathbf{D}^{+}\mathbf{p}$), the error in the modal solution can be written in terms of the measurement noise as follows:

$$\mathbf{e}=\mathbf{D}^{+}\mathbf{n}. \tag{11}$$

The cross-spectral matrix of errors in the least-squares solution $\hat{\mathbf{S}}_{\mathbf{ee}}$ resulting from measurement noise at the sensors, $\mathbf{S}_{\mathbf{nn}}$ can therefore be written as

$$\mathbf{S}_{\mathbf{ee}}=\lim_{T\to\infty}E\left[\frac{1}{T}\hat{\mathbf{e}}\hat{\mathbf{e}}^{\mathbf{H}}\right]=\mathbf{D}^{+}\mathbf{S}_{\mathbf{nn}}(\mathbf{D}^{+})^{\mathbf{H}}. \tag{12}$$

If $\kappa(\mathbf{D})$ denotes the condition number of the directivity matrix $\mathbf{D}$, it may be shown[8] that

$$\frac{\|\mathbf{S}_{\mathbf{ee}}\|}{\|\mathbf{S}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}\|}\leqslant\kappa^{2}(\mathbf{D})\frac{\|\mathbf{S}_{\mathbf{nn}}\|}{\|\mathbf{S}_{\hat{\mathbf{p}}\hat{\mathbf{p}}}\|}. \tag{13}$$

Equation (13) is, for broadband noise, the equivalent expression to Eq. (17) in Castres and Joseph[6] for tonal noise. In other words, the condition number of the directivity matrix bounds the error in the solution from the error found in the measurements both for tonal and broadband noise.

III. DESIGN OF A ROBUST SENSOR ARRAY

As demonstrated in Castres and Joseph,[6] at frequencies not too close to the modal cut-on frequencies, the condition number $\kappa(\mathbf{D})$ will depend critically upon the sensor locations in the radiated field. A comparatively small value of $\kappa(\mathbf{D})$ is indicative of good coupling between the modal information radiated from the duct inlet and the sensor positionings. Before presenting experimental modal inversion results, guidelines about the positioning and the number of microphones required for robust inversion are now discussed.

Five different array geometries comprising 126 sensors positioned over a hemispherical surface, whose axis is aligned with the duct axis, are presented in Fig. 1. Alongside each array, their condition number at $ka=20$ for the inversion of 109 modes, where a is the duct radius is presented. It is clear that the condition number can be reduced by many orders of magnitude by a judicious choice of sensor array.

Figure 1 shows that a uniform distribution of microphones, both in the azimuthal and polar directions, is an important requisite for low $\kappa(\mathbf{D})$ and hence robust inversion. The directivity matrices formed from "star" and "ring" arrays are poorly conditioned. These sensor arrays, shown in Figs. 1(a) and 1(b), possess a sparse distribution of microphones in polar and azimuthal directions, respectively. The sensor arrays presented in Figs. 1(c) and 1(d) have similar geometries to the geometries in Figs. 1(a) and 1(b) but offer a fine sampling of microphones in both the polar and azimuthal directions simultaneously.

It is also important to understand how $\kappa(\mathbf{D})$ varies as the number of microphones is increased in a particular array geometry. Figure 2 shows the variation of $\kappa(\mathbf{D})$ plotted against the ratio of the number of sensors to the number of cut-on modes K/L at $ka=20$ in a starfish array.

It can be seen that the condition number decreases significantly as the number of microphones is increased slightly above the number of cut-on modes. The condition number $\kappa(\mathbf{D})$ ceases to improve for ratios K/L above about 1.3 suggesting that the array should have at least 30% more sensors than modes at the highest frequency of interest. Very similar behavior is observed for the other arrays shown in Fig. 1.

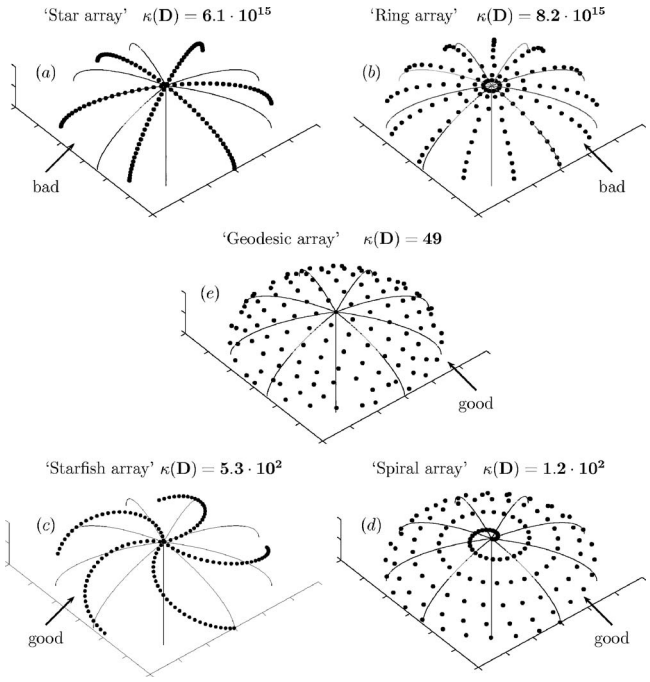


FIG. 1. Examples of array geometry and their resulting condition number $\kappa(\mathbf{D})$ at $ka=20$: (a) star array, (b) ring array, (c) starfish array, (d) spiral array, and (e) geodesic array.

The most robust sensor array, i.e., that with the lowest $\kappa(\mathbf{D})$ value, is the array with microphones occupying equal surface area over the TCS, as shown in Fig. 1(e). It is referred to here as a geodesic array following the definition of a geodesic sphere given by Fuller.[9,10] This array is based on the geometry of an icosahedron inscribed within a sphere. The icosahedron has 20 faces, each of which is an isosceles triangle. Each of these triangles may be divided further into a

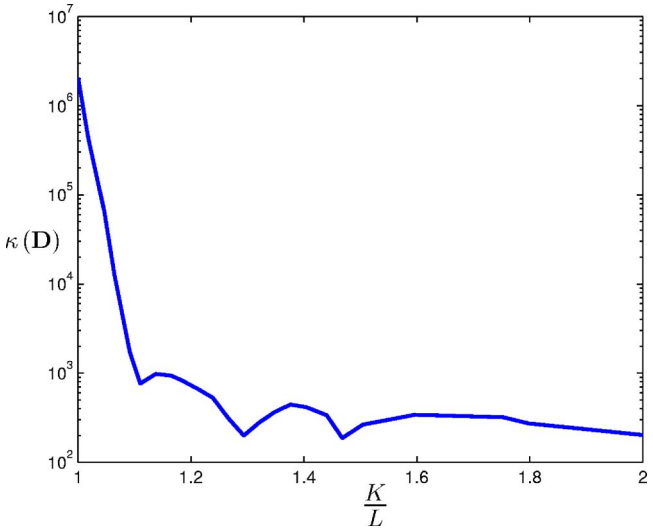


FIG. 2. (Color online) The condition number $\kappa(\mathbf{D})$ vs the ratio of number of sensors to number of cut-on modes K/L at $ka=20$ for a starfish sensor array.

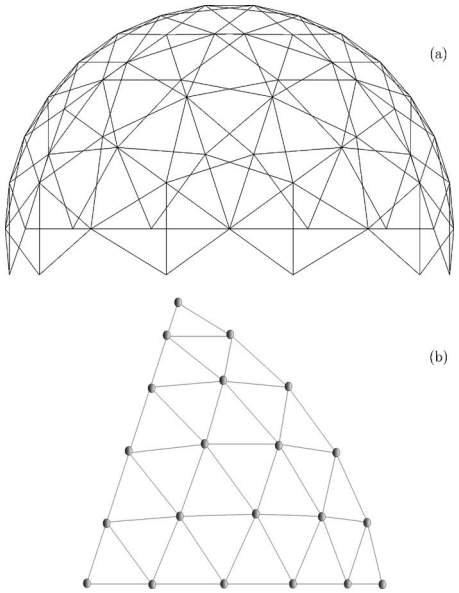


FIG. 3. (a) A geodesic sensor array. (b) One face of an icosahedron.

given number of smaller isosceles triangles. Figure 3(a) illustrates an example of such a geometry. Figure 3(b) shows a single face of the icosahedron. The microphones are placed on the vertices of each subface of the icosahedron.

This array geometry, by definition, has the property of a perfectly uniform spreading of microphones over the hemispherical surface. However, the number of microphones satisfying this arrangement is limited to certain discrete values by the number of subdivisions. It is interesting to note that the TCS usually applied on real engines during ground testing has the geometry of an irregular geodesic sphere, see for example Fig. 1 in Castres and Joseph.[6]

Figure 4 shows the behavior of $\kappa(\mathbf{D})$ versus ka for the three arrays with lowest conditioning, i.e., the geodesic, spiral, and starfish arrays. Here, the modal radiation matrix $\mathbf{D}$ was computed for the duct inlet investigated experimentally in Sec. IV from a finite element (FE) and infinite element (IE) analysis and the ratio of the number of sensors to the number of cut-on modes K/L decreases as ka increases.

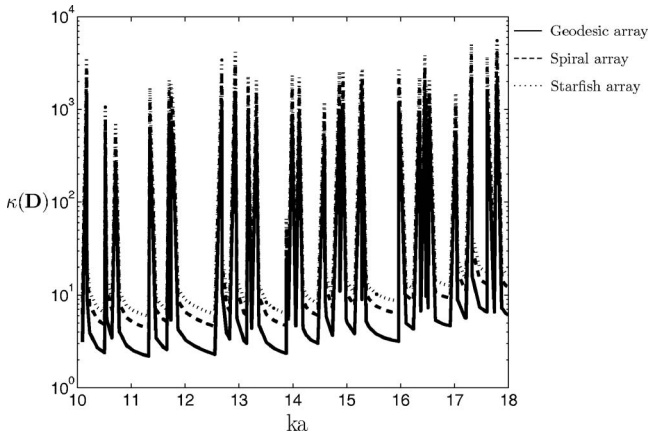
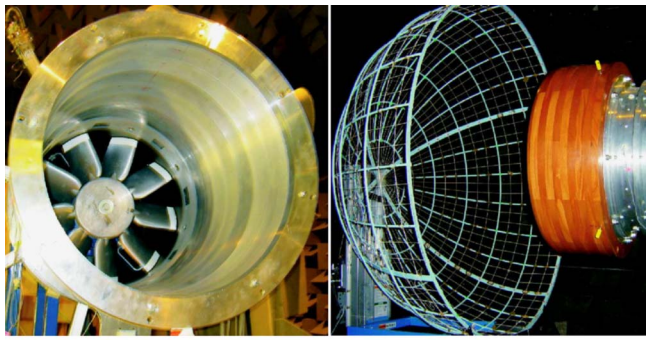


FIG. 4. Variation of the condition number $\kappa(\mathbf{D})$ with normalized frequency ka accounting for the inversion of all cut-on modes by the "geodesic," "spiral," and "starfish" arrays.



(a) View of the fan (b) Inlet with bellmouth and hemispherical rig

FIG. 5. (Color online) The experimental rig.

The κ spectrum in Fig. 4 shows the presence of a number of peaks coinciding with the modal cut-on frequencies, as discussed in Castres and Joseph.[6] At frequencies in between these peaks, the geodesic array is seen to afford a more robust inversion with a κ value of approximately two to three times smaller than either the spiral or starfish arrays.

IV. EXPERIMENTAL METHOD

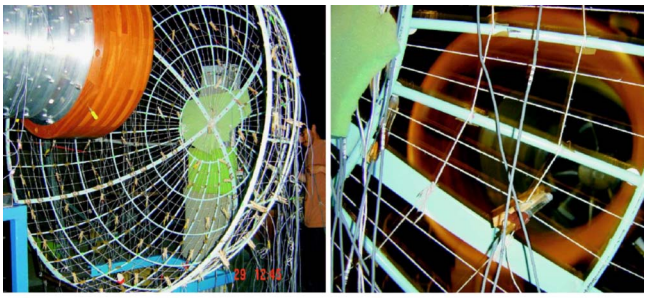
A. Experimental setup

The above-described inversion theory was experimentally investigated. An array comprising 91 microphones was used to measure the sound field radiated from a low speed fan in a circular duct of 0.315 m radius. The fan rig is located in a semianechoic chamber that enables measurements of the acoustic radiation from the duct inlet to be made under nearly free-field conditions. The fan comprises 9 blades situated at a distance of 0.84 m from the duct exit as shown in Fig. 5(a).

The maximum rotational frequency of the fan is 50 Hz, which produces an axial flow speed of $M_z \approx 0.1$. A single ring of 24 wall flush mounted microphones is used to perform a spinning modal decomposition of the broadband sound field by way of validating the mode amplitudes inverted from the TCS pressure measurements. The bellmouth shown in Fig. 5(b) is fixed to the end of the duct to reproduce the acoustic behavior of the lip usually present in the turbofan inlets of real aircraft engines.

A hemispherical wire-frame made from a series of steel bars bent into an arc and positioned at every 15°, both in the polar and azimuthal angles, is used for the pressure measurements of the radiated field. In order to allow more accurate positioning of the sensors, strings are stretched over the rig at every 5° in both angles, as shown in Fig. 6. The hemispherical array is located over a hemisphere of radius $r=1$ m, which corresponds to a ratio $r/a=3.2$. The TCS used on real engines corresponds to a ratio $r/a=6$. The model TCS is therefore located slightly closer in the near field of the duct than that of a full-scale TCS.

Acoustic pressure measurements were made simultaneously using 91 1/4 in. Brüel & Kjaer Falcon microphones arranged over the model TCS according to the geodesic ge-



(a) Microphones mounted on the TCS (b) Close look of a microphone clipped on the TCS

FIG. 6. (Color online) Experimental setup of microphones on the laboratory-scale TCS.

ometry discussed earlier, with their diaphragms facing toward the center of the duct inlet. The mounting of the microphones on the TCS is shown in Fig. 6.

Signals captured by the microphones are then fed into four custom-built preamplifiers, each comprising 33 input channels. Amplifier outputs are then connected to two SONY SIR 1000 digital recording systems, each with 64 channels. These are synchronized to record signals simultaneously on DAT magnetic tapes. Low pass filters are used to prevent aliasing of the high frequency signals. The TCS sensors are calibrated relative to a reference microphone. This is done by measurements of the transfer function between the reference microphone colocated with each microphone on the array using a loudspeaker facing the sensors and driven by a white noise signal. The pressure at each array microphone was then corrected for magnitude and phase relative to the reference microphone via this transfer function. The calibration procedure allows the gain and phase of the microphones to be determined to better than 10% and 2°, respectively. The robustness of the array ensures that the accuracy of the technique is not affected by these small errors. It was found to be important to calibrate the in-duct microphones *in situ* since they are mounted within plastic holders prior to insertion into the holes in the duct wall and their sensitivities were found to change during insertion. In order to do so, a reference microphone is flush-mounted onto a rigid plate and the loudspeaker is directly placed onto the reference microphone with an insulating layer to prevent any cavity resonances. The calibration was performed relative to a reference sensor by the use of a transfer function between the loudspeaker signal and the calibrated reference microphone signal. The loudspeaker was then placed onto each microphone placed inside the duct and the transfer function between the loudspeaker signal and any in-duct microphone signal was measured.

B. Experimental procedure

The modal directivity matrix **D** of the experimental fan inlet was computed using the FE/IE code ACTRAN. The FE/IE mesh to the sound field radiated from the experimental fan inlet is shown in Fig. 7.

Cross power spectral densities between each microphone with every other microphone were computed using the Welch method[11] from the acquisition of 60 s of data. The

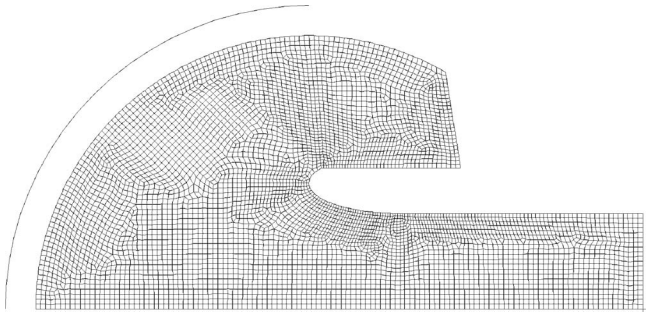


FIG. 7. FE/IE mesh of the experimental inlet.

signals are discretized at a sampling rate f_s of 12 kHz. The data sequence is divided into 1500 segments of 43 ms, on which a $N_w=512$ point fast Fourier transform is performed to give a frequency resolution of $\Delta f=f_s/N_w=23.44$ Hz. In nondimensional duct units, this frequency resolution corresponds to $\Delta ka=0.1364$. This allows the estimation of the power spectra at 257 different frequency bins over the frequency range 0–6 kHz.

V. EXPERIMENTAL RESULTS

A. Spectrum of sound power

By making the usual far-field approximation and using the property of the geodesic array that each microphone occupies an equal surface area of $\Delta A=2\pi r^2/K$, the spectral density of sound power radiated from the fan inlet is approximately given by

$$W(\omega)=\frac{\Delta A}{\rho c}\sum_{i=1}^{K} S_{\hat{p}_i\hat{p}_i}(\theta_i,\varphi_i,\omega). \tag{14}$$

The sound power radiated from the inlet at four fan speeds of 20, 30, 40, and 50 Hz, measured by the geodesic sensor array, is plotted in Fig. 8 against the dimensionless frequency ka.

Peaks in the spectrum are found to occur at multiples of the BPF, which corresponds to 180, 270, 360, and 450 Hz

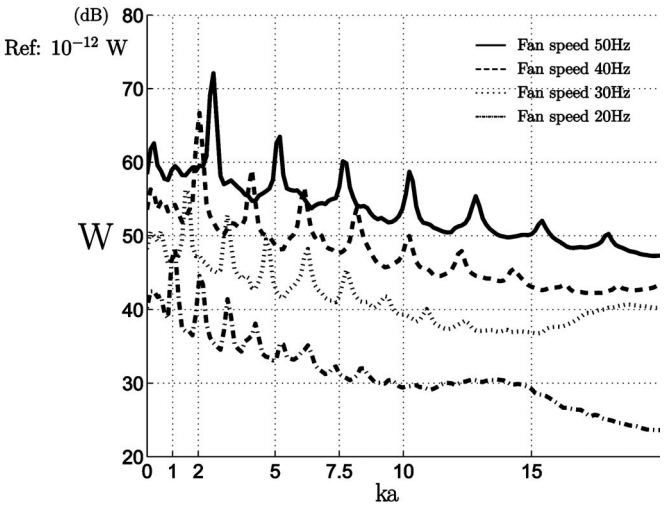


FIG. 8. The power radiated W from the fan inlet and measured by the geodesic sensor array for fan speeds 20, 30, 40, and 50 Hz.

(i.e., $ka=1.05$, 1.57, 2.1, and 2.62) for the four fan speeds. However, apart from the first tone in each spectrum with a level of 10 dB above the broadband level, the peaks of subsequent tones are only about 5 dB above the broadband level. These tones also have a significant frequency bandwidth of around 60 Hz. These peaks therefore correspond to pseudotones that are most likely due to the interaction between upstream flow distortion and the fan blades. Here, the frequencies of these peaks are referred to as tones to distinguish them from the frequencies at which purely broadband noise occurs. Finally, signal to noise ratios (SNR) of 25 and 30 dB are found for the fan speeds of 20 and 30 Hz, while 40 and 45 dB SNR is achieved for the 40 and 50 Hz fan speeds. The modal inversion is therefore performed at the highest fan speed of 50 Hz to ensure maximum SNR levels. In this paper, the modal inversion procedure will be applied at the following three frequencies:

1. Broadband part of the spectrum: $ka=13.78$.
2. Pseudotone frequency: $ka=12.82$.
3. Modal cut-on frequencies of the (±7,1) and (±8,1) modes: $ka=12.93$ and $ka=14.12$.

B. Comparison of measured and reconstructed directivities

As an indication of the quality of the inversion, we now compare the measured mean squared pressures with the diagonal elements of $\bar{\mathbf{S}}_{\hat{\mathbf{p}}\hat{\mathbf{p}}}$ deduced from the reconstructed modal matrix $\mathbf{S}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}$ as follows:

$$\bar{\mathbf{S}}_{\hat{\mathbf{p}}\hat{\mathbf{p}}} = \mathbf{D}\mathbf{S}_{\hat{\mathbf{a}}\hat{\mathbf{a}}}\mathbf{D}^{\mathbf{H}}. \tag{15}$$

In nondimensionalized form, the mean square pressure over the hemispherical array may be written as

$$Q(\theta_i,\varphi_i,\omega) = \frac{S_{\hat{p}\hat{p}}(\theta_i\varphi_i,\omega)A}{\rho c W(\omega)} = \frac{S_{\hat{p}\hat{p}}(\theta_i\varphi_i,\omega)}{\frac{1}{A}\int_A S_{\hat{p}\hat{p}}(\theta_i\varphi_i,\omega)dA} \tag{16}$$

such that $(1/A)\int_A Q dA = 1$. Here, θ and φ represent the azimuthal and the polar coordinates, respectively.

Figure 9 shows maps of Q and $\bar{Q}$, the measured and reconstructed dimensionless mean square pressure variation, respectively, obtained from the geodesic array measurements at $ka=12.82$, 13.78, and 14.12, as viewed looking along the axis of the array.

Comparison between $\hat{Q}$ and $\bar{Q}$ allows an assessment of the residual error, $\|\hat{\mathbf{p}}-\mathbf{p}\|$ resulting from modeling errors and errors due to noise on the sensors. Good agreement between Q and $\bar{Q}$ can generally be observed, which suggests that the least-square estimate of mode amplitudes defined in Eq. (6) provides a close match to the measured pressures when the condition number is low [i.e., $\kappa(\mathbf{D})=3.6$ and $\kappa(\mathbf{D})=3.88$ for $ka=12.82$ and $ka=13.78$, respectively]. Agreement of 3 dB or less is achieved across the surface of the microphone array at most frequencies. An exception occurs at the cut-on frequency, $ka=14.12$, where differences of 6–8 dB are observed in the sideline directions (i.e., $\varphi \in [80°, 90°]$). It was shown that the inefficient transformed pressures strongly contaminated by measurement noise radiate most strongly to the sideline direction.[6] The mean square pressure variation may therefore be reconstructed to within 3 dB of the actual pressure, suggesting that the modal amplitudes should be inverted to the same level of accuracy using arrays with good conditioning.

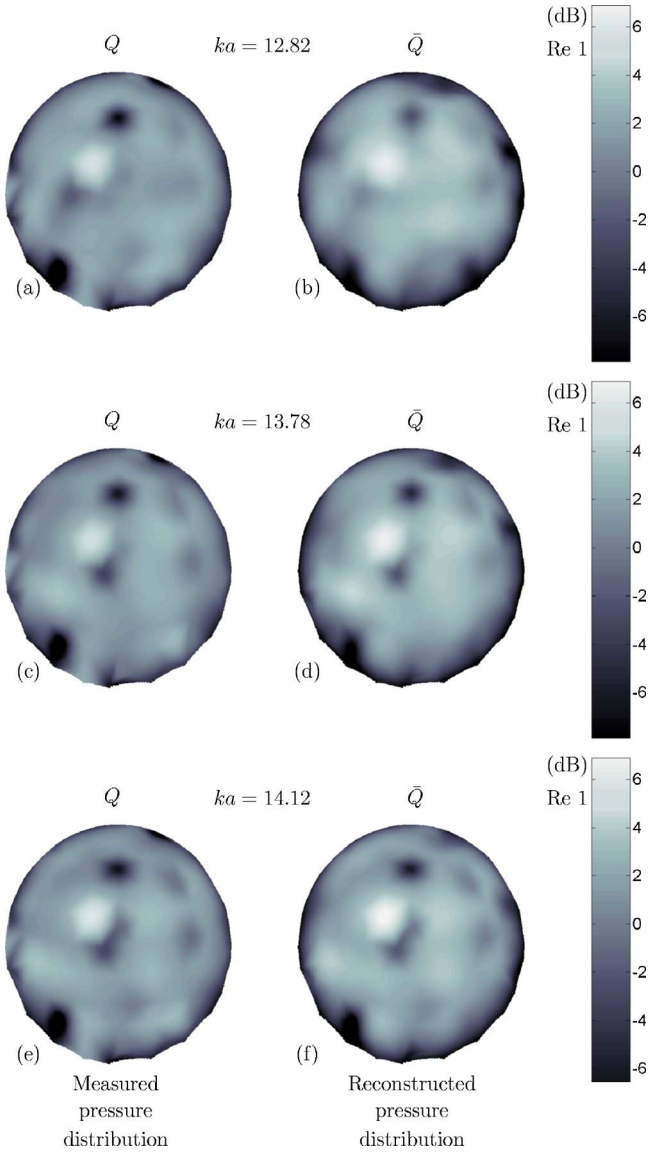


FIG. 9. (Color online) Normalized mean-square pressure representation over the TCS hemispherical surface looking from above at (a),(b) $ka=12.82$, (c),(d) $ka=13.78$, and (e),(f) $ka=14.12$.

C. The mean square mode amplitudes inverted from TCS measurements

Figure 10 shows the inverted mean square mode amplitudes obtained from the diagonal elements of Eq. (6) at $ka=12.82$, 12.93, 13.78, and 14.12 plotted against their cut-on ratio α_{mn}. For clarity of presentation, modes with negative spinning mode numbers are identified by a negative cut-on ratio. Note that modal amplitudes for modes not too close to cutoff have an α_{mn} distribution that appears to follow reasonably well an "equal energy per mode" model, which for zero flow is of the form[12]

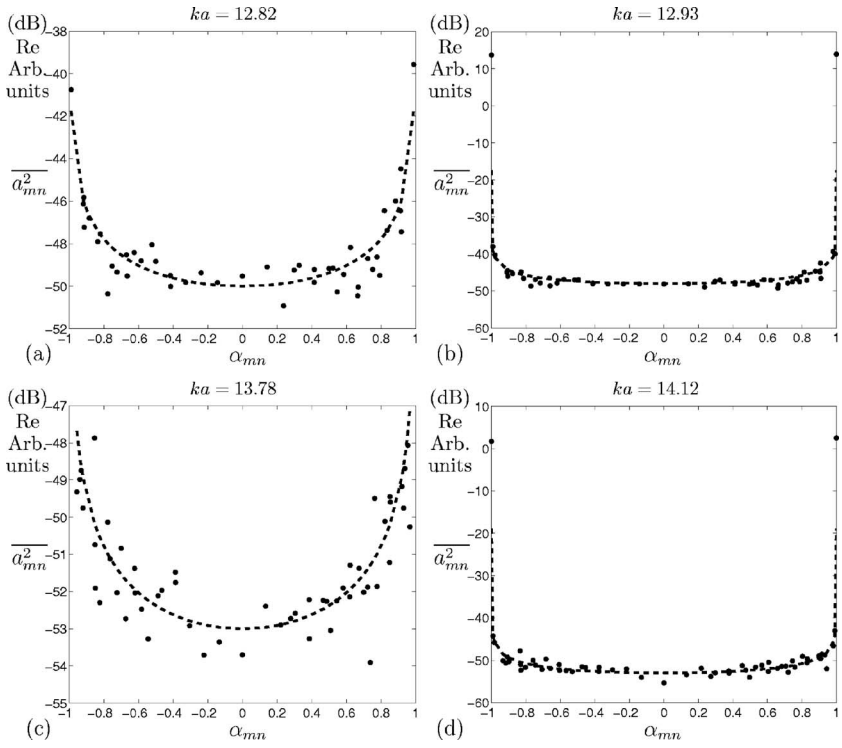


FIG. 10. Mean square mode amplitudes $\overline{a_{mn}^2}$ against the modal cut-on ratio α_{mn} inverted from geodesic array and Eq. (17) is plotted by a dashed line (a) $ka=12.82$, (b) $ka=12.93$, (c) $ka=13.78$, and (d) $ka=14.12$.

$$\overline{a_{mn}^2} = \rho c \frac{\varpi}{S\sqrt{1-\alpha_{mn}^2}}, \tag{17}$$

where S is the duct cross-section area. Equation (17) is shown in Fig. 10 as a dashed curve (where ϖ is the time-averaged acoustic sound power carried by each cut-on mode at the frequency of interest). This model is widely used for representing the mode distribution in fan broadband noise.

However, this mode distribution model is observed to provide a poor fit to the experimental data for the nearly cut-off modes $|\alpha_{mn}|>0.99$ at the modal cut-on frequencies $ka=12.93$ and 14.12 shown in Figs. 10(b) and 10(d). These cases are characterized by relatively high condition numbers, i.e., $\kappa(\mathbf{D})=1746$ and $\kappa(\mathbf{D})=1135$, respectively. At these frequencies, the nearly cut-off modes, $|\alpha_{mn}|>0.99$, are considerably overestimated by the inversion procedure, as presented from the theoretical study detailed in Castres and Joseph.[6]

While the least-squares solution given in Eq. (6) is optimal when low conditioning is achieved with the use of a geodesic array, it was shown in Castres and Joseph[6] that a constrained least-squares solution is required to deduce the mode amplitudes at frequencies close to the cut-on frequencies. A novel regularization procedure that enhances the robustness of these inversion results at these cut-on frequencies is presented in the Appendix. The essence of the regularization technique is to constrain the modal solution for the nearly cut-off modes in order to make it more stable to variations of the noise perturbing the measurements.

As an alternative representation of the inverted mode amplitudes, Fig. 11 shows the mean square mode amplitudes plotted on a black and white scale against their spinning and radial mode number m and n, deduced from the geodesic sensor array at $ka=12.82$, 12.93, 13.78, and 14.12, respectively.

In general, the mean square mode amplitudes are observed to vary by no more than 10–12 dB. As in Fig. 10, modes near cutoff, i.e., those modes along the edge of the modal triangle, have greatest amplitude. Another interesting observation is that corotating modes, $m>0$, generally have a slightly higher amplitude than contrarotating modes relative to the direction of the fan rotation. This phenomenon is well-demonstrated for fan broadband noise, for example by Ganz *et al.*[13] However, this phenomenon is anticipated to be weak in the laboratory fan rig due to the relatively slow rotational speed of the fan.

D. The modal coherence function

The modal coherence function, defined here in Eq. (18), which quantifies the degree of statistical interdependence between any two modes $\hat{a}_i$, and $\hat{a}_j$, can be defined as follows:

$$\gamma_{\hat{a}_i\hat{a}_j}^2(\omega) = \frac{|S_{\hat{a}_i\hat{a}_j}(\omega)|^2}{S_{\hat{a}_i\hat{a}_i}(\omega)S_{\hat{a}_j\hat{a}_j}(\omega)} \tag{18}$$

and has the property $0 \leqslant \gamma_{\hat{a}_i\hat{a}_j} \leqslant 1$.

It is often assumed that the modes generated in an engine duct, in the broadband part of the pressure spectrum, are incoherent. This confirms the results found by Lewy *et al.*[14] Figure 12 shows the modal coherence function of Eq. (18) computed from the geodesic sensor arrays at the frequencies investigated in the previous sections. Note that the values $\gamma_{\hat{a}_i\hat{a}_i}=1$ have been omitted from Fig. 12 to aid clarity of presentation.

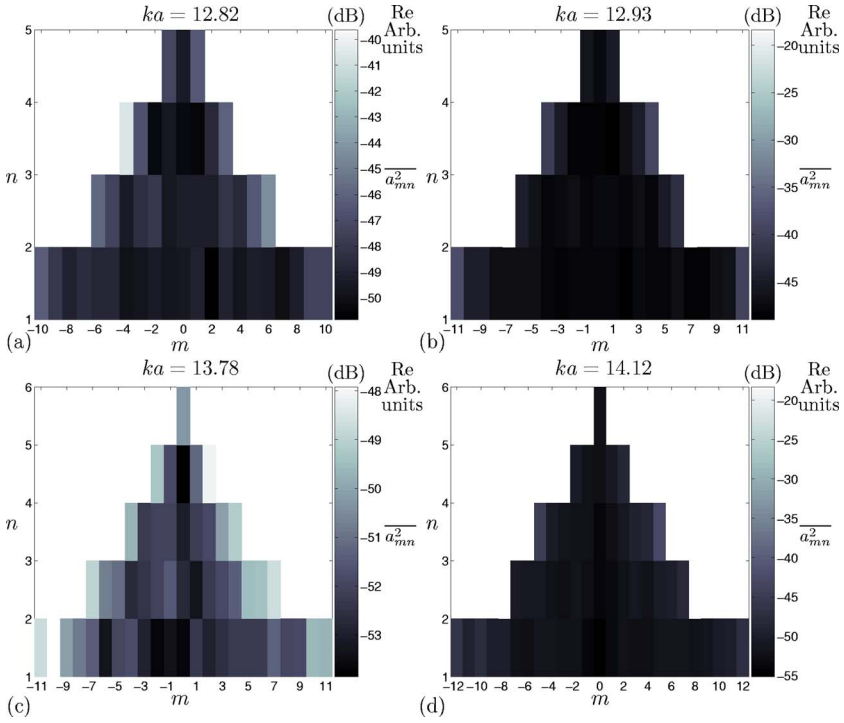


FIG. 11. (Color online) Mean square mode amplitudes $\overline{a_{mn}^2}$ against spinning and radial mode number m and n detected by the geodesic array: (a) $ka=12.82$, (b) $ka=12.93$, (c) $ka=13.78$, and (d) $ka=14.12$.

The modal coherence is generally found to be less than about 0.2 for most of the modal combinations at all frequencies investigated. Note that no significant coherence values between any modes is observed at $ka=12.82$ corresponding to 5 BPF at the 50 Hz fan speed. The absence of good modal coherence at these pseudotones is essentially due to the fact that the peaks observed in Fig. 8 are not pure tones created by rotor-stator interaction but the result of an interaction between upstream flow distortion in the duct and the fan blades. It is also possible that coherence levels may be improved by reducing the analysis bandwidth, which is currently $\Delta ka=0.1364$.

E. Comparison with the mean square mode amplitudes deduced from in-duct modal analysis

In order to provide independent validation of the mode amplitude inversion results obtained using the near-field sensor measurements presented in Sec. V C, a single ring of 24 wall flush-mounted microphones is used in the duct to perform a spinning modal decomposition. This measurement only allows the determination of amplitudes of individual spinning modes at the duct wall and not individual radial modes. The maximum spinning mode number M that can be inverted from the 24 microphones according to the sampling theorem is $M=11$. In the present duct, this allows the determination of the amplitudes of all cut-on modes up to $ka=13.8$. The cut-on frequency $ka=14.12$ studied in the previous sections is therefore not investigated in this section. Comparison will also be made at the cut-on frequency $ka=11.35$.

At a single frequency, the in-duct pressure field at the i^{th} microphone may be written as

$$p(\theta_i) = \sum_{m=-M}^{+M} a_m e^{-jm\theta_i}, \tag{19}$$

where a_m are the pressure mode amplitudes at the duct wall.

The acoustic pressure that propagates inside the duct, neglecting flow and reflections from the duct terminations, can be written in the cylindrical coordinate system (r,θ,z) as follows:

$$p(r,\theta,z) = \sum_{m=-\infty}^{+\infty} \sum_{n=1}^{+\infty} a_{mn} \Psi_{mn}(r,\theta) \exp(jk\sqrt{1-\alpha_{mn}^2}z), \tag{20}$$

where $\Psi_{mn}(r,\theta)$ are the normalized mode shape functions, which for a hard-walled cylindrical duct is given by

$$\Psi_{mn}(r,\theta) = \frac{1}{\sqrt{\Lambda_{mn}}} J_m(\kappa_{mn} r) e^{-jm\theta}, \tag{21}$$

where $J_m(x)$ is the Bessel function of first kind of order m and Λ_{mn} is chosen to ensure normalization condition $S^{-1}\int_S |\Psi|^2 dS=1$. If the ring of sensors is located at $z=0$, comparison of Eq. (19) with Eq. (20) allows the amplitudes a_m at the duct wall to be related to the mode amplitudes a_{mn} by

$$a_m = \sum_{n=1}^{+N} a_{mn} \frac{J_m(\kappa_{mn} a)}{\sqrt{\Lambda_{mn}}}, \tag{22}$$

where N is the maximum radial mode numbers cut on at a given frequency. Inverting Eq. (19) and taking the expectation $E[\overline{a_m^2}]$, the mean square mode amplitudes at the ring of microphones is given by

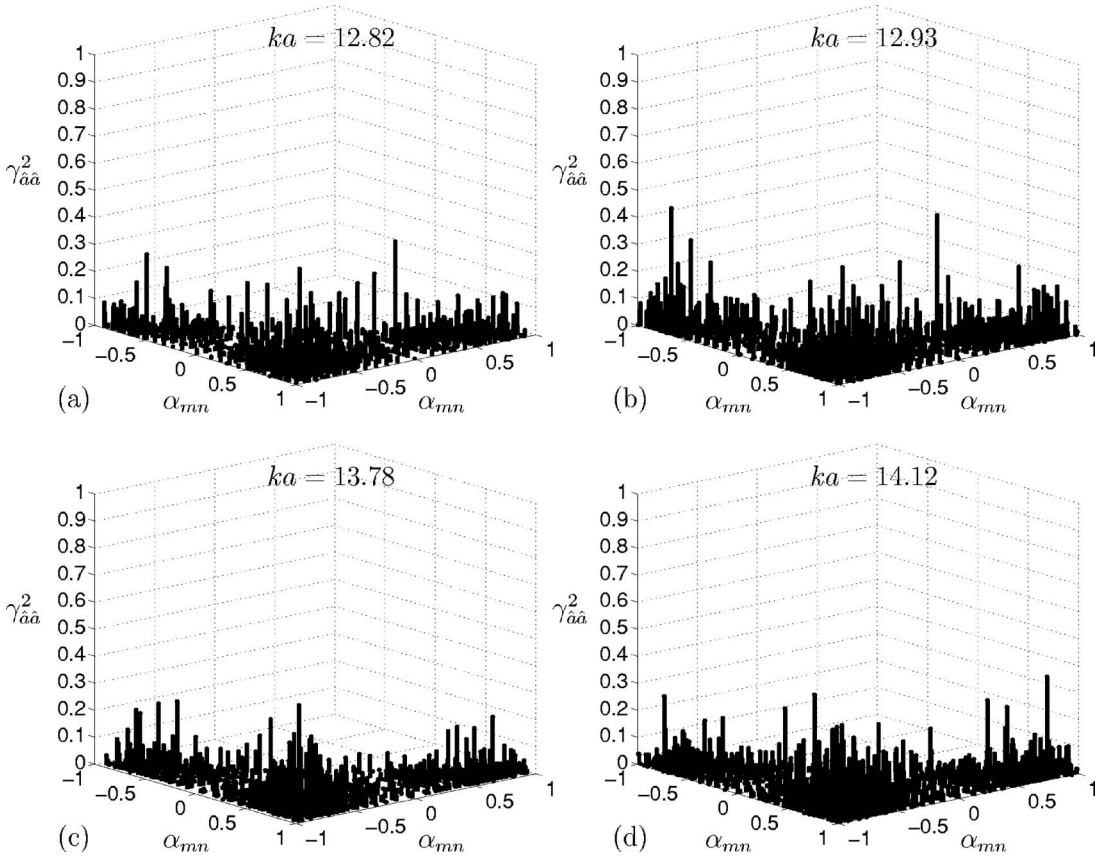


FIG. 12. Modal coherence function $\gamma^2_{\hat{a}\hat{a}}$ against the modal cut-on ratio α_{mn} detected by the geodesic array: (a) $ka=12.82$, (b) $ka=12.95$, (c) $ka=13.78$, and (d) $ka=14.12$.

$$\overline{a_m^2}(\omega) = \left(\frac{1}{K}\right)^2 \sum_{i=1}^{K} \sum_{j=1}^{K} S_{\hat{p}\hat{p}}(\theta_i, \theta_j, \omega) e^{jm(\theta_i - \theta_j)}, \tag{23}$$

where $S_{\hat{p}\hat{p}}(\theta_i, \theta_i, \omega)$ is the cross spectrum of measured pressures between any two microphones located at angular positions θ_i and θ_j at the duct wall. Assuming incoherent radial modes, as demonstrated in Fig. 12, following Eq. (22), the mean square mode amplitudes deduced from the in-duct microphones can readily be compared to that from the near-field sensor array by the following relationship:

$$\overline{a_m^2} = \sum_{n=1}^{+N} \overline{a_{mn}^2} \left| \frac{J_m(\kappa_{mn} a)}{\sqrt{\Lambda_{mn}}} \right|^2. \tag{24}$$

Figure 13 shows a comparison between the mean square mode amplitudes computed from the in-duct measurements and those from the geodesic array computed using Eq. (24), expressed in decibels relative to arbitrary units. To allow a clearer comparison, vertical arrows are used to connect two data points at the same m value. Note that the Tikhonov regularization scheme described in the Appendix is applied to the modal solution inverted from the near-field pressure measurements at the cut-on frequencies, namely at $ka=11.35$ and $ka=12.93$. Figure 13 shows generally good agreement between the two mode amplitude estimates to within 6 dB, with many amplitudes found to agree to within 3 dB or better.

VI. CONCLUSION

This paper has experimentally investigated an inverse technique for deducing the mode amplitudes of the broadband sound field radiated from a ducted fan. Pressure measurements were made under nearly free field conditions on a laboratory-scale fan inlet using a hemispherical structure to mount the microphones arranged in a geodesic geometry. The inverse technique, which used a FE/IE analysis to model the modal directivities of a laboratory-scale fan inlet in no-flow conditions, was performed to deduce the mode amplitudes of the broadband sound field. The estimated mean square mode amplitudes were plotted against their cut-on ratio. The distribution of the mode amplitudes was found to closely follow a model in which there is equal energy per mode. The modal inversion from the geodesic sensor array was compared with the results from conventional spinning mode decomposition using a single ring of in-duct microphones. Spinning mode amplitudes obtained using the two techniques were found to agree to within 6 dB, with many of the modes found to agree to within 3 dB or less. The inverse technique also allowed the estimation of the modal coherence, which quantifies the level of statistical interdependence between any two mode amplitudes. Low coherence values were observed at most frequencies, thereby confirming the commonly made assumption of uncorrelated mode amplitudes. A geodesic microphone array geometry has been identified as providing the most robust detection of mode amplitudes. This measurement technique, which allows for the first

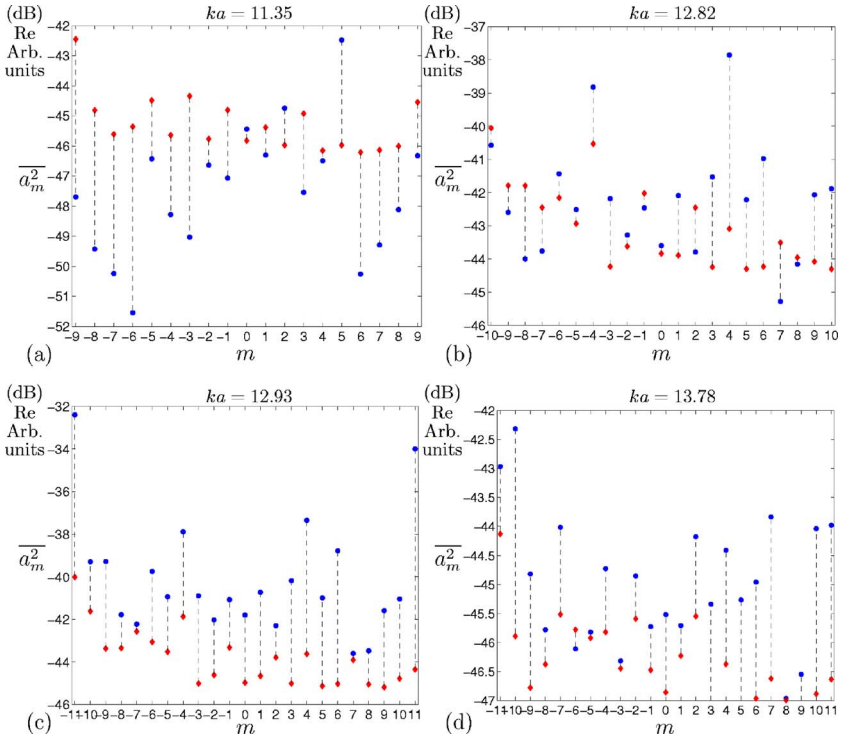


FIG. 13. (Color online) Mean square mode amplitudes $\overline{a_m^2}$ against spinning mode number m inverted from in-duct microphones (diamonds) and geodesic sensor array (circles): (a) $ka=11.35$, (b) $ka=12.82$, (c) $ka=12.93$, and (d) $ka=13.78$.

time an accurate estimate of broadband mode distribution, is a stepping-stone toward better understanding of the source mechanisms of broadband fan noise and forms an essential tool in developing strategies to control the sound field radiating at the inlet of the engine. However, there is significant scope for improving the accuracy of the technique. A forward problem which models the effects of realistic flows as well as the presence of liners within the inlet could be implemented. Such forward problem would account for an increasing number of cut-on modes. The robustness and accuracy of the inverse problem would then have to be studied. The unique microphone spreading of geodesic sensor arrays allows good detection of complicated modal radiation patterns. Therefore, further investigation of other classes of geodesic geometry[10] could allow a better coupling of the TCS sensor array with the modal information radiated by the inlet with different flow and liner configurations.

ACKNOWLEDGMENT

The authors would like to acknowledge Rolls-Royce plc, which through the Rolls-Royce University Technology Centre in Gas Turbine Noise at ISVR, supported this work.

APPENDIX: A REGULARIZATION TECHNIQUE FOR THE STABILIZATION OF THE MODAL INVERSION AT FREQUENCIES NEAR CUT-OFF

The constrained least-squares estimates for the mode amplitude cross spectral matrix $\mathbf{S}_{\hat{\mathbf{a}}_\mathbf{R}\hat{\mathbf{a}}_\mathbf{R}}$ is given by

$$\mathbf{S}_{\hat{\mathbf{a}}_\mathbf{R}\hat{\mathbf{a}}_\mathbf{R}} = \mathbf{D}^{\#}\mathbf{S}_{\hat{\mathbf{p}}\hat{\mathbf{p}}}(\mathbf{D}^{\#})^{\mathbf{H}}, \tag{A1}$$

where $\mathbf{D}^{\#}$ is the regularized inverse matrix of $\mathbf{D}$ given by Hansen[15] as

$$\mathbf{D}^{\#} = [\mathbf{D}^{\mathbf{H}}\mathbf{D} + \beta\mathbf{R}^{\mathbf{H}}\mathbf{R}]^{-1}\mathbf{D}^{\mathbf{H}} \tag{A2}$$

and now includes an additional term $\beta\mathbf{R}^{\mathbf{H}}\mathbf{R}$ that penalizes the norm of the modal solution. It was discussed in Castres and Joseph[6] that, in an ideally conditioned directivity matrix, the singular value spectrum would be perfectly flat and the solution would be maximally robust to measurement noise. It was also demonstrated that with the geodesic sensor array geometry, two small singular values associated to transformed modes, which only contained acoustic modes very close to cut-off (i.e., $\alpha_i > 0.99$), were responsible for ill-conditioning at frequencies near cut-off frequencies. In order

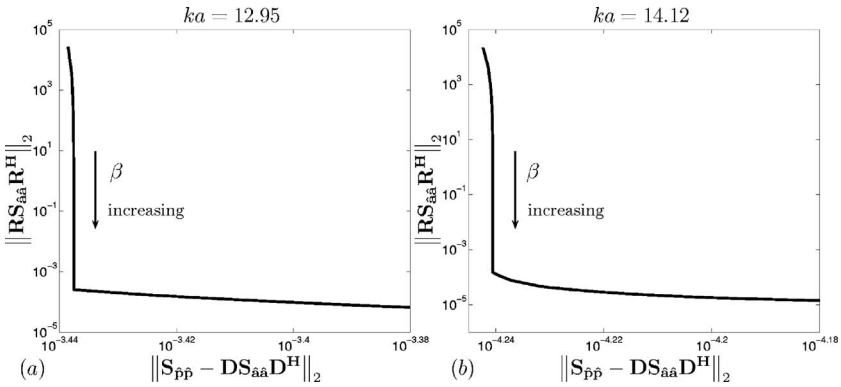


FIG. 14. L curve for optimal β parameter for broadband sound field at the cut-on frequencies.

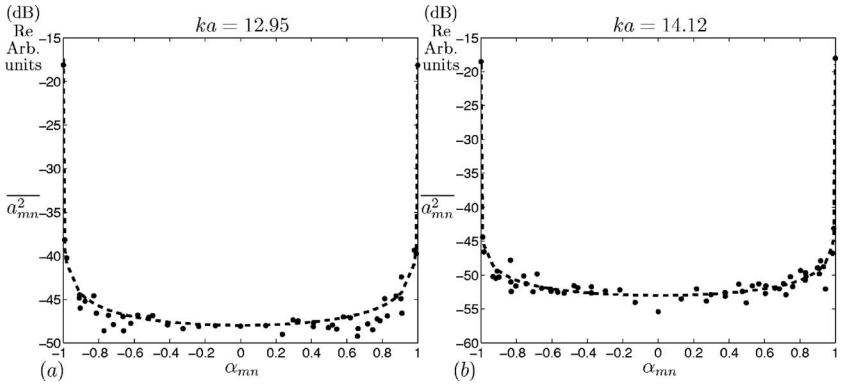


FIG. 15. Mean square mode amplitudes $\overline{a_{mn}^2}$ against the modal cut-on ratio α_{mn} inverted from geodesic array after regularization and Eq. (17) is plotted by a dashed line (a) $ka=12.93$ and (b) $ka=14.12$.

to assign a greater penalty to these near cut-off modes, a general form Tikhonov regularization scheme is applied to the broadband experimental data with the following regularizing matrix defined as follows:[16]

$$\mathbf{R} = \begin{bmatrix} \mathbf{I_{L-Q}} & \mathbf{0} \\ \mathbf{0} & \mathbf{S_Q} \end{bmatrix}, \quad \mathbf{S_Q} = \mathrm{diag}(\mathbf{1}/\|\mathbf{D_i}\|_2). \tag{A3}$$

Here, $\mathbf{S_Q}$ is a diagonal matrix whose elements comprise the norm of the column vectors of $\mathbf{D}$ for Q modes with $\alpha_i \geqslant 0.99$.

The appropriate perturbation bound κ_β, which is the equivalent of the condition number $\kappa(\mathbf{D})$ for the constrained least-squares solution, is given by[15,16]

$$\kappa_\beta = \frac{\|\mathbf{D}\|_2 \|\mathbf{R^{-1}}\|_2}{2\sqrt{\beta}}, \tag{A4}$$

where β is a specified weighting factor.

A generalization of the Hansen L-curve technique,[17] in which $\|\mathbf{S_{\hat{p}\hat{p}} - DS_{\hat{a}\hat{a}}D^H}\|_2$ is plotted against $\|\mathbf{RS_{\hat{a}\hat{a}}R^H}\|_2$, is used to determine the regularization parameter β that satisfies the best trade-off between the residuals and the errors introduced by constraining the solution. From Eq. (12), if the inlet sound field is not correlated with the external noise measured by the sensors, the norm of the residuals is given by $\|\mathbf{S_{\hat{p}\hat{p}} - DS_{\hat{a}\hat{a}}D^H}\|_2$ while the norm of errors introduced by constraining the solution is $\|\mathbf{RS_{\hat{a}\hat{a}}R^H}\|_2$. Figures 14(a) and 14(b) show the L-curves for $ka=12.93$ and $ka=14.12$, respectively, applied to the broadband sound field resulting from the general form of Tikhonov regularization. Using the assumption that the noise and the fan inlet measurements are uncorrelated, as given by Eq. (12), the error in the solution can therefore be written as follows:

$$\mathbf{S_{aa} - S_{\hat{a}\hat{a}} = S_{aa} - D^{\#}S_{pp}(D^{\#})^H - D^{\#}S_{nn}(D^{\#})^H}. \tag{A5}$$

Substituting Eq. (16) into Eq. (A5) and performing a generalized singular value decomposition (GSVD) of the pair $(\mathbf{D},\mathbf{R})$, is detailed in Hansen[15]

$$\mathbf{D = U\Sigma X^{-1}}, \quad \mathbf{R = VMX^{-1}} \tag{A6}$$

where Σ and $\mathbf{M}$ are diagonal matrices given by

$$\Sigma = \mathrm{diag}(\sigma_i), \quad \mathbf{M} = \mathrm{diag}(\mu_i) \tag{A7}$$

lead to the following solution error:

$$\mathbf{S_{aa} - S_{\hat{a}\hat{a}} = S_{aa} - X\Lambda X^H S_{aa} X\Lambda X^H - D^{\#}S_{nn}(D^{\#})^H}, \tag{A8}$$

where $\mathbf{X}$ is a nonsingular matrix that is the equivalent of $\mathbf{V}$ in the SVD of $\mathbf{D}$[6] for the GSVD of $(\mathbf{D},\mathbf{R})$ and Λ is a diagonal matrix, whose elements are the filter factors from the Tikhonov filter function given in terms of the generalized singular values γ_i and the Tikhonov parameter β as follows:[17]

$$f(\gamma_i) = \frac{\gamma_i^2}{\gamma_i^2 + \beta} = \frac{1}{1 + \beta/\gamma_i^2}. \tag{A9}$$

It can be readily seen from Eq. (A8) that, in the case of broadband noise, the errors due to regularization are controlled by the elements of the matrix $\mathbf{X\Lambda X^H}$, which represents the filtered modal components, i.e., $\Lambda_i = \gamma_i^2/(\gamma_i^2+\beta)$, while the term $\mathbf{D^{\#}S_{nn}(D^{\#})^H}$ represents error due to the presence of background noise in the measurements. When very little regularization is introduced, most of the filter factors tend to unity so that $\mathbf{X\Lambda X^H \mapsto I_L}$. The errors are then dominated by $\mathbf{D^{\#}S_{nn}(D^{\#})^H}$. The solution is said to be undersmoothed. This corresponds to the uppermost part of the L-curve, while the rightmost part corresponds to the oversmoothed solution (i.e., $\beta \gg \gamma_i$) when the filter factors are small. Figure 14 shows that the general form of Tikhonov regularization displays a sharp corner on the L curve. This regularizing scheme thus allows a sharp trade-off between the minimization of the residuals and that of the errors introduced by the constraint applied to the solution. In this scheme, a regularization parameter of $\beta = 8.5 \times 10^{-5}$ for $ka=12.93$ and $ka=14.12$ coincides with the L-curve corner and allows the perturbation bounds κ_β of 264 and 273, respectively, which, compared to the $\kappa(\mathbf{D})$ values of 1746 and 1135, respectively, found with the unconstrained least-squares method, are significantly reduced. Figures 15(a) and 15(b) show the mean square mode amplitudes resulting from the Tikhonov regularized solution of Eq. (A2) at $ka=12.93$ and 14.12, respectively. The mode amplitudes for the nearly cut-off modes, which are overestimated in Figs. 10(b) and 10(d), now follow the "equal energy per mode" model much more closely, suggesting that these new results are much closer to their actual values than when regularization is absent.

[1]M. Tyler and T. Sofrin, "Axial flow compressor noise studies," Soc. Automot. Eng. [Spec. Publ.] **70**, 309–332 (1962).
[2]R. Thomas, F. Farassat, L. Clark, C. Gerhold, J. Kelly, and L. Becker, "A

mode detection using the azimuthal directivity of a turbofan model," in *the Fifth AIAA/CEAS Aeroacoustics Conference*, Bellevue, WA, 1999, Paper No. AIAA 99-1954.
[3]F. Farassat, D. Nark, and H. Russel, "The detection of radiated modes from ducted fan engine," in *the Seventh AIAA/CEAS Aeroacoustics Conference*, Maastricht, The Netherlands, 2001, Paper No. AIAA 2001-2138.
[4]S. Lewy, "Inverse method predicting spinning modes radiated by a ducted fan from free-field measurements," J. Acoust. Soc. Am. **117**, 744–750 (2005).
[5]L. Enghardt, L. Neuhaus, and C. Lowis, "Broadband sound power determination in flow ducts," in *the AIAA/CEAS Aeroacoustics Conference*, Monterey, CA, 2004, Paper No. 2004-2940.
[6]F. Castres and P. Joseph, "Mode detection in turbofan inlets from near field sensor arrays," J. Acoust. Soc. Am. **121**, 796–807 (2007).
[7]G. Golub and C. Van Loan, *Matrix Computations* (North Oxford Academic, Oxford, 1983).
[8]P. Nelson and S. Yoon, "Estimation of acoustic source strength by inverse methods. I. Conditioning of the inverse problem," J. Sound Vib. **233**, 643–668 (2000).
[9]B. Fuller, "The age of the dome," Build International **2**(6), 7–15 (1969).
[10]H. Kenner, *Geodesic Math and How to Use It* (University of California Press, 1976).
[11]P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," IEEE Trans. Audio Electroacoust. **AU-15**, 70–73 (1967).
[12]P. Joseph, C. Morfey, and C. Lowis, "Multi-mode sound transmission in ducts with flow," J. Sound Vib. **264**, 523–544 (2003).
[13]U. Ganz, P. Joppa, T. Patten, and D. Scharpf, "Beoing 18-inch fan rig broadband noise test," Technical Rep. No. NASA/CR - 1998 - 208704, Boeing Commercial Airplane Group, 1998.
[14]S. Lewy, J. Lambourion, C. Malarmey, M. Perulli, and B. Rafine, "Direct experimental verification of the theoretical model predicting rotor noise generation," in *AIAA fifth Aeroacoustics Conference*, Seattle, WA, 1979, Paper No. AIAA 79-0658.
[15]C. Hansen, "Regularization, gsvd and truncated gsvd," Behring Inst. Mitt **29**, 491–504 (1989).
[16]C. Hansen, "Perturbations bounds for discrete Tikhonov regularization," Inverse Probl. **5**, L41–L44 (1989).
[17]C. Hansen, "Rank-deficient and discrete ill-posed problems," *SIAM Monographs on Mathematical Modeling and Computation*, SIAM, Philadelphia, 1988, p. 243.

Modeling of the roundabout noise impact

Rufin Makarewicz and Roman Golebiewski
Institute of Acoustics, A. Mickiewicz University, 61-614 Poznan, Umultowska 85, Poland

(Received 11 October 2006; revised 7 May 2007; accepted 24 May 2007)

A roundabout is a very popular tool used by town planners for carrying smooth and stationary road traffic flow. In this study it is shown that the replacement of a classical road intersection by a roundabout, under certain conditions, may produce a traffic noise decrease. These conditions are expressed in terms of the roundabout speed and the receiver location. The A-weighted sound exposure level is used to describe noise reduction.
 [DOI: 10.1121/1.2749700]

PACS number(s): 43.50.Lj [KA] Pages: 860–868

I. INTRODUCTION

Does the change of a road intersection into a roundabout reduce noise or not? The answer to this question is a main purpose of this study.

To calculate the roundabout noise, Lewis and James[1] replaced the circle by a series of contiguous arcs and calculated the total noise by summing the contributions due to individual arcs. Then, To and Chan[2] assumed that the roundabout can be represented as a linear source of circular shape, located on the ground. They derived a simple equation that relates the noise level at any distance from the roundabout to the noise level measured at the center of the roundabout. The roundabout traffic noise was measured and analyzed by Berengier[3] and Bertoni.[4] Recently, Picaut *et al.*[5] have calculated the roundabout noise with the nondirectional point source model of a vehicle and the ground effect. In the present study we follow their way of reasoning and use the vehicle sound power calculated on the basis of the Japanese ASJ RTN model.[6]

II. ROAD INTERSECTION VERSUS ROUNDABOUT

The roundabout of the radius R (Fig. 1) replaces a regular intersection (Fig. 2) with four arms of length R and the stop signs at its center ($x=0; y=0$). Along the arms ($-R \le x < 0$), ($0 \le x < +R$), ($-R \le y < 0$), ($0 \le y < +R$), both deceleration (vehicle approach) and acceleration (vehicle departure) take place. Suppose the probabilities of turning right, moving straight ahead, and turning left are identical for each arm (Fig. 3). Consequently, the resultant probability of deceleration (solid line) and acceleration (dashed line) are equal to each other (Fig. 4).

Now, we introduce the A-weighted sound exposure,

$$E = \int_0^\tau p_A^2(t)\, dt, \tag{1}$$

where τ is the time interval of noise emission and p_A^2 symbolizes the A-weighted squared sound pressure of noise generated by a single vehicle. The deceleration and acceleration noise, emitted from the intersection arms, are characterized by

$$E_d = E_d(-R \le x < 0) + E_d(0 \le x < +R) + E_d(-R \le y < 0) + E_d(0 \le y < +R) \tag{2}$$

and

$$E_a = E_a(-R \le x < 0) + E_a(0 \le x < +R) + E_a(-R \le y < 0) + E_a(0 \le y < +R), \tag{3}$$

respectively. Finally, the representative noise of the entire intersection is given by

$$E_{in} = E_d + E_a. \tag{4}$$

Similarly to the road intersection, a roundabout is characterized by identical probabilities of turning right, moving straight ahead, and turning left (Fig. 5). There are four possible approach roads. Summing up the above probabilities, one gets a representative event of the roundabout noise: a single vehicle completes the circle ($0 \le \phi < 2\pi$) at a constant speed V_R (Fig. 6). Introducing four quadrants of the roundabout, ($0 \le \phi < \pi/2$), ($\pi/2 \le \phi < \pi$), ($\pi \le \phi < 3\pi/2$), and ($3\pi/2 \le \phi < 2\pi$), one arrives at the corresponding value of the A-weighted sound exposure,

$$E_{ro} = E_{ro}(0 \le \phi < \pi/2) + E_{ro}(\pi/2 \le \phi < \pi) + E_{ro}(\pi \le \phi < 3\pi/2) + E_{ro}(3\pi/2 \le \phi < 2\pi). \tag{5}$$

The condition of noise reduction by a roundabout is as follows [Eqs. (4) and (5)]:

$$\Delta L = 10 \log \left\{ \frac{E_{ro}}{E_d + E_d} \right\} < 0. \tag{6}$$

We will derive the explicit form of the above relationship.

Actually, there are a few categories of road vehicles. To show the salient features of the model, only automobiles are considered.

III. A-WEIGHTED SQUARED SOUND PRESSURE

Experimental data are available for automobile noise directivity; however, in Ref. 7 it has been shown that the nondirectional monopole is a feasible representation of a real automobile. Tire-road interaction is the primary source of sound waves at steady speed and automobile deceleration. When an automobile accelerates, its engine becomes the predominant source. We assume that in any operating condi-

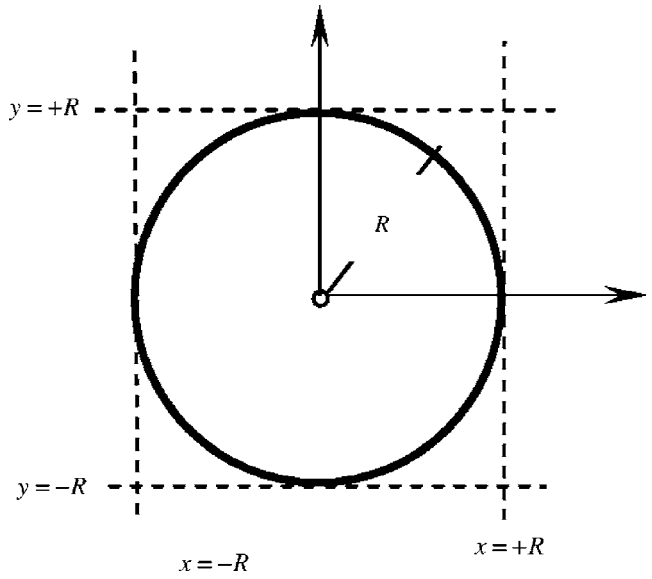


FIG. 1. A roundabout of radius R in the rectangular coordinate system (x,y).

tions, the receiver height exceeds the source height, $H \gg h$ (Fig. 7). Accordingly, the propagation distance between the source $S(x,y,h)$ and the receiver $O(X,Y,H)$ can be approximated by

$$d \approx \sqrt{(x-X)^2 + (y-Y)^2 + H^2}. \tag{7}$$

Under a semi-free field type of propagation conditions, the A-weighted squared sound pressure of a monopole is

$$p_A^2 = \frac{W_A(V)\rho c}{4\pi d^2} G(d), \tag{8}$$

where ρc denotes the characteristics impedance of air, $W_A(V)$ expresses the dependence of the A-weighted sound power on the vehicle speed V, and the function $G(d)$ quantifies the ground effect. With $H \gg h$, one can write (Refs. 8–10)

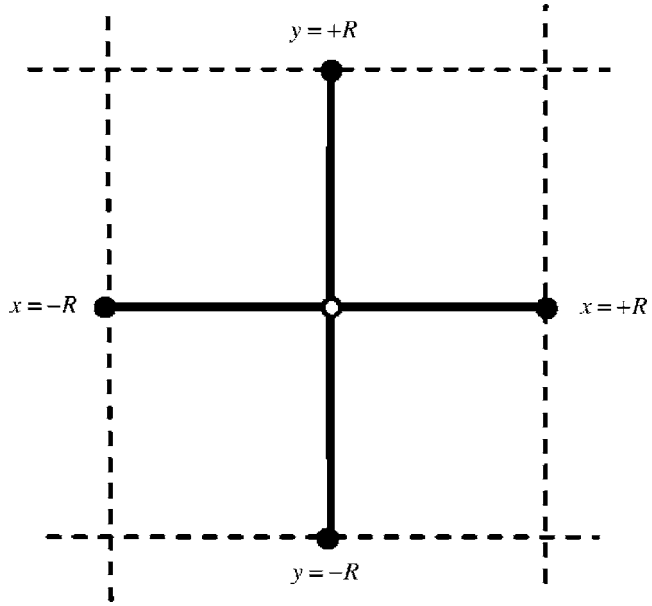


FIG. 2. A road intersection equivalent to a roundabout (Fig. 1).

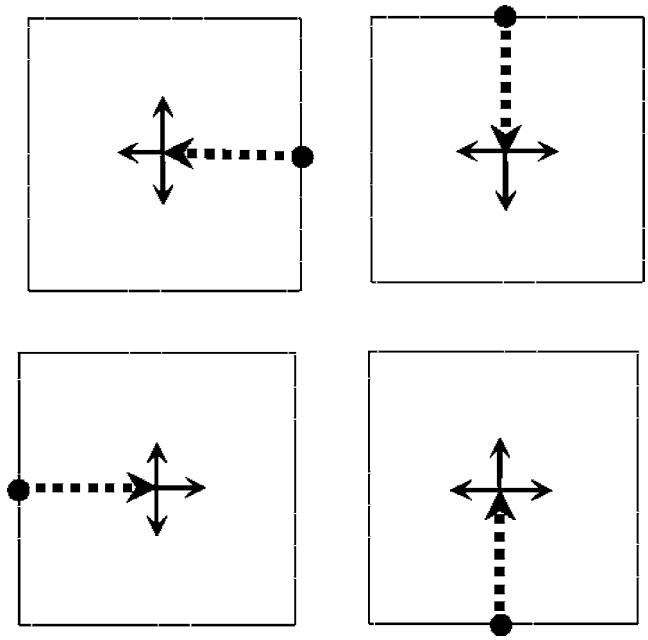


FIG. 3. Probabilities of turning right, moving straight ahead, and turning left are the same for four arms of the road intersection (Fig. 2). Black point denotes the intersection entrance.

$$G(d) \approx \frac{\beta}{1 + (1/\varsigma)(d/H)^2}. \tag{9}$$

The value of the nondimensional ground parameter ς grows with the surface impedance. Above a hard surface with large value of ς, with the receiver so close to the source that

$$d \ll \sqrt{\varsigma} H, \tag{10}$$

Eq. (9) yields $G \to \beta$ and Eq. (8) simplifies to the form

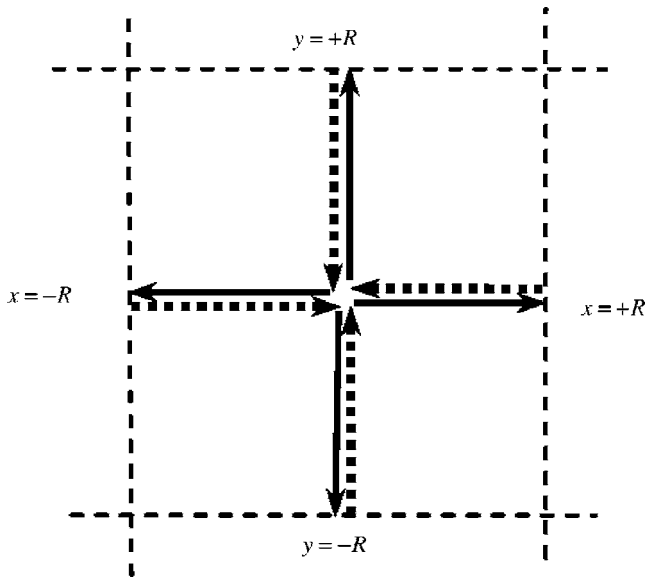


FIG. 4. Probabilities of deceleration (dotted line) and acceleration (solid line) along the arms of the road intersection are equal to each other.

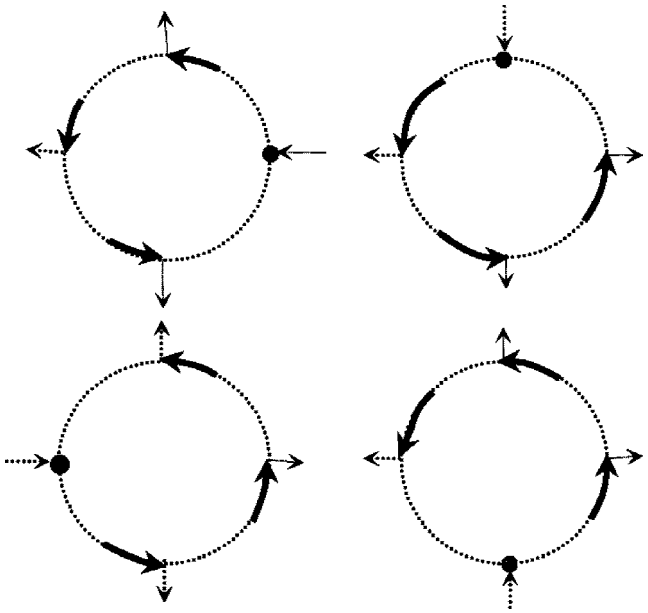


FIG. 5. Probabilities of turning right, moving straight ahead, and turning left (and then leaving the roundabout) remain the same as the probabilities for the road intersection (Fig. 3). Black point denotes the roundabout entrance.

$$p_A^2 = \frac{\beta W_A(V)\rho c}{4\pi d^2}. \tag{11}$$

One can see that the ground reflection at a hard surface brings about a virtual change of the A-weighted sound power, $W_A \rightarrow \beta W_A$. In other words, the sound wave propagation is governed solely by geometrical spreading. To separate the geometrical spreading and the reflection at a real ground surface, we rewrite Eq. (8) as follows,

$$p_A^2 = \frac{\beta W_A(V)\rho c}{4\pi}\left[\frac{1}{d^2} - \frac{1}{d^2 + \varsigma H^2}\right]. \tag{12}$$

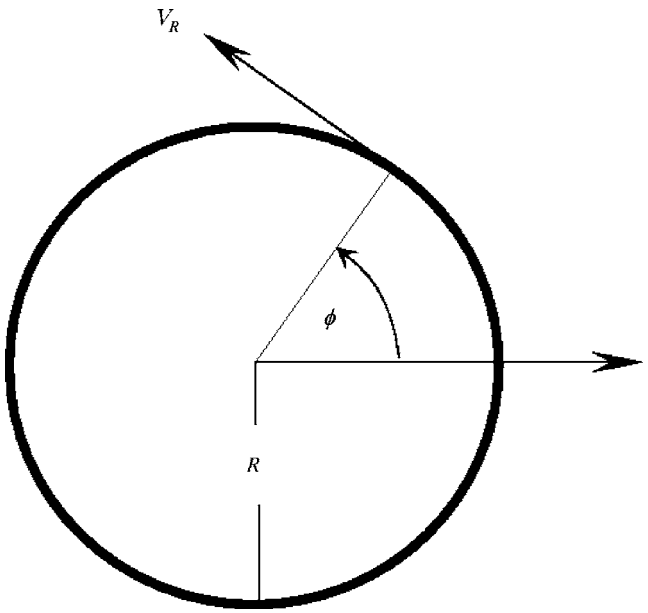


FIG. 6. Resultant distribution of probabilities shown in Fig. 5 corresponds to the completion of the full circle by a vehicle ($0 \le \phi < 2\pi$).

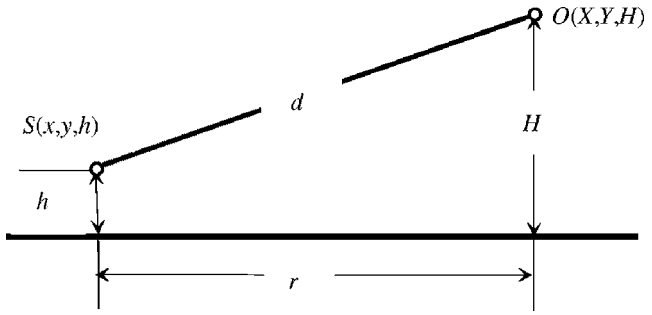


FIG. 7. The distance d [Eq. (7)] between the source $S(x,y,h)$ and the receiver $O(X,Y,H)$.

The A-weighted squared sound pressure p_A^2 and the directly measurable A-weighted sound pressure level L_A are related to each other,

$$L_A = 10 \lg \{p_A^2/p_o^2\}, \tag{13}$$

where the reference sound pressure $p_o = 20\ \mu$Pa.

The simultaneous measurement of the A-weighted sound pressure levels, L_{A1} and L_{A2}, by two microphones, (d_1, H_1) and (d_2, H_2), gives the value of ground parameter [Eqs. (8), (9), and (13)]

$$\varsigma = \frac{(d_2/H_2)^2 - m(d_1/H_1)_2}{m-1}, \tag{14}$$

where

$$m = \left(\frac{d_1}{d_2}\right)^2 10^{0.1[L_{A1} - L_{A2}]}. \tag{15}$$

Example: Let us take the microphone heights $H_1 = H_2 = 1$ m and the distances $d_1 = 15$, $d_2 = 30$ m. For the difference $L_{A1} - L_{A2} = 6.1$ dB (typical of a concrete pavement) and $L_{A1} - L_{A2} = 8.0$ dB (typical of a grassy area), one gets $\varsigma = 36\,360$ (hard ground) and $\varsigma = 944$ (soft ground). Thus, the inequalities

$$5 \times 10^2 < \varsigma < 5 \times 10^4 \tag{16}$$

seem to define the variability range of the ground parameter.

IV. A-WEIGHTED SOUND EXPOSURE

The results of noise measurements in many countries lead to the following relationship between the A-weighted sound power level and the vehicle speed:

$$L_{WA} = L + m \lg (V/V_o). \tag{17}$$

To get the nondimensional argument of the logarithmic function, the vehicle speed V is divided by the unit velocity, $V_o = 1$ km/h. The constant L is discussed below. Japanese and European measurements of automobile noise (Refs. 6 and 11) indicate that at a *s*teady speed V and during deceleration (with predominant tire-road noise)

$$L_{WA}^{(sd)} = L_{sd} + 30 \lg (V/V_o), \tag{18}$$

where $L_{sd} = 46.7$ dB. Measurements of noise produced by automobiles during acceleration (with predominant engine noise) yield (Refs. 6 and 11)

$$L_{WA}^{(a)} = L_a + 10\lg(V/V_o), \tag{19}$$

where $L_a = 82.3$ dB is a function of the average engine speed. However, the vehicle speed V and the engine speed are related to each other, therefore one can write $L_{WA}^{(a)}$ as a function of the vehicle speed V only. Obviously, the values of L_{sd} [Eq. (18)] and L_a [Eq. (19)] depend on the type of pavement.

Making use of the definition of the A-weighted sound power level, one gets the A-weighted sound power for a steady speed and deceleration (superscripts *sd*) and the A-weighted sound power for acceleration (superscript *a*),

$$W_A^{(sd)} = W_{sd} \cdot (V/V_o)^3, \quad W_A^{(a)} = W_a \cdot (V/V_o). \tag{20}$$

Here

$$W_{sd} = W_o \times 10^{0.1L_{sd}}, \quad W_a = W_o \times 10^{0.1L_a}, \tag{21}$$

where the reference sound power $W_o = 10^{-12}$ (Watts). The values of L_{sd}, L_a are supposed to be known from measurements [Eqs. (18) and (19)].

For the instantaneous speed V, the infinitesimal displacement of a vehicle along its trajectory equals $dl = Vdt$, and the A-weighted sound exposure [Eq. (1)] can be rewritten as

$$E = \int_l \frac{p^2 A}{V}\, dl, \tag{22}$$

where l is length of the road segment. Accordingly, the A-weighted sound exposure for steady speed and deceleration can be expressed as follows [Eqs. (12), and (20)–(22)]:

$$E_{sd} = \frac{\rho c}{4\pi}\int_0^l S_{sd}\left[\frac{1}{d^2} - \frac{1}{d^2 + \varsigma H^2}\right] dl. \tag{23}$$

Similarly, the acceleration noise is characterized by the A-weighted sound exposure

$$E_a = \frac{\rho c}{4\pi}\int_0^l S_a\left[\frac{1}{d^2} - \frac{1}{d^2 + \varsigma H^2}\right] dl. \tag{24}$$

Here the noise emission quantifies the functions of the vehicle speed,

$$S_{sd} = \frac{W_o}{V_o}10^{0.1L_{sd}}\left(\frac{V}{V_o}\right)^2, \quad S_a = \frac{W_o}{V_o}10^{0.1L_a}. \tag{25}$$

Dimensional analysis shows that S_{sd} and S_a are related to the sound energy emitted from the line source of unit length (Joules per meter). The independence of S_a on the instantaneous value of vehicle speed during acceleration V is a little strange; however, it is based on thousands of measurements (Refs. 6 and 11).

The A-weighted sound exposure of noise emitted during either steady motion or deceleration can be calculated from [Eqs. (23)–(25)]

$$E_{sd} = \frac{W_o \rho c}{4\pi V_o R}10^{0.1L_{sd}}[F_{sd}(0) - F_{sd}(\varsigma)], \tag{26}$$

where

$$F_{sd}(\varsigma) = \frac{R}{V_o^2}\int_l \frac{V^2\, dl}{d^2 + \varsigma H^2}, \tag{27}$$

with the distance d defined by Eq. (7). Similarly, Eqs. (23)–(25) yield the A-weighted sound exposure of noise emitted during acceleration,

$$E_a = \frac{W_o \rho c}{4\pi V_o R}10^{0.1L_a}[F_a(0) - F_a(\varsigma)], \tag{28}$$

where

$$F_a(\varsigma) = R\int_l \frac{dl}{d^2 + \varsigma H^2}. \tag{29}$$

When a vehicle moves along the x axis or y axis (Fig. 2), $dl = \pm dx$ and $dl = \pm dy$, respectively. If the vehicle moves along the roundabout arc of radius R, then $dl = Rd\phi$, where $d\phi$ is the angle increment (Fig. 6).

A. Road intersection

1. Deceleration noise

Solid lines in Fig. 4 represent the vehicle deceleration along four arms of the intersection: $(-R \le x < 0)$, $(0 \le x < +R)$, $(-R \le y < 0)$, $(0 \le y < +R)$. Because of the stop sign, the vehicle speed $V = V_R$ (at the entrance into an arm) decreases to $V = 0$ at the intersection center $(x = 0, y = 0)$. At any distance $0 \le l \le R$ from the center the vehicle speed V can be calculated from

$$V^2(l) = 2a_d l, \tag{30}$$

where a_d denotes a constant deceleration. For $V(\pm R) = V_R$ one gets $V_R^2 = 2a_d R$ and Eq. (30) takes the from

$$V^2(l) = V_R^2 \frac{l}{R}. \tag{31}$$

Ultimately, Eq. (27) can be rewritten as

$$F_d(\varsigma) = \left(\frac{V_R}{V_o}\right)^2 \int_l \frac{l\, dl}{d^2 + \varsigma H^2}. \tag{32}$$

The propagation of noise, generated during deceleration along the x axis (from $x = -R$ to $x = 0$ and from $x = +R$ to $x = 0$), corresponds to the distance d defined by Eq. (7) with $y = 0$. Thus,

$$F_d(-R \le x < 0, \varsigma) = \left(\frac{V_R}{V_o}\right)^2 \int_{-R}^0 \frac{x\, dx}{(x - X)^2 + Y^2 + (\varsigma + 1)H^2} \tag{33}$$

and

$$F_d(+R \le x < 0, \varsigma) = \left(\frac{V_R}{V_o}\right)^2 \int_{+R}^0 \frac{x\,(-dx)}{(x - X)^2 + Y^2 + (\varsigma + 1)H^2}. \tag{34}$$

Finally, the deceleration noise coming from the x axis is characterized by the sum

$$F_d^{(x)}(\varsigma) = F_d(-R \le x < 0, \varsigma) + F_d(+R \le x < 0, \varsigma), \quad (35)$$

which equals [Eqs. (33)–(35)]

$$F_d^{(x)}(\varsigma) = \left(\frac{V_R}{V_o}\right)^2 \left\{ \ln \sqrt{\frac{(X-R)^2 + r_y^2}{(X+R)^2 + r_y^2}} + \frac{X}{r_y}\left[\arctan\left(\frac{R+X}{r_y}\right) + \arctan\left(\frac{R-X}{r_y}\right)\right]\right\}, \quad (36)$$

where

$$r_y^2 = Y^2 + (1+\varsigma)H^2. \quad (37)$$

Following the same way of reasoning, the deceleration noise coming from the y axis ($-R \le y < 0$) and ($+R \le y < 0$) is described by the function

$$F_d^{(y)}(\varsigma) = \left(\frac{V_R}{V_o}\right)^2 \left\{ \ln \sqrt{\frac{(Y-R)^2 + r_x^2}{(Y+R)^2 + r_x^2}} + \frac{Y}{r_x}\left[\arctan\left(\frac{R+Y}{r_x}\right) + \arctan\left(\frac{R-Y}{r_x}\right)\right]\right\}, \quad (38)$$

where

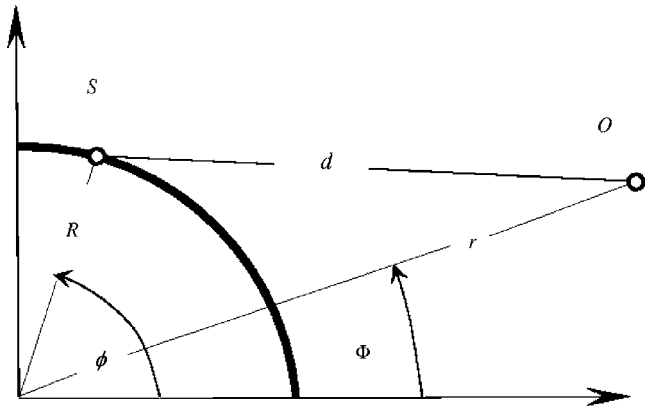


FIG. 8. The distance d (Fig. 7), in the polar coordinate system (r, ϕ), for the vehicle moving along the first quadrant of the roundabout [Eqs. (49), (51), and (52)].

$$r_x^2 = X^2 + (1+\varsigma)H^2. \quad (39)$$

Introducing the polar coordinate system with $X = r\cos(\Phi)$ and $Y = r\sin(\Phi)$ (see Figs. 7 and 8), the A-weighted sound exposure of the deceleration noise emitted from four arms of the road intersection (Fig. 2) is determined by [Eqs. (2), (26), (36), and (38)]

$$E_d = \frac{W_o \rho c}{4\pi V_o R} 10^{0.1 L_{sd}} \left(\frac{V_R}{V_o}\right)^2 [A(0) - A(\varsigma)], \quad (40)$$

where

$$A(\varsigma) = \ln\left\{ \sqrt{\frac{r^2 - 2Rr\cos\Phi + R^2 + (1+\varsigma)H^2}{r^2 + 2Rr\cos\Phi + R^2 + (1+\varsigma)H^2}} \sqrt{\frac{r^2 - 2Rr\sin\Phi + R^2 + (1+\varsigma)H^2}{r^2 + 2Rr\sin\Phi + R^2 + (1+\varsigma)H^2}}\right\} + \frac{r\cos\Phi}{\sqrt{r^2\sin^2\Phi + (1+\varsigma)H^2}} \arctan\left(\frac{2R\sqrt{r^2\sin^2\Phi + (1+\varsigma)H^2}}{r^2 - R^2 + (1+\varsigma)H^2}\right) + \frac{r\sin\Phi}{\sqrt{r^2\cos^2\Phi + (1+\varsigma)H^2}} \arctan\left(\frac{2R\sqrt{r^2\cos^2\Phi + (1+\varsigma)H^2}}{r^2 - R^2 + (1+\varsigma)H^2}\right) \quad (41)$$

2. Acceleration noise

Along the four arms of the intersection ($-R \le x < 0$), ($0 \le x < +R$), ($-R \le y < 0$), ($0 \le y < +R$), both deceleration and acceleration occur. The noise generated during the acceleration along the x axis ($-R \le x < 0$ and $0 \le x < +R$) is quantified by two functions [Eq. (7) with $y = 0$ and Eq. (29)],

$$F_a(-R \le x < 0, \varsigma) = R\int_{-R}^{0} \frac{dx}{(x-X)^2 + Y^2 + (\varsigma+1)H^2} \quad (42)$$

and

$$F_a(0 \le x < +R, \varsigma) = R\int_{+R}^{0} \frac{x\,(-dx)}{(x-X)^2 + Y^2 + (\varsigma+1)H^2}, \quad (43)$$

respectively. The sum of the above functions,

$$F_a^{(x)}(\varsigma) = F_a(-R < x < 0, \varsigma) + F_a(0 < x < +R, \varsigma), \quad (44)$$

can be written as [Eqs. (42) and (43)]

$$F_a^{(x)}(\varsigma) = \frac{R}{r_y}\left[\arctan\left(\frac{R+X}{r_y}\right) + \arctan\left(\frac{R-X}{r_y}\right)\right], \quad (45)$$

where the distance r_y is determined by Eq. (37). The acceleration noise from the y axis ($-R \le y < 0$ and $0 \le y < +R$) corresponds to the function

$$F_a^{(y)}(\varsigma) = \frac{R}{r_x}\left[\arctan\left(\frac{R+Y}{r_x}\right) + \arctan\left(\frac{R-Y}{r_x}\right)\right], \quad (46)$$

where r_x is given by Eq. (39). Consequently, Eqs. (3) and (28) yield the A-weighted sound exposure of the acceleration noise,

$$E_a = \frac{W_o \rho c}{4\pi V_o R} 10^{0.1 L_a}[B(0) - B(\varsigma)]. \quad (47)$$

With the receiver coordinates $X = r\cos\Phi$ and $Y = r\sin\Phi$ (Figs. 7 and 8), Eqs. (45) and (46) combine into

$$B(\varsigma) = \frac{R}{\sqrt{r^2\sin^2\Phi + (1+\varsigma)H^2}} \arctan\left(\frac{2R\sqrt{r^2\sin^2\Phi + (1+\varsigma)H^2}}{r^2 - R^2 + (1+\varsigma)H^2}\right) + \frac{R}{\sqrt{r^2\cos^2\Phi + (1+\varsigma)H^2}} \arctan\left(\frac{2R\sqrt{r^2\cos^2\Phi + (1+\varsigma)H^2}}{r^2 - R^2 + (1+\varsigma)H^2}\right) \quad (48)$$

B. Roundabout

The transformation of the road intersection into the roundabout changes the vehicle trajectory (Figs. 2 and 6). It has been shown (Sec. II) that the deceleration and acceleration along four intersection arms (Fig. 4) is equivalent to completion of the circle by a single vehicle in the roundabout at a constant speed $V = V_R$. Applying the cosine law to Fig. 8, one gets the vehicle-receiver distance [Eq. (7)],

$$d^2 = r^2 + R^2 + 2Rr\cos(\phi - \Phi) + H^2. \quad (49)$$

Then, introducing the increment of the circle length $dl = Rd\phi$, Eq. (27) can be rearranged into

$$F_{ro}(\varsigma) = \left(\frac{V_R}{V_o}\right)^2 R^2 \int_0^{2\pi} \frac{d\phi}{d^2(\phi) + \varsigma H^2}. \quad (50)$$

To find the above integral, we write

$$a = r^2 + R^2 + (1+\varsigma)H^2, \quad b = 2Rr, \quad (51)$$

and divide the roundabout into four quadrants, $(0 \le \phi < \pi/2)$, $(\pi/2 \le \phi < \pi)$, $(\pi \le \phi < 3\pi/2)$, and $(3\pi/2 \le \phi < 2\pi)$.

The noise contribution of the first quadrant $(0 \le \phi < \pi/2)$ (Fig. 8) can be calculated from

$$F_{ro}(0 \le \phi < \pi/2, \varsigma) = \left(\frac{V_R}{V_o}\right)^2 R^2 \int_0^{\pi/2} \frac{d\phi}{a - b\cos(\phi - \Phi)}, \quad (52)$$

so that

$$F_{ro}(0 \le \phi < \pi/2, \varsigma) = \left(\frac{V_R}{V_o}\right)^2 \frac{2R^2}{\sqrt{a^2 - b^2}}\left\{\arctan\left[\frac{a+b}{\sqrt{a^2-b^2}}\tan\left(\frac{\pi}{4} - \frac{\Phi}{2}\right)\right] + \arctan\left[\frac{a+b}{\sqrt{a^2-b^2}}\tan\left(\frac{\Phi}{2}\right)\right]\right\}. \quad (53)$$

The noise contributions from the second, third, and fourth quadrants are described by

$$F_{ro}(\pi/2 \le \phi < \pi, \varsigma) = \left(\frac{V_R}{V_o}\right)^2 \frac{2R^2}{\sqrt{a^2 - b^2}}\left\{\arctan\left[\frac{a}{\sqrt{a^2-b^2}}\tan\left(\frac{\pi}{4} - \frac{\Phi}{2}\right) + \frac{b}{\sqrt{a^2-b^2}}\right] + \arctan\left[\frac{a}{\sqrt{a^2-b^2}}\tan\left(\frac{\Phi}{2}\right) - \frac{b}{\sqrt{a^2-b^2}}\right]\right\}, \quad (54)$$

$$F_{ro}(\pi \le \phi < 3\pi/2, \varsigma) = \left(\frac{V_R}{V_o}\right)^2 \frac{2R^2}{\sqrt{a^2 - b^2}}\left\{\arctan\left[\frac{a-b}{\sqrt{a^2-b^2}}\tan\left(\frac{\pi}{4} - \frac{\Phi}{2}\right)\right] + \arctan\left[\frac{a-b}{\sqrt{a^2-b^2}}\tan\left(\frac{\Phi}{2}\right)\right]\right\} \quad (55)$$

and

$$F_{ro}(3\pi/2 \le \phi < 2\pi, \varsigma) = \left(\frac{V_R}{V_o}\right)^2 \frac{2R^2}{\sqrt{a^2 - b^2}}\left\{\arctan\left[\frac{a}{\sqrt{a^2-b^2}}\tan\left(\frac{\pi}{4} - \frac{\Phi}{2}\right) - \frac{b}{\sqrt{a^2-b^2}}\right] + \arctan\left[\frac{a}{\sqrt{a^2-b^2}}\tan\left(\frac{\Phi}{2}\right) + \frac{b}{\sqrt{a^2-b^2}}\right]\right\}, \quad (56)$$

respectively. Mindful that

$$F_{ro}(0 \le \phi < \pi/2, \varsigma) + F_{ro}(\pi \le \phi < 3\pi/2, \varsigma) = 2\left(\frac{V_R}{V_o}\right)^2 \frac{R^2}{\sqrt{a^2 - b^2}}\left\{\pi - \arctan\left[\frac{a\sqrt{a^2-b^2}}{b^2\sin\phi\cos\phi}\right]\right\} \quad (57)$$

and

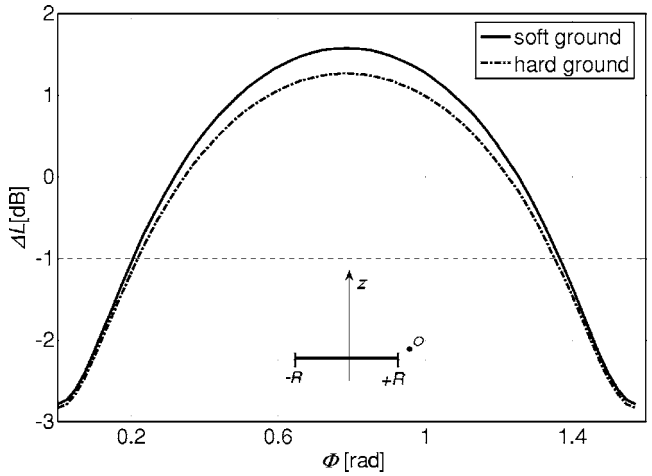


FIG. 9. Noise reduction ΔL [Eqs. (61) and (62)] for the roundabout traffic speed V_R=30 (km/h), the soft (ς=944) and hard (ς=36 360) ground, the receiver $O(r,H,\phi)$ close to the roundabout, r=0.1R, and close to the ground surface, H=0.1R.

$$F_{ro}(\pi/2 \le \phi < \pi, \varsigma) + F_{ro}(3\pi/2 \le \phi < 2\pi, \varsigma) = 2\left(\frac{V_R}{V_o}\right)^2 \frac{R^2}{\sqrt{a^2-b^2}} \arctan\left[\frac{a\sqrt{a^2-b^2}}{b^2 \sin\phi \cos\phi}\right], \quad (58)$$

one gets the A-weighted sound exposure of noise coming from the cruising vehicle (at a constant speed V_R) along the roundabout [Eqs. (5), (26), (51), (57), and (58)]:

$$E_{ro} = \frac{W_o \rho c}{4\pi V_o R} 10^{0.1 L_{sd}} \left(\frac{V_R}{V_o}\right)^2 [C(0) - C(\varsigma)], \quad (59)$$

where

$$C(\varsigma) = \frac{2\pi R^2}{\sqrt{(r+R)^2 + (1+\varsigma)H^2}\sqrt{(r-R)^2 + (1+\varsigma)H^2}}. \quad (60)$$

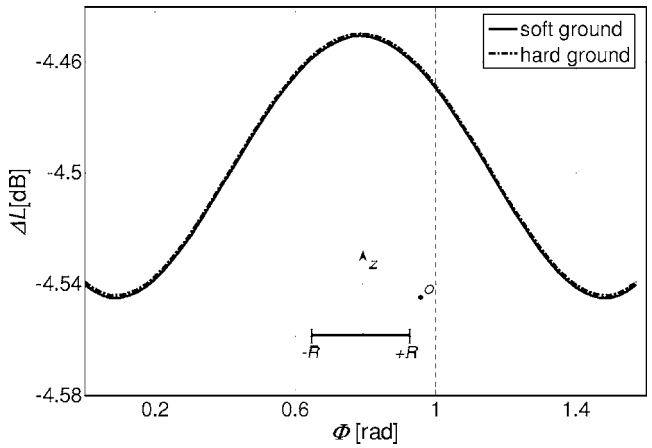


FIG. 10. Noise reduction ΔL [Eqs. (61) and (62)] for the roundabout traffic speed V_R=30(km/h), the soft (ς=944) and hard (ς=36 360) ground, the receiver $O(r,H,\Phi)$ close to the roundabout, r=0.1R, and high above the ground surface, H=R.

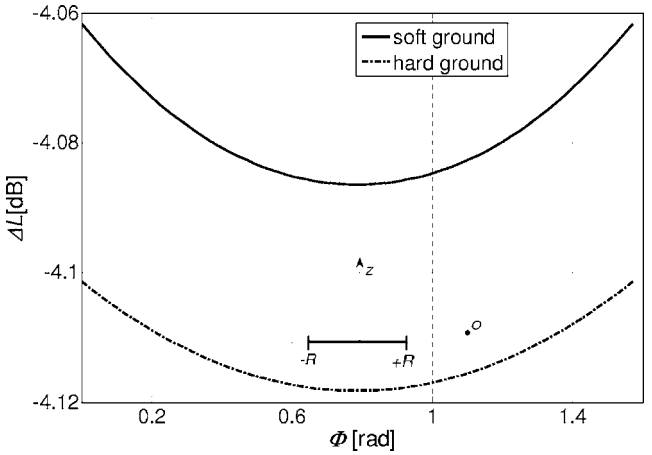


FIG. 11. Noise reduction ΔL [Eqs. (61) and (62)] for the roundabout traffic speed V_R=30 (km/h), the soft (ς=944) and hard (ς=36 360) ground, the receiver $O(r,H,\phi)$ far away from the roundabout, r=10R, and close to the ground surface, H=0.1R.

V. RESULTS

Equations (6), (40), (47), and (59) imply the noise reduction caused by replacement of a road intersection by a roundabout,

$$\Delta L = 10 \lg \left\{ \frac{C(0) - C(\varsigma)}{[A(0) - A(\varsigma)] + \sigma(V_R)[B(0) - B(\varsigma)]} \right\}, \quad (61)$$

where the functions $A(\varsigma)$, $B(\varsigma)$, and $C(\varsigma)$ are defined by Eqs. (41), (48), and (60), and

$$\sigma = 10^{0.1(L_a - L_{sd})} \left(\frac{V_o}{V_R}\right)^2. \quad (62)$$

With L_a=82.3 and L_{sd}=46.7 (see Sec. IV) the above equation takes the form

$$\sigma = 3631 \left(\frac{V_o}{V_R}\right)^2, \quad (63)$$

where V_o=1 km/h and V_R (km/h) is the steady speed of the roundabout traffic.

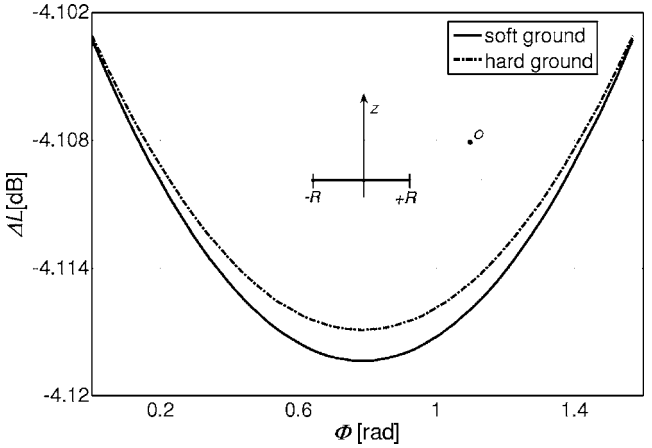


FIG. 12. Noise reduction ΔL [Eqs. (61) and (62)] for the roundabout traffic speed V_R=30 (km/h), the soft (ς=944) and hard (ς=36 360) ground, the receiver $O(r,H,\Phi)$ far away from the roundabout, r=10R, and high above the ground surface, H=R.

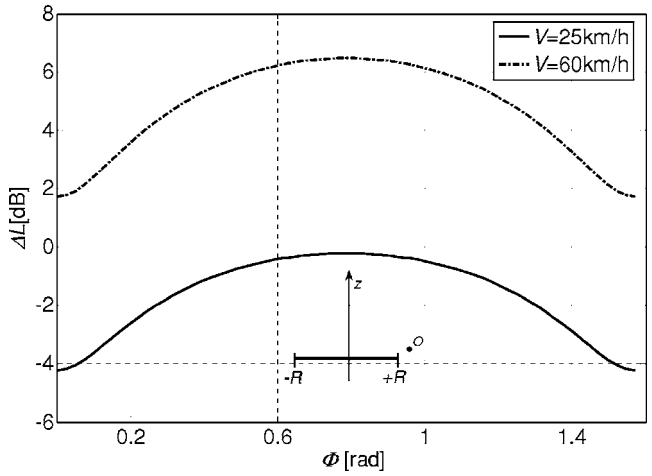


FIG. 13. Noise reduction ΔL [Eqs. (61) and (62)] for the roundabout traffic speed V_R=60 (km/h), the hard ground (ς=36 360), and the critical speed V_*=25 (km/h), at the receiver $O(r,H,\Phi)$ close to the roundabout, r=0.1R, and close to the ground surface, H=0.1R.

Figures 9–12 depict the noise reduction ΔL [Eqs. (61) and (62)] caused by a roundabout of a radius R, when an automobile cruises at a steady speed V_R=30 km/h. The receiver location $O(r,H,\Phi)$ is determined by the angle $0 \leq \Phi \leq \pi/2$, the relative distance r/R, and the relative height H/R. It should be noted that for the receiver location determined by inequalities $0.1 < r/R < 10$ and $0.1 < H/R < 1.0$, the noise reduction ΔL is almost identical for the soft ground (ς=944) and the hard ground (ς=36 360).

The noise reduction becomes real, i.e., $\Delta L < 0$, when the roundabout speed is sufficiently low, $V_R < V_*$. The critical speed V_* depends on the receiver location:

- close to the roundabout (r=0.1R) and close to the ground (H=0.1R), one gets V_*=25 (km/h) (Fig. 13),
- close to the roundabout (r=0.1R) and high above the ground (H=R), one obtains V_*=52 (km/h) (Fig. 14),
- far away from the roundabout (r=10R) and close to the ground (H=0.1R) one arrives at V_*=48 (km/h) (Fig. 15), and, finally,

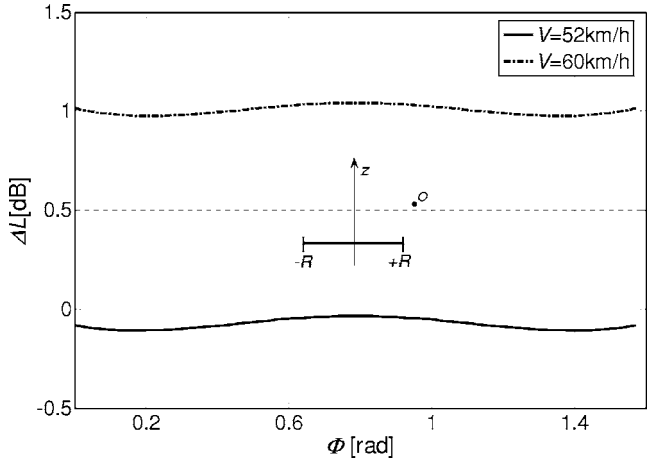


FIG. 14. Noise reduction ΔL [Eqs. (61) and (62)] for the roundabout traffic speed V_R=60 (km/h), the hard ground (ς=36 360), and the critical speed V_*=52 (km/h) at the receiver $O(r,H,\Phi)$ close to the roundabout, r=0.1R, and high above the ground surface, H=R.

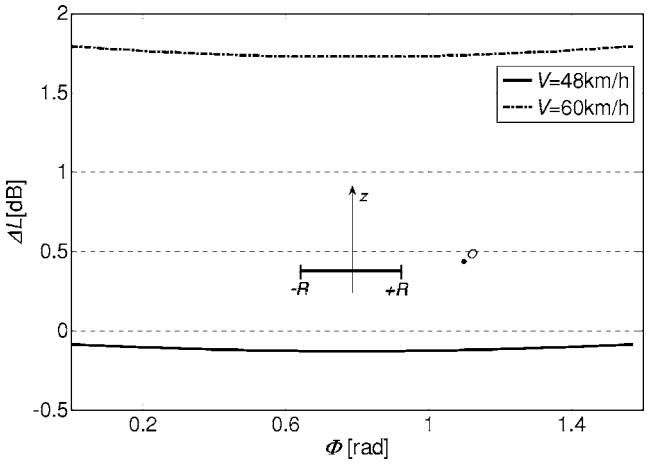


FIG. 15. Noise reduction ΔL [Eqs. (61) and (62)] for the roundabout speed V_R=60 (km/h), the hard ground (ς=36 360), and the critical speed V_*=48 (km/h), at the receiver $O(r,H,\Phi)$ far away from the roundabout, r=10R, and close to the ground surface, H=0.1R.

- far away from the roundabout (r=10R) and high above the ground (H=R) one arrives again at V_*=48 (km/h).

For any location of the receiver ($r > R, H > 0$) the calculation of the critical speed V_* is possible with the use of Eqs. (61) and (62).

VI. CONCLUSIONS

Replacement of a road intersection (Fig. 2) by a roundabout (Fig. 1) gives a noise reduction, ΔL [Eqs. (61) and (62)]. It is real, $\Delta L < 0$, when the roundabout traffic speed V_R is less than the critical speed V_*. The model presented in this study is based on the following assumptions:

- traffic flow is identical in all possible directions (Figs. 3 and 5),
- there is only one category of vehicles, i.e., automobiles,
- each automobile stops at the center of the road intersection (Fig. 2),
- there is a constant deceleration and acceleration rate,
- during deceleration and acceleration, the A-weighted sound power is a cubic function (αV^3) and a linear function (αV), respectively, of the automobile speed [Eq. (20)],
- an automobile is represented by a nondirectional point source, and
- the homogeneous atmosphere is at rest.

Although rather restrictive, these assumptions are realistic and it seems that the model presented here captures salient features of noise generation and propagation. The model presented here would become more precise by refined consideration of the above assumption. For example, by taking into consideration the waves reflected from the building walls (Ref. 12).

ACKNOWLEDGMENTS

The authors are grateful to two anonymous reviewers for their insightful corrections and constructive comments.

[1]P. T. Lewis and A. James, "Noise levels in the vicinity of traffic roundabouts," J. Sound Vib. **72**, 51–69 (1980).
[2]W. M. To and T. M. Chan, "The noise emitted from vehicles at roundabouts," J. Acoust. Soc. Am. **107**, 2760–2763 (2000).
[3]M. Berengier, "Acoustical impact of traffic flowing equipments in urban area," Forum Acusticum 2002, September 16–20, 2002, Sevilla.
[4]D. Bertoni, "Experience in planning of noise mitigation measures in evaluating their effectiveness in the city of Modena," Inter-Noise 2004, August 22–25, Prague.
[5]J. Picaut, M. Berengier, and E. Pousseau, "Noise impact modeling of a roundabout," Inter-Noise 2005, August 7–10, Rio de Janerio.
[6]S. Kono, Y. Oshino, T. Iwase, T. Sone, and H. Tachibana, "Road traffic noise prediction model ASJ RTN-Model 2003 proposed by the Acoustical Society of Japan, Part 2," ICA 2004, April 4–9, Kyoto.
[7]B. M. Favre, "Noise emission of road vehicles: Evaluation of simple model," J. Sound Vib. **91**, 571–582 (1983).
[8]R. Makarewicz and P. Kokowski, "Simplified model of ground effect," J. Acoust. Soc. Am. **101**, 372–376 (1997).
[9]K. Attenborough, T. Waters-Fuller, K. M. Li, and J. A. Lines, "Acoustical properties of farmland," J. Agric. Eng. Res. **76**, 183–195 (2000).
[10]K. Attenborough, "A comparison of engineering methods for predicting ground effect," Forum Acusticum, Berlin, March 14–19, 1999.
[11]Y. Oshino, K. Tsukui, C. Roovers, G. Blokland, and H. Tachibana, "Possibility of international standardization of road traffic noise prediction model," Inter-Noise 2004, August 22–25, Prague.
[12]J. Kang, "Numerical modeling of the sound fields in urban squares," J. Acoust. Soc. Am. **117**, 3695–3606 (2005).

Enhancing low frequency sound transmission measurements using a synthesis method

Teresa Bravo[a] and Cédric Maury
Université de Technologie de Compiègne, Laboratoire Roberval FRE-CNRS 2833, Secteur Acoustique, BP 20529, F-60205 Compiègne Cedex, France

(Received 23 August 2006; revised 4 June 2007; accepted 11 June 2007)

The characterization of low frequency sound transmission between two rooms via a flexible panel is investigated experimentally in this work. Previously, the individual effects of the transmission suite on the measured sound reduction index have been studied analytically, and the results have been compared with the ideal case of having free field radiation conditions on both sides of the panel. A new approach is proposed using a near-field array of loudspeakers driven by a set of optimized signals such that a diffuse pressure field is reproduced on the surface of the partition to be tested. The practical effectiveness of this method is assessed when using a set of 16 acoustic sources located in the source reverberant room in close proximity to an aluminium panel. The experimental results obtained confirm the dependence of the characterized sound reduction index on the particular test chamber considered in the low frequency range. They also validate the proposed synthesis method for providing an estimate that only depends on the properties of the partition itself. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2754062]

PACS number(s): 43.55.Rg, 43.55.Br, 43.60.Fg, 43.60.Cg [MCH] Pages: 869–880

I. INTRODUCTION

One important topic in room acoustics concerns the characterization of the degree of diffuseness in finite sized rooms.[1] When dealing with reverberant test chambers, a limited number of modes exists in the low frequency range that influences the room spectral response and imposes a limit in the applicability of the diffuse field theory. There have been many investigations on the interaction between sources, receivers and room modes at low frequency, focused on specific applications. Several authors[2,3] have conducted research on the suitability of diffusing surfaces to scatter sound while suppressing specular reflections. For instance, in music performance spaces, diffusion may enhance the sound perceived by the listeners. But the diffusors can also be used in rooms for music reproduction, such as recording studios, to provide a more neutral listening environment.[4] In particular, Cox *et al.*[5] have presented an optimization method for the room size and layout of the loudspeakers and listeners to minimize the coloration effects of low-frequency modes.

The questions of how to improve and characterize sound-field diffuseness have been widely studied regarding measurements and tests carried out in reverberant rooms. Of particular importance is the problem of determining the low frequency sound reduction index of partitions. In several industries, high performance acoustic insulating lightweight materials are often demanded for providing acoustic insulation or for protecting people against extraneous airborne noise. The prediction of the sound insulation properties of structures, as described in the International Standard ISO 140-3,[6] is based on the diffuse field theory, where the incidence of the sound energy is equally probable in all directions. However, in practice, panels are normally tested in finite-sized reverberant transmission suites, where an acoustic field is created by several loudspeakers in the source room and transmitted to the receiving room via the test panel. At low frequency, a problem appears when the sound field in the measurement rooms is not diffuse.

Several inter-laboratory comparisons[7–10] carried out in different facilities have demonstrated that the sound insulation at low frequencies can experience significant variations. The current normatives[6] recognize that the diffuse-field theory cannot be applied in the low frequency bands, especially when the room volumes are equal to or less than 50 m^3. An informative annex *F* has been included to ISO 140-3 proposing some guidance for improving the measurements in the low frequency range. The recommendations concern the spatial sampling of the sound field, the number of sources and the averaging time. It is also recommended to increase the minimum separation distances between the loudspeakers and the microphones, and between the microphones and the surfaces of the test rooms. However, in practice it is sometimes difficult to follow these guidelines, especially for small transmission suites.

As alternatives to the above measurement recommendations, a number of papers have discussed results obtained with different testing methods. These include the use of absorbing materials covering the walls facing the partition in both rooms[11] to make the pressure field more uniform and to improve the modal overlapping, or the use of sound intensity measurements in the receiving chamber to minimize its modal influence.[12–14] Kropp *et al.*[15] have evaluated the main parameters affecting the estimation of the reduction index at low frequency, and have concluded that a change of the measurement procedure will hardly solve the problem. It has been found that the limiting factor for the prediction of the transmission loss is not the measurement accuracy, but the

[a]Author to whom correspondence should be addressed. Electronic mail: teresa.bravo-maria@utc.fr

description of the sound insulation itself which is only valid for the specific case under consideration. They suggest the measurement be performed in identical source and receiving rooms as the results would represent the worst case.

In a previous paper, Bravo and Elliott[16] have presented a theoretical investigation on the variability of low frequency sound transmission measurements using a modal approach. An analysis was performed when the panel was mounted in an infinite baffle with free field radiation conditions on both sides, and when the effects of the source and receiving rooms were each taken into account using a fully coupled model. The results obtained on the sound reduction index showed important discrepancies in the low frequency range, thus depending on the specific room parameters. Instead of employing absorbing materials for increasing the diffuseness of the pressure field, or evaluating the incident power on a greater number of points over the whole volume, the authors proposed to use a near-field array of loudspeakers driven by a set of optimized signals such that a diffuse pressure field is reproduced on the surface of the partition to be tested. Preliminary predictions were carried out from a set of frequency response functions (FRFs) measured between loudspeakers and microphones located in a sound transmission suite. It was shown that in the case where the panel is connected to a source room and radiates into an anechoic chamber, the loudspeakers' array is able to compensate for the modal influence of the source room in the low frequency range.[17]

In parallel, the feasibility of synthesizing a number of random pressure fields with given spatial correlation characteristics has been studied theoretically,[18] considering a one-dimensional array of acoustic sources radiating in free field. An experimental setup was designed to reproduce the statistics of two-dimensional random pressure fields using a near-field array of 4×4 loudspeakers located in a semianechoic chamber. The physical limitation performances were assessed for the experimental reproduction of an acoustic diffuse field, a grazing plane wave and a turbulent boundary layer pressure field.[19,20]

Using the array of 4×4 loudspeakers, the previous study on the reproduction of random pressure fields in semianechoic conditions has been extended to a much more reactive environment, a reverberant room, and an experimental investigation is reported in this paper to verify the predictions of the previous theoretical analysis[16] in a real transmission suite. More specifically, this study discusses results on the experimental reproduction of a diffuse pressure field over the surface of a test panel mounted in a sound transmission suite with a small reverberant chamber as a source room, and a large anechoic chamber as the receiving room, and the estimation of the material sound insulation characteristics with the new methodology. The objective is to validate the proposed simulation procedure so that the results obtained at low frequencies on the sound insulation are not restricted to the specific measurement conditions. Of special interest is the determination of the frequency range over which the array of loudspeakers is able to accurately reproduce the diffuse pressure field, as this frequency range must extend up to the source room Schroeder frequency, below which the diffuse field theory is not applicable.

The rest of the paper is organized as follows. The first part describes the determination of the panel sound insulation properties and introduces the real transmission suite. A brief review of both the classical and the new methodology for the determination of the sound reduction index of flexible partitions in terms of the modal characteristics of the uncoupled subsystems is presented. In the second part experimental results are presented comparing the estimated sound reduction index in accordance with the normative and using the proposed synthesis method. The results obtained show that the new method provides an improved estimate of the insulation characteristics of the test panel, ensuring the required reproducibility and repeatability.

II. SOUND REDUCTION INDEX IN A REAL TRANSMISSION SUITE

When a sound field impinges upon the surface of a building element, part of it is reflected back, part propagates to other connecting elements (flanking transmission), part dissipates as heat within the material and part transmits through the element. The sound reduction index is defined in terms of the fraction of incident energy transmitted through the test element. The major problem of providing a reliable estimate of this quantity at low frequencies in a real transmission facility has been the subject of several studies with different approaches. The objectives of this section are twofold: first to describe and characterize a laboratory setup used to determine the sound insulation characteristics of a test partition, and second to present the main differences for its estimation between the normative procedure and the use of a synthesis technique based on the generation of a diffuse pressure field in the source room with an array of suitably driven loudspeakers close to the test panel. Analytical predictions are performed to outline the limitations due to the classical method and to show that the new approach can provide a solution for sound insulation measurements that only depends on the properties of the partition.

A. Characterization of the experimental setup

The measurements are carried out in the transmission laboratory of the Department of Mechanical Engineering at the Université de Technologie de Compiègne. This facility is used to undertake work research in building acoustics, and occasionally commercial consultancy. A schematic representation of the transmission suite is presented in Fig. 1. The experimental setup consists of a box-shaped reverberant room with dimensions $2.72\ \text{m}\times5.31\ \text{m}\times2.96\ \text{m}$ and wide band reverberation time $T_{60}=2.16$ s. The walls are covered with a hard gloss paint to increase the reverberation time and improve the diffuseness of the pressure field. This room is connected to a large anechoic chamber, with dimensions $7.10\ \text{m}\times6.37\ \text{m}\times2.65\ \text{m}$ and cutoff frequency of 80 Hz, through an opening in which the partition is mounted. The building element to be tested is an aluminium panel of length and width $0.97\ \text{m}\times0.67\ \text{m}$, respectively, and with a thickness of 0.003 m. It is centrally positioned between the two rooms. All the joints between the panel and the cavities have been sealed to prevent acoustic leakage.

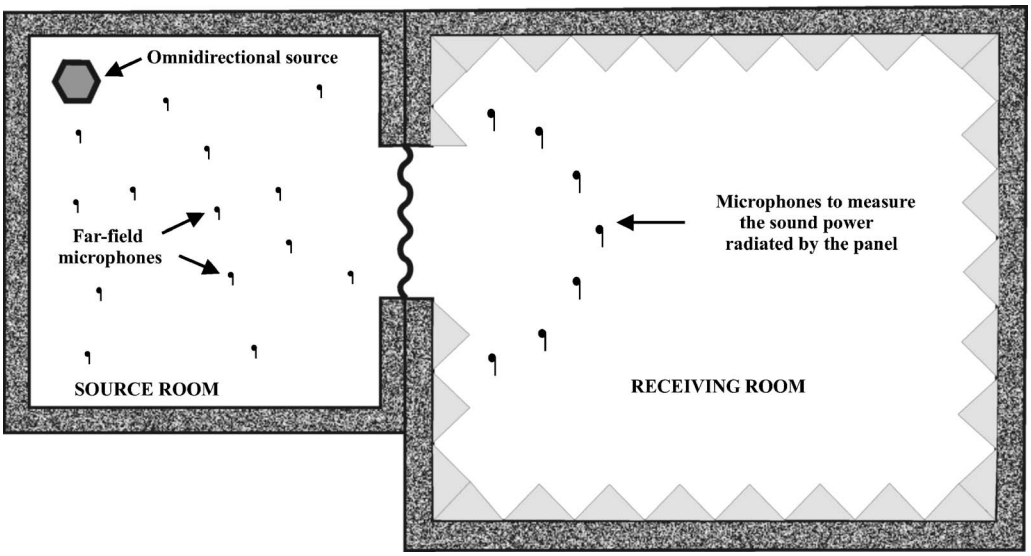


FIG. 1. Geometric arrangement of the source and receiving room connected through a test partition for sound transmission loss measurement.

Prior to the experimental verification, an analytical model initially developed[16,17] for understanding the problems involved at low frequency in the determination of the sound transmission properties of test panels has been adapted to the parameters of the current laboratory setup. It can provide useful prediction results on the improvements and the physical limitations that can be expected from the use of the new measurement method for the case under consideration. This model has been formulated in terms of the modal characteristics of the source, receiving room and partition subsystems and it constitutes an effective approach for estimating the behavior of low frequency coupled systems. More specifically, the complex sound pressure at a point $\mathbf{M}$ in an enclosure of volume V, and the structural velocity at a point $\mathbf{P}$ on a panel of surface S_p can be expressed at frequency ω, respectively, as a sum of modal terms of the form[21]

$$p(\mathbf{M};\omega) = \sum_{\ell mn=0}^{\infty} a_{\ell mn}^{(a)}(\omega)\psi_{\ell mn}(\mathbf{M}), \tag{1}$$

$$v(\mathbf{P};\omega) = \sum_{qr=1}^{\infty} a_{qr}^{(p)}(\omega)\phi_{qr}(\mathbf{P}), \tag{2}$$

where $a_{\ell mn}^{(a)}(\omega)$ is the complex amplitude of the ℓmn acoustic pressure mode defined by the mode shape function $\psi_{\ell mn}(\mathbf{M})$ and the natural frequency $\omega_{\ell mn}$, and $a_{qr}^{(p)}(\omega)$ is the complex amplitude of the qr structural mode defined by the mode shape function $\phi_{qr}(\mathbf{P})$ and the natural frequency ω_{qr}.

In order to obtain reliable predictions, it is important to make an adjustment of the modal parameters introduced into the vibro-acoustic model with those observed from the measurements. A modal analysis is carried out for both the panel and the cavity. For the partition, a single input—single output testing is performed *in situ* using an electromagnetic shaker attached *via* a stiff drive rod to the source room side of the panel at a corner point, which is not likely to coincide with many structural modes. It is equipped with an impedance head to measure both force and acceleration at the excitation point. A small-sized accelerometer is attached on the receiving room side of the panel to measure mobility frequency response functions (FRFs). The number of nodes employed (12×8 evenly spaced points, respectively, along the length and the width of the panel) has been determined taking into account the frequency range of interest for the experiment, i.e., below 1 kHz. Figure 2 shows a comparison of the experimental kinetic energy of the panel estimated from the output of the accelerometer positions, and calculated with the modal model in which the panel is assumed to have a Young's modulus $E=7.110^{10}$ Pa, a density $\mu=2650$ kg/m^3 and a Poisson's ratio $\nu=0.35$. As the mechanical damping of a test panel is not only dependent upon the internal damping of the panel but also upon the mounting conditions on the wall, the modal coefficients $a_{qr}^{(p)}(\omega)$ are calculated assuming either a simply supported panel or a clamped panel. They are extracted by circle-fitting the Nyquist plots of the FRFs in the vicinity of each resonance. It can be appreciated from Fig. 2 that, for this particular configuration, the model with simply supported boundary conditions provides a better agreement with the experimental results.

A modal analysis has also been carried out to correlate with measured data the formulation for the source room-panel system with the panel characterized by the above updated model. A sound field is generated by a high power dodecahedron loudspeaker located in one corner of the reverberant room and the potential energy in the cavity is estimated from the sound pressure measured at 14 microphone positions randomly distributed throughout the total volume. An initial estimate of the room modal damping coefficients is given by the 3 dB bandwidth $B_{\ell mn}$ of the ℓmn acoustic mode. It is deduced from measurements of the room reverberation time $T_{60,\Delta\omega}$ in a set of frequency bands $\Delta\omega$ such that $B_{\ell mn}\approx 6/(2T_{60,\Delta\omega}\log_{10}(e))$, where e is the constant of Neper. This relation has been obtained from considerations about the decay of the transient energy in the room. The modal damping factors are then further adjusted by circle fitting the Nyquist plots of the FRFs between the microphones and the

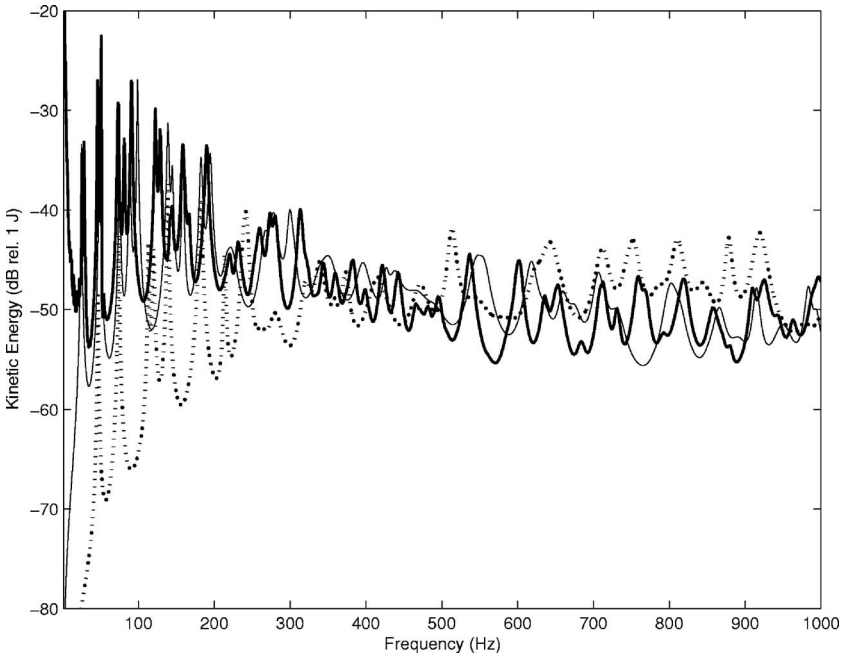


FIG. 2. Kinetic energy of the test panel: measured (bold) and predicted with two boundary conditions (clamped: dotted; simply supported: thin).

loudspeaker in the vicinity of each distinguishable resonance, separated in frequency by at most the full 3 dB bandwidth of the resonance, $2B_{\ell mn}$, i.e., below $f_b \approx 269$ Hz where $f_b = c^{3/2}(T_{60}\log_{10}(e)/12V)^{1/2}$, with c the sound speed in air. It can be seen from Fig. 3 that the analytical model for the source room-panel configuration is representative of the measured potential energy over the frequency range of interest, i.e., well below the room Schroeder frequency $f_s \approx 450$ Hz where $f_s = c^{3/2}(T_{60}\log_{10}(e)/4V)^{1/2}$. Once the source room and the panel subsystems have been characterized and their parameters accordingly adjusted in the model, the two uncoupled subsystems can be coupled via modal theory.[16] The analytical formulation is used to predict the sound insulation properties of the test panel with either the classical or the new methodology.

B. Determination of the sound insulation properties

1. Classical approach

The sound insulation properties of a test partition are usually characterized in terms of a sound reduction index, R, expressed by[22]

$$R(\omega) = 10\log_{10}\left(\frac{\Pi_{\text{inc}}(\omega)}{\Pi_{\text{rad}}(\omega)}\right)\ (\text{dB}), \tag{3}$$

where $\Pi_{\text{inc}}(\omega)$ and $\Pi_{\text{rad}}(\omega)$ are the sound power incident and radiated by the partition respectively, at frequency ω.

The determination of these quantities depends on the particular arrangement considered. Measurements carried out using the traditional ISO 140-3 method suppose the use of two adjacent chambers connected though an aperture where

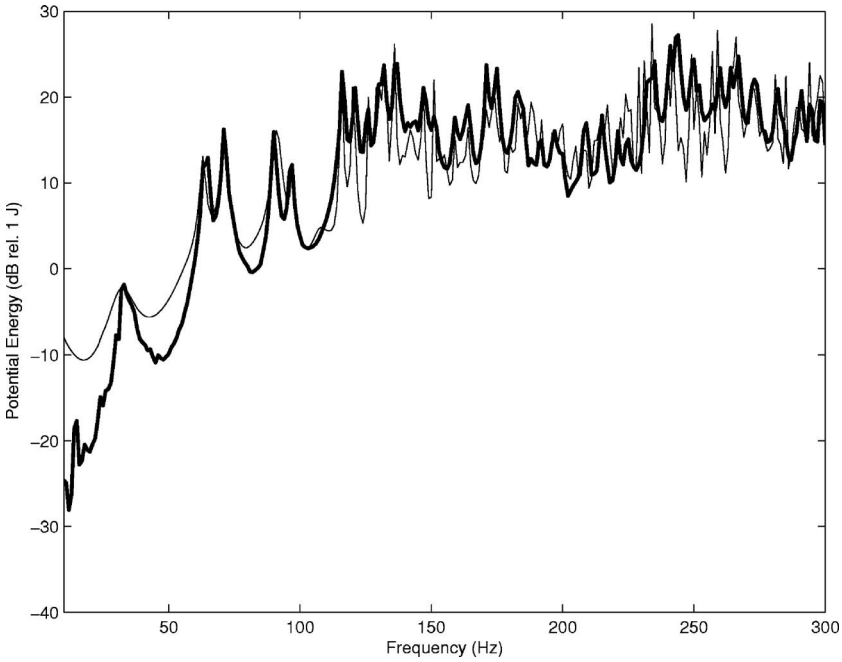


FIG. 3. Potential energy of the source room: measured (bold) and predicted (thin).

FIG. 4. (Color online) Photograph of the acoustic source and the test partition for the classical methodology.

FIG. 5. (Color online) Photograph of the near-field array of loudspeakers for the synthesis of an acoustic diffuse field on the source room side of the test partition for the new methodology.

the test panel is located. The procedure assumes diffuse sound fields in both rooms. As quoted in the introduction, recommendations are specified in terms of a minimum distance, equal to a quarter of a wavelength at the lowest frequency of interest, between the microphones and the surface of the test rooms and between different loudspeaker or microphone positions. Such limitations are difficult to follow for typical reverberant chambers with a volume less than 50 m,[3] as is the case in the current study, or only authorize a very limited number of transducer positions. Several authors[23] have suggested the use of sound intensity measurements in the receiving room to determine the sound power transmitted by the panel. The receiving environment should then be very quiet, as is the case for an anechoic room. Using this methodology it is possible to neglect the influence of the normal modes of the receiving room on the sound transmission measurements, but the problem of the lack of diffuseness of the sound field in the source room at low frequency still remains.

To highlight the sensitivity of the classical method to variations in the source room conditions, a model has been developed to calculate the sound reduction index on the basis of the modal theory.[16] The system consists of a test panel excited by a number of loudspeakers in the source room and radiating in free field, which is an approximation to the experimental setup when the receiving room is replaced by an anechoic chamber. The subsystems are coupled since the pressure field in the source room is influenced by the panel motion and the panel structural response is excited by the pressure field in the source room. The radiated power has been calculated in terms of the velocities of an array of elemental radiators on the surface of the panel.

The prediction model is applied to the transmission lab assuming that the source room is excited by an omnidirectional loudspeaker to create an isotropic sound field. Figure 4 shows a picture of both the speaker and the panel mounted on the source room side wall. In practice, the acoustic source is located on the back wall corner of the room at a distance of 30 cm from the wall's surfaces so that it well couples into a large number of low frequency normal modes. A set of 14 microphone positions are randomly distributed throughout the cavity to provide an estimate of the potential energy using the average sound pressure level. In the receiving room, the sound power radiated by the panel is determined by the enveloping surface method[24] from the pressure field measured at ten microphone positions located over a hemispherical surface of radius 1 m, surrounding the partition. Using this configuration, a prediction of the partition sound reduction index is obtained according to the normative in one-third-octave frequency bands. In the following subsection, these results are compared with those that can be achieved with the synthesis method.

2. Synthesis approach

A novel approach based on the generation of an acoustic diffuse pressure field by an array of loudspeakers suitably driven and situated in the near field of the object to be tested is assessed for the lab transmission suite. The complete system consists of a set of 4×4 loudspeakers of 0.21 m diameter, mounted in a wooden enclosure with dimensions 1 m $\times 1$ m $\times 0.2$ m, and centered in front of the test panel. The source's drive signals are optimized so that a spatially correlated random pressure field is generated with a spatial correlation function which best fits the one due to an acoustic diffuse pressure field at a number of microphone positions situated in close proximity to the panel. The microphones are 6 mm miniature electret capsules mounted over one stainless steel rod, evenly spaced along the width of the panel. During the experimental verification, the bar is sequentially displaced along the panel length so that a total number of 7×9 microphone positions are uniformly distributed over a grid distant 1 cm apart from the partition. Figure 5 shows a photograph of the experimental setup, with the loudspeaker's box situated in front of the panel, and the microphone's bar at a distance of 1 cm apart from the partition. The grid of microphones represents the surface over which an ideal dif-

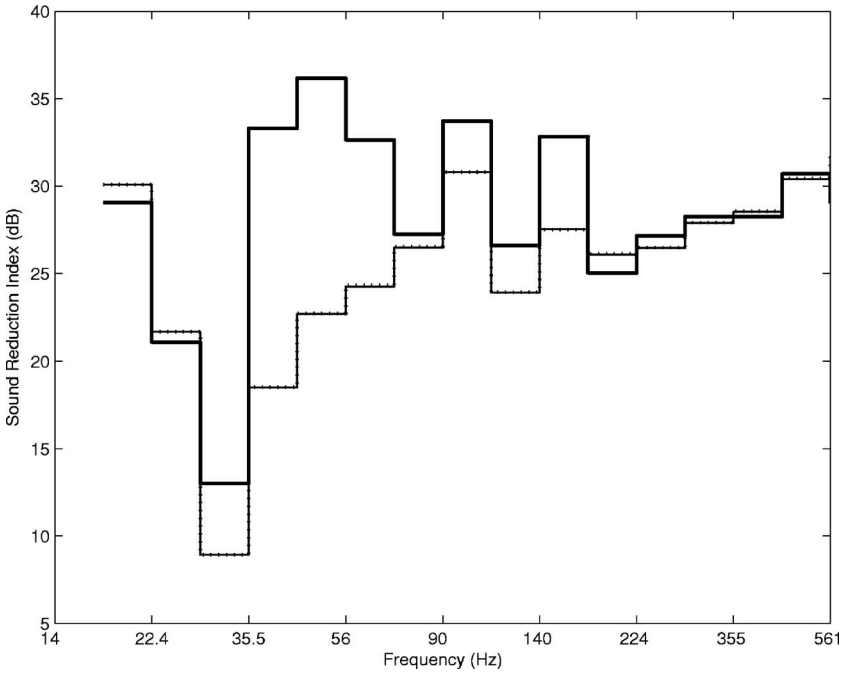


FIG. 6. Prediction of the sound reduction index in third-octave bands for the test partition mounted in an infinite baffle under diffuse field excitation conditions (thin), and in the source room-panel configuration obtained with the near-field optimized loudspeakers (dotted) and with a far-field acoustic source such as in the classical methodology (bold).

fuse field is reproduced. The size, the number and the location of the transducers have been considered in previous works in order to achieve an accurate reproduction of the diffuse field correlation function with enough spatial resolution within a semianechoic room.[19,20] Further work has assessed how the simulation accuracy is modified if the synthesis is achieved in a much more reactive environment like in a reverberant chamber.[25] In particular, the distance between the loudspeaker's front face and the microphones is taken as the separation distance between two adjacent loudspeakers in the array in order to lower the conditioning of the plant matrix to be equalized between the loudspeakers and the microphones [see Eq. (6)]. Of special interest is to determine if the frequency range for an accurate reproduction of the diffuse pressure field extends up to the source room Schroeder frequency, where the diffuse field assumption is not yet valid.

The new method for the determination of the sound reduction index of the test partition differs considerably from the classical method. The incident power in the source room is obtained from pressures generated at the grid microphone positions close to the test object. During the synthesis process, the microphone's output signals provide a measure of the spatial variation of the pressure field, which ideally corresponds to that of a diffuse field, **d**. This can be assumed to be derived from a set of uncorrelated white unit variance reference signals, **x**, *via* a matrix of shaping filters **D**, calculated from an eigen-factorization of the cross-spectral density (CSD) matrix between the elements of the diffuse field pressures at the microphones

$$\mathbf{S}_{dd} = E[\mathbf{d}\mathbf{d}^H], \tag{4}$$

where $\mathbf{d}=\mathbf{D}\mathbf{x}$, E denotes the expectation operator and H denotes the Hermitian, complex conjugate transpose. The explicit dependence of the variables on frequency, ω, has been dropped for notational convenience. The diagonal terms of $\mathbf{S}_{dd}$ correspond to the power spectral densities of each of the microphones, which are the same in this case, and the off diagonal elements correspond to the CSDs, which for two microphones A and B are assumed to be of the form[18]

$$S_{AB}(r;\omega) = S_{AA}(\omega)\frac{\sin kr}{kr}, \tag{5}$$

where $S_{AA}(\omega)$ is the power spectral density at a single microphone, k is the acoustic wave number, ω/c, and r is the distance between the microphones A and B.

A matrix of control filters, **W**, is designed to determine the optimum input signals to an array of loudspeakers, which drive the microphone outputs to be as close as possible to those due to a diffuse field, **d**. Using a least-squares optimization algorithm, the optimal matrix of filters is given by[18]

$$\mathbf{W}_{\text{opt}} = [\mathbf{G}^H\mathbf{G}]^{-1}\mathbf{G}^H\mathbf{D} = \mathbf{G}^{\dagger}\mathbf{D}, \tag{6}$$

where **G** is the matrix of acoustic responses between the near field loudspeakers and the near field microphones and $\mathbf{G}^{\dagger}$ is the pseudo-inverse of **G**.

Using the analytical formulation presented in Sec. II A for the source room-panel system, the sound reduction index is calculated from the proposed synthesis method, and compared with the one obtained using the normative recommendations. The results shown in Fig. 6 correspond to three different measurement conditions. First, an ideal diffuse field is assumed, incident on the partition mounted in an infinitely baffle, and with free field radiation conditions in both the source and the receiving sides. This estimation has been taken as the reference case, as it only contains the characteristic response of the panel. For the classical method, the omnidirectional source is driven by a random noise and the incident power is calculated using an approximation to the potential energy in the source room over the far-field microphones. As it can be seen at low frequency, the modal influence of the small room is clearly visible, with differences in the estimated sound reduction index of the panel greater than

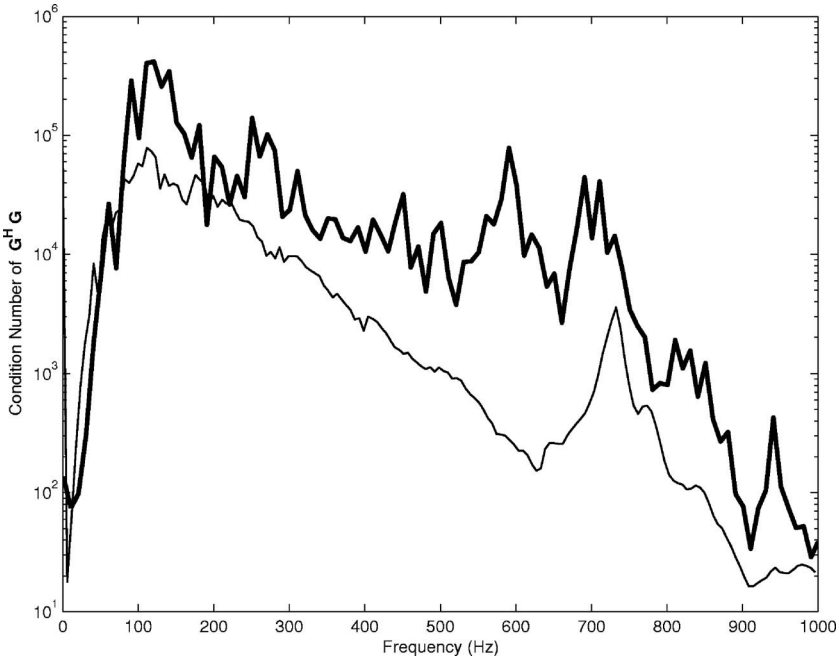


FIG. 7. Condition number associated with the plant response matrix measured between a near-field array of 4×4 loudspeakers and a grid of 7×9 microphones covering the surface of a test panel and located in a reverberant (bold) or in a semianechoic (thin) acoustic environment.

15 dB. Using the synthesis approach, the 16 near-field loudspeakers are driven by optimal signals for the reproduction of a diffuse pressure field, and the power incident is proportional to the mean square sound pressure averaged on the positions of the microphones' grid over the partition. From these prediction results it is remarkable that the sound reduction index obtained when using the near-field array of loudspeakers is almost identical to the reference result in the required frequency range. Above the Schroeder frequency the three results converge towards the same value, as expected.

C. Correlation results from the measured plant matrices

In order to implement the synthesis technique, a set of frequency response functions (FRFs) is required between the near-field loudspeakers and the microphones to calculate the optimal matrix of control filters, $\mathbf{W}_{\text{opt}}$, that generate the loudspeakers' driving signals. The FRFs are measured between each of the 16 near-field loudspeakers and each of the 7×9 microphones. A special attention is paid to the invertibility of the physical plant matrix between the loudspeakers and the microphones, accounting for the modal structure of the sound field in the reverberant room. Apart from the number of sources required per unit acoustic wavelength, the plant matrix invertibility is another factor limiting the accuracy of the simulation. It is characterized by the condition number of the matrix $\mathbf{G}^H\mathbf{G}$ to be inverted [see Eq. (6)]. For a given source/sensor geometry, the condition number is depicted in Fig. 7 in either a reverberant or a semianechoic acoustic environment for comparison. It can be seen that the conditioning of the plant matrix becomes poorer in the reverberant environment since the condition number is much larger (by up to a factor 10^3) than that in the semianechoic case. The highest condition number appears in both cases at low frequencies below 200–300 Hz, decreasing when frequency increases, although not uniformly. In particular, it increases significantly at frequencies corresponding to the lightly damped resonances of the reverberant chamber, which could then be equalized with difficulty by the optimum controller.

For a successful reconstruction of an ideal diffuse field at the microphones' positions, it is preferable to keep the plant matrix condition number as low as possible. However, there are no general quantitative thresholds that have been expressed on the condition number below which a satisfactory multiple-point reconstruction can be achieved since it depends on the specific source/sensor configurations as well as the acoustic environment under consideration. In the following, it is shown that the least-squares solution that minimizes the reproduction error at the microphones positions, as given by Eq. (6), yields good performances without the need to use regularization methods.

Experimental correlation results are illustrated in Fig. 8 between the pressure measured at a microphone in the middle point M of the right edge along the panel width and the pressure measured at a reference microphone in the center point $M_{\text{ref.}}$ of the panel surface as a function of the nondimensional frequency normalized by the microphones' separation distance L between M and $M_{\text{ref.}}$. The upper limit, $kL=7.7$, corresponds to a frequency of 1 kHz. To plot the normalized spatial correlation, the following relation has been used:

$$C_{yy}(kL) = \frac{\Re[S_{yy}(M,M_{\text{ref.}};\omega)]}{\sqrt{S_{yy}(M,M;\omega)S_{yy}(M_{\text{ref.}},M_{\text{ref.}};\omega)}}, \tag{7}$$

where $S_{yy}(M,M_{\text{ref.}};\omega)$ is the CSD function evaluated at frequency ω between the corresponding microphone outputs when the sound field is generated either by the suitably driven near-field loudspeakers or by the omnidirectional sound source using the classical method. The corresponding CSD functions are, respectively, extracted from the following CSD matrices $\mathbf{S}_{yy}=\mathbf{G}\mathbf{W}_{\text{opt}}\mathbf{W}_{\text{opt.}}^H\mathbf{G}^H$ or $\mathbf{S}_{yy}=\mathbf{G}_{\text{omni.}}\mathbf{G}_{\text{omni.}}^H$, where $\mathbf{G}_{\text{omni.}}$ is the plant matrix between the omnidirectional source and each of the 7×9 microphones.

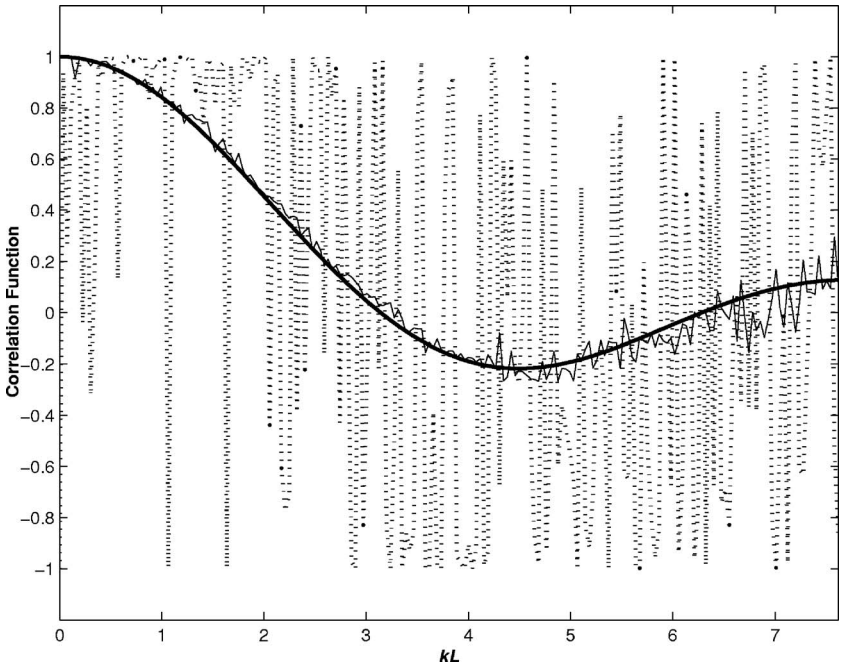


FIG. 8. Spatial correlation function along the panel length when perfect reproduction of an acoustic diffuse field at the microphones' positions is assumed (bold), that achieved in a reverberant room when using either a near-field array of optimized loudspeakers (thin) or an omnidirectional sound source (dotted).

Figure 8 shows the ability of the near-field array of sources to generate at the microphone outputs a random sound field with statistical properties matching those due to an ideal diffuse field up to about $kL=5.6$, and corresponding to an upper frequency of 680 Hz. Above this frequency, noticeable discrepancies start to appear between the approximate and the ideal correlation functions, caused by an insufficiently dense array of sources. This is in accordance with a quantitative criterion[20] according to which two sources per unit acoustic wavelength are required to optimally reproduce the CSD matrix for an acoustic diffuse field. It corresponds to an upper frequency limit of 843 Hz below which an acoustic diffuse field can be approximated over the panel area assuming an array of 4×4 sources.

Using the classical approach, the normalized spatial correlation function varies between +1 and −1. It shows that the sound field generated in the reverberant chamber by a pure-tone omnidirectional source is fully coherent over the panel, in the sense that only one room mode is resonant at the driving frequency and the others are coupled to the excitation through a fixed phase relationship. In this case, the chamber modes do not contribute independently to the cross correlation of the pressure at two microphone positions, especially over the panel. This conclusion has already been drawn from computational[26,27] and experimental[28] approaches, but for pressures away from the room surface walls where the sound field is more likely to be diffuse. However, if several sound sources are simultaneously in operation, averaging over a number of uncorrelated far-field sources positions would generate an equivalent incoherent sound field for which the phase relationships between the room modes would be averaged out, thus eliminating cross-coupling terms between the modes.

For a single source position at one corner of the room, the random nature of the excitation is accounted for if the auto- and cross spectra of the pressures at the microphones are averaged over frequency bands containing a sufficient number of chamber modes which are resonant, i.e., when performing a narrowband analysis of the correlation results, as follows:

$$C_{yy,\Delta\omega}(kL)=\frac{\Re\left[\sum_{\omega=\omega_{\min}}^{\omega_{\max}} S_{yy}(M,M_{\text{ref.}};\omega)\right]}{\sqrt{\sum_{\omega=\omega_{\min}}^{\omega_{\max}} S_{yy}(M,M;\omega)\sum_{\omega=\omega_{\min}}^{\omega_{\max}} S_{yy}(M_{\text{ref.}},M_{\text{ref.}};\omega)}} \tag{8}$$

with $\Delta\omega=\omega_{\max}-\omega_{\min}$. Figure 9 shows how the correlation results are modified when averaged over one-third-octave bands. As expected, below the room Schroeder frequency (450 Hz or $kL=3.7$), the classical approach shows significant deviations from ideal diffuseness, whereas the new approach with the near-field optimized sources already provides an excellent reconstruction of the ideal correlation function for a diffuse field. The performances of both approaches will now be compared but in terms of the real-time measurements of the test panel sound reduction index.

III. EXPERIMENTAL VERIFICATION

Once the physical parameters of the cavity-panel system have been identified to update the model and prediction results have assessed the improvements that can be achieved with the use of a near-field array of transducers, a practical verification of the new methodology is performed in the lab transmission suite and compared with the predicted results. The hardware used for the generation and the acquisition of the time-domain signals at the microphone outputs can be appreciated in Fig. 10. The loudspeakers are driven by a number of partially correlated optimal signals synthesized by a 16-channel arbitrary waveform generator (AWG) programmed by a PC. In the frequency domain, the drive signals to the actuators, $\mathbf{u}_{\text{opt}}$, are obtained from a set of white noise

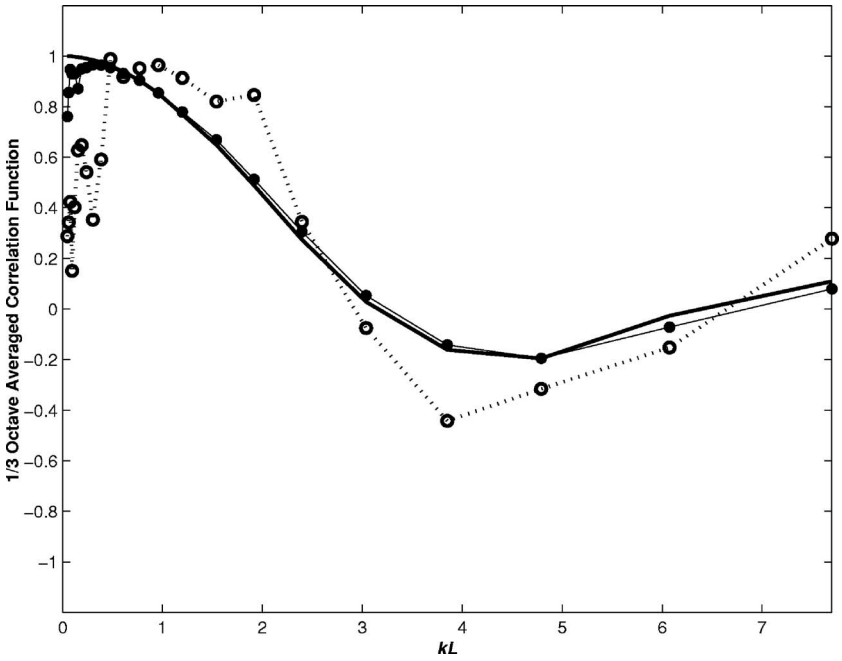


FIG. 9. Third-octave averaged spatial correlation function along the panel length when perfect reproduction of an acoustic diffuse field at the microphones' positions is assumed (bold), that achieved in a reverberant room when using either a near-field array of optimized loudspeakers (thin) or an omnidirectional sound source (dotted).

reference signals $\mathbf{x}$ such that $\mathbf{u}_{opt}=\mathbf{W}_{opt}\mathbf{x}=\mathbf{G}^{\dagger}\mathbf{D}\mathbf{x}$. The corresponding time-domain signals $\tilde{\mathbf{u}}_{opt}$ are obtained by performing on each channel an inverse discrete Fourier transform of $\mathbf{u}_{opt}$. The signals are generated offline, stored in 16 memory cards, and then synchroneously played out during the synthesis process. The offline synthesis of the random pressures does not constrain the control filter $\mathbf{W}_{opt}$ to be causal. The AWG system and its output processor unit are shown in Fig. 11.

In Sec. II C, predictions from the measured FRFs have shown that an acoustic diffuse field could be generated with enough accuracy on the source room side of the panel up to about 680 Hz. Therefore, if the generation bandwidth is setup to 1 kHz, the optimal drive signals can be generated from time records of length $N=12{,}804$ samples with a sampling period of 500 μs over a duration $T_G=6.4$ s. They are generated only from a single reading of the memories in order to avoid cyclic snapshots that may produce discontinuous time histories at the end sections of each block data. The acquisition time of the microphones' output signals is set to a value lower than the duration T_G of the drive signals and is chosen to be $T_A=2.56$ s. The influence of a greater acquisition time provided that the acquisition bandwidth stays above 1 kHz ($T_A \leqslant T_G$) does not significantly influence the results. The repeatability of the new method has also been analyzed to ensure that the simulation is robust to small variations that necessarily occur between the plant matrix initially measured to calculate the drive signals and the one that is equalized during the synthesis process.

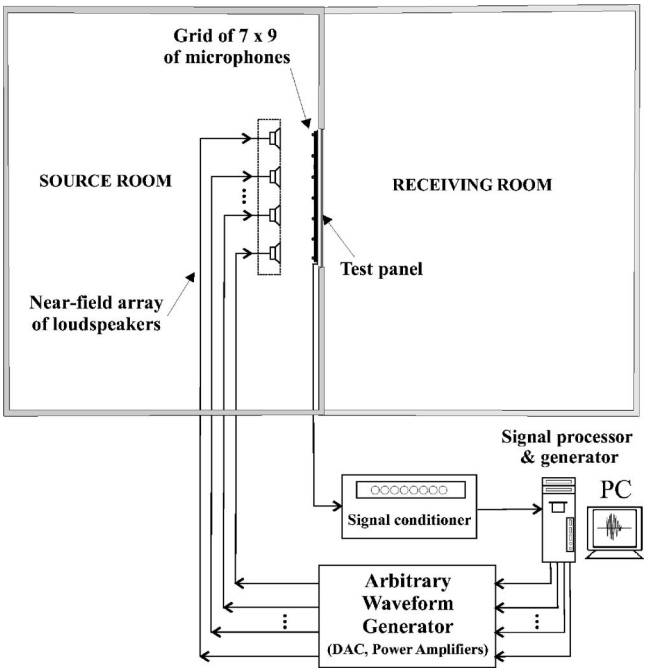


FIG. 10. Experimental setup for the laboratory synthesis of a diffuse sound pressure field.

FIG. 11. (Color online) Photograph of the arbitrary waveform generator system.

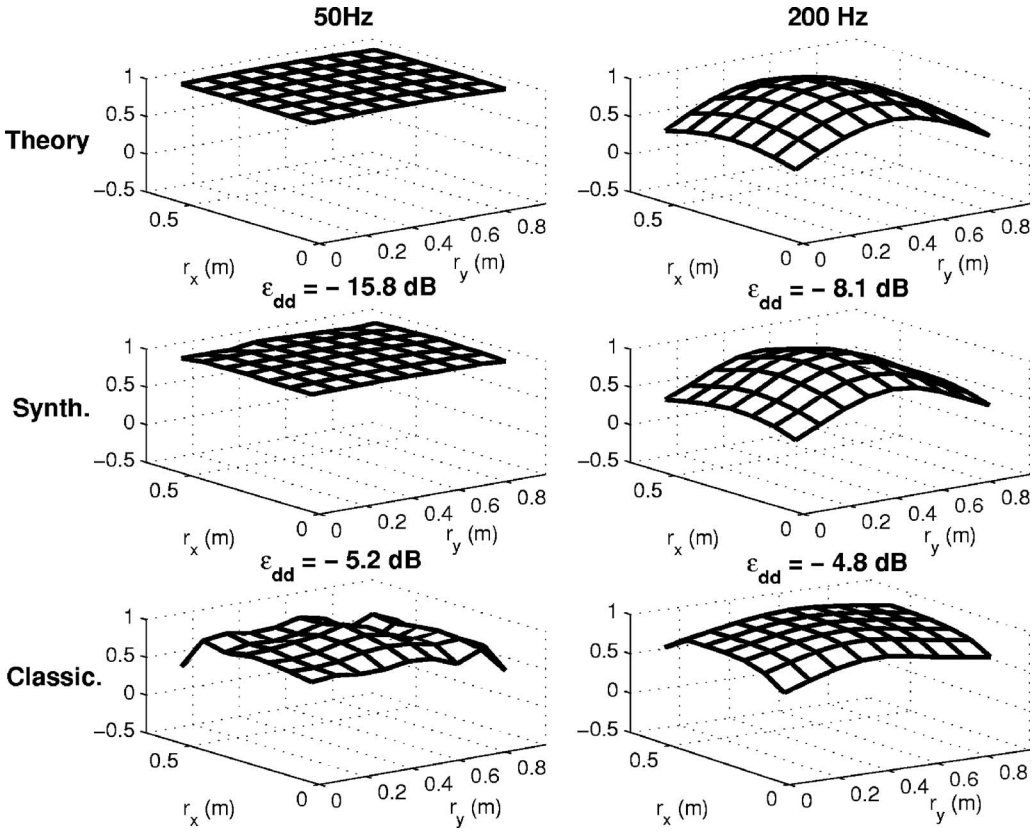


FIG. 12. Normalized amplitudes of the spatial correlation structures obtained at 50 Hz (left column) and 200 Hz (right column) when perfect reproduction of an acoustic diffuse field at 7×9 microphones positions is assumed (top row), that estimated from the optimal time-domain signals at the microphones outputs due to the near-field optimized sources (middle row) and that estimated from the time-domain signals due to the omnidirectional source (bottom row).

Figure 12 compares the spatial correlation patterns estimated from the measured time-domain microphone outputs at frequencies below the source room Schroeder frequency using either the synthesis or the classical approach. The CSDs are estimated using the Welch's method of averaging over short modified periodograms. The accuracy with which the correlation structures are approximated is quantified by a normalized residual error, ε_{dd}, between the assumed and measured correlation functions.[20] It is defined as

$$\varepsilon_{dd} = \frac{\|C_{dd} - C_{yy}\|}{\|C_{dd}\|}, \tag{9}$$

where $\|\cdot\|$ is the Frobenius norm. It is associated to the residual matrix $C_{dd} - C_{yy}$ between the diffuse field correlation matrix C_{dd} and the one estimated from experimental synthesis C_{yy}, defined in Eq. (7). A large reduction in the residual error results in an accurate reconstruction of the assumed correlation structure. In particular, at 50 and 200 Hz, one observes that the reduced set of loudspeakers is clearly able to provide a satisfactory approximation to the correlation function due to an ideal diffuse field, whereas the deviation from the theory is more pronounced as for the statistical properties of the random pressures due to the far-field source.

Figure 13 presents a comparison in third-octave frequency bands between the reference sound reduction index predicted analytically with ideal diffuse field conditions and the experimental measured estimators obtained using the classical and the synthesis methods. At 50 and 200 Hz, the differences between the theoretical sound reduction index and the one measured using the classical method in a small reverberant chamber corresponds, respectively, to 16 and 6 dB. However, the discrepancies with theory are much more reduced when one considers the sound reduction index obtained from the synthesized CSDs which almost reproduce an ideal diffuse field over the panel surface up to 680 Hz. Some differences can still be observed up to the 400 Hz third-octave band, especially in the 160 Hz band that are about 5–6 dB. These errors could be reduced if the acoustic design of the excitation system was improved. During the real-time simulations, it is likely that cross coupling occurs to some extent between the drivers through the low frequency resonances of the enclosure. In particular, the first resonance of the enclosure at 170 Hz is not sufficiently damped by the fiberglass material filling the cabinet, mostly efficient above 800 Hz. An improved design of the excitation system would require isolating each driver from the others using individual uncoupled enclosures.

Above the 15th third-octave band, 450–560 Hz, which contains the Schroeder frequency, there are enough modes in the measurement band for the sound field in the source room to be considered as diffuse and both the normalized and the new method show equivalent performances. Also, we note that the sound reduction index obtained from the measured time-domain signals follows well the trend predicted from the updated analytical formulation and shown in Fig. 6. So the experimental results obtained constitute a validation of

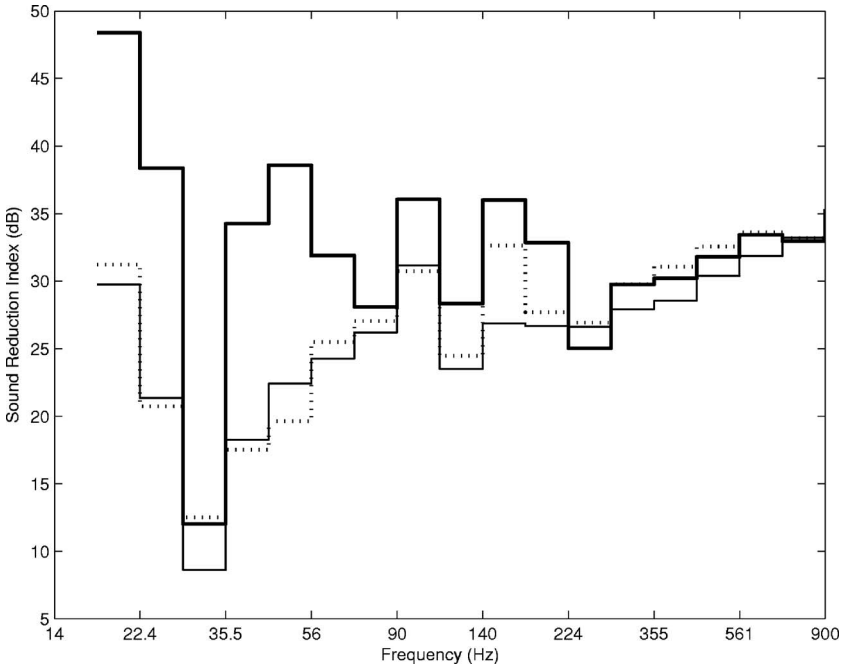


FIG. 13. Sound reduction index in one-third-octave bands for the test partition in an infinite baffle under diffuse field excitation conditions (thin), and measured in the source room-panel configuration from experimental synthesis of a diffuse pressure field (dotted) and using the classical approach (bold).

the new methodology presented, providing an estimator of the sound reduction index that only depends on the properties of the partition itself, avoiding any influence of the transmission suite modal behavior.

IV. CONCLUSIONS

This work has studied the experimental characterization of the sound insulation properties of flexible partitions in the low frequency range. Numerical and experimental evidence has shown that the modal properties of the transmission suite can influence the sound reduction index so that the predictions may not be accurate. A new characterization method has been proposed: it is based on the synthesis of a diffuse pressure field with a near-field array of optimally driven loudspeakers over a grid of microphones close to the test panel.

First, an analytical investigation has been discussed considering a small reverberation room with a total volume of about 43 m^3 connected to an anechoic chamber through the partition to be tested. The dimensions and physical parameters of the model have been adjusted so that they fit those of the real experimental lab. This analytical study enabled to quantify the improvements that can be achieved with the diffuse field synthesis method. The results obtained have predicted that, for this particular case, differences of up to 15 dB can be encountered when the classical and the new method proposed are used.

Subsequently, an experimental verification has been performed in which the loudspeakers are driven by a set of partially correlated optimal signals calculated from the measured transfer functions between the sources and the sensors. The spatial correlation patterns reproduced over the microphones have been plotted at different frequencies, and the authors have shown that in the frequency range of interest, i.e., below the source room Schroeder frequency, the synthesis method is able to compensate for the modal properties of the reverberant room and, thus, provide an estimated sound reduction index very similar to the one obtained under the assumption of a pure incident diffuse pressure field.

The results obtained provide an experimental validation of previous analytical predictions: they could be used to quantify the accuracy of the simulation that can be achieved in different facilities and provide an estimate of the sound reduction index corresponding to ideal diffuse field excitation conditions. In the latter, the only required input data is an experimental modal analysis of the panel.

We note that the measured sound insulation properties are not only valid for the specific lab transmission suite under consideration, but constitute an estimate that only depends on the properties of the material tested. This suggests that, when using the new methodology, the dimensions of the source room could be further reduced, thus avoiding the use of large volumes whose temperature and physical conditions are difficult to control over the whole measurement procedure and which are not exempt from the modal cavity influence in the very low frequency range.

In practice, the new methodology can also be used for the analysis of specimens with larger sizes. In this case, the number of sources per unit acoustic wavelength should be increased according to the theoretical criterion.[20] The drivers should then be uniformly distributed over the reproduction area. Finally, further work could be focused on the experimental validation of the synthesis method to enhance transmission loss measurements in other types of acoustical environments such as in ducts or in anechoic environments.

ACKNOWLEDGMENT

The authors acknowledge financial support for this research from the ANVAR (French Agency for Innovation) under Contract No. SR-04-129.

[1]J. S. Bradley, "Hot topics in architectural acoustic," J. Acoust. Soc. Am. **88**, S66–S67 (1990).

[2]T. J. Cox and P. D'Antonio, "Acoustic phase gratings for reduced specular reflections," Appl. Acoust. **60**, 167–186 (2000).
[3]T. J. Hargreaves, T. J. Cox, Y. W. Lam, and P. D'Antonio, "Surface diffuse coefficients for room acoustics: Free-field measurements," J. Acoust. Soc. Am. **108**, 1710–1720 (2000).
[4]P. D'Antonio and T. J. Cox, "Diffusor application in rooms," Appl. Acoust. **52**, 113–142 (2000).
[5]T. J. Cox, P. D'Antonio, and M. R. Avis, "Room sizing and optimization at low frequencies," J. Audio Eng. Soc. **52**, 640–651 (2004).
[6]BS EN ISO 140–3. Acoustics, Measurement of Sound Insulation in Buildings and of Buildings Elements. Part 3: Laboratory Measurements of Airborne Sound Insulation of Building Elements (1995).
[7]M. Vorländer and A. C. C. Warnock, "Inter-laboratory comparisons of low-frequency sound transmission: I. Conventional and intensity measurements," J. Acoust. Soc. Am. **93**, 2343 (1993).
[8]A. Schmitz, A. Meier, and G. Raabe, "Intercomparison test of sound insulation measurement in test facilities. I.," J. Acoust. Soc. Am. **105**, 1199 (1999).
[9]A. Schmitz, A. Meier, and G. Raabe, "Intercomparison test of sound insulation measurement in test facilities. II.," J. Acoust. Soc. Am. **105**, 1199 (1999).
[10]H. S. Olesen, "Laboratory measurement of sound insulation in the frequency range 50 Hz to 160 Hz-A. Nordic intercomparison," Nordtest Report No. TR 489, (2003). Available from: www.nordtest.org.
[11]J. Roland, "Adaptation of existing test facilities to low frequencies measurements," *Proc. of Internoise 95*, Bolton, Newport Beach, California, 1113–1116 (1995).
[12]M. J. Crocker, P. K. Raju, and B. Forssen, "Measurement of transmission loss of panels by the direct determination of transmitted acoustic intensity," Noise Control Eng. **17**, 6–11 (1981).
[13]D. B. Pedersen, J. Roland, G. Raabe, and W. Maysenhölder, "Measurement of the low-frequency sound insulation of building components," Acust. Acta Acust. **86**, 495–505 (2000).
[14]V. Hongisto, M. Lindgren, and J. Keränen, "Enhancing maximum measurable sound reduction index using sound intensity method and strong receiving room absorption," J. Acoust. Soc. Am. **109**, 254–265 (2001).
[15]W. Kropp, A. Pietrzyk, and T. Kihlman, "On the meaning of the sound reduction index at low frequencies," Acta Acust. **2**, 379–392 (1994).
[16]T. Bravo and S. J. Elliott, "Variability of low frequency sound transmission measurements," J. Acoust. Soc. Am. **115**, 2986–2997 (2004).
[17]T. Bravo and S. J. Elliott, "A simulation analysis of low frequency sound transmission measurements," *Proceedings of the 18th International Congress on Acoustics*, Kyoto, Japan, 4–9 April, 2004 (ISBN: 9901915-6-0).
[18]S. J. Elliott, C. Maury, and P. Gardonio, "The synthesis of spatially correlated random pressure fields," J. Acoust. Soc. Am. **117**, 1186–1201 (2005).
[19]T. Bravo and C. Maury, "The experimental synthesis of random pressure fields: Methodology," J. Acoust. Soc. Am. **120**, 2702–2711 (2006).
[20]C. Maury and T. Bravo, "The experimental synthesis of random pressure fields: Practical feasibility," J. Acoust. Soc. Am. **120**, 2712–2723 (2006).
[21]P. M. Morse, *Vibration and Sound*, 2nd ed. (McGraw–Hill, New York, 1948) (reprinted in 1981 by the Acoustical Society of America).
[22]M. D. Egan, *Concepts in Architectural Acoustic* (McGraw–Hill, New York, 1972).
[23]R. E. Halliwell and A. C. C. Warnock, "Sound transmission loss: Comparison of conventional techniques with sound intensity techniques," J. Acoust. Soc. Am. **77**, 2094–2103 (1985).
[24]ISO 3744. Acoustics, Determination of sound power levels of noise sources using sound pressure—engineering method in an essentially free field over a reflecting plane (1994).
[25]T. Bravo and C. Maury, "The synthesis of a diffuse sound field with a near field array of loudspeakers," *Proceedings of the 13th International Congress on Sound and Vibration*, Vienna, July 2–6, 2006 (ISBN: 3-9501554-5-7, J. Eberhardsteiner, H. A. Mang, and H. Waubke, Editors; Publisher: Vienna University of Technology, Austria), Paper No. 258 in Structured Session SS22: Inverse Problems in Vibro-Acoustics.
[26]W. T. Chu, "Eigenmode analysis of the interference patterns in reverberant sound fields," J. Acoust. Soc. Am. **68**, 184–190 (1980).
[27]H. Nélisse and J. Nicolas, "Characterization of a diffuse field in a reverberant room," J. Acoust. Soc. Am. **101**, 3517–3524 (1997).
[28]W. T. Chu, "Comments on the coherent and incoherent nature of a reverberant sound field," J. Acoust. Soc. Am. **69**, 1710–1715 (1981).

The bag-of-frames approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music

Jean-Julien Aucouturier[a)]
Ikegami Lab, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan

Boris Defreville[b)]
ORELIA, 77300 Fontainebleau, France

François Pachet[c)]
SONY CSL, 6 rue Amyot 75005 Paris, France

(Received 13 April 2006; revised 30 January 2007; accepted 29 May 2007)

The "bag-of-frames" approach (BOF) to audio pattern recognition represents signals as the long-term statistical distribution of their local spectral features. This approach has proved nearly optimal for simulating the auditory perception of natural and human environments (or soundscapes), and is also the most predominent paradigm to extract high-level descriptions from music signals. However, recent studies show that, contrary to its application to soundscape signals, BOF only provides limited performance when applied to polyphonic music signals. This paper proposes to explicitly examine the difference between urban soundscapes and polyphonic music with respect to their modeling with the BOF approach. First, the application of the same measure of acoustic similarity on both soundscape and music data sets confirms that the BOF approach can model soundscapes to near-perfect precision, and exhibits none of the limitations observed in the music data set. Second, the modification of this measure by two custom homogeneity transforms reveals critical differences in the temporal and statistical structure of the typical frame distribution of each type of signal. Such differences may explain the uneven performance of BOF algorithms on soundscapes and music signals, and suggest that their human perception rely on cognitive processes of a different nature. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2750160]

PACS number(s): 43.60.Cg, 43.50.Rq, 43.60.Lq, 43.66.Jh, 43.66.Ba [BSF] Pages: 881–891

I. INTRODUCTION

A. Soundscapes

In 1977, composer R. Murray Schafer coined the term *soundscape* as an auditory equivalence to landscape.[1] He proposed to consider soundscapes as musical compositions, in which the sound sources are musical instruments. Nowadays, the concept of soundscape is used as a methodological and theoretical framework in the field of rural or urban sound quality, notably for the assessment of noise annoyance.[2] Psycho-physic experiments on the perception of soundscapes[3–5] indicate that the cognitive processes of recognition and similarity operate on the basis of the identification of the physical sources. For instance, a given soundscape can be classified as a "park," when specific and localized audio events such as "birds singing," or "children playing" are identified.[6] This also holds for semantic categorization,[7] i.e., the subjective "unpleasantness" of urban soundscapes increases when more mechanical sound sources (e.g., vehicles) are identified than natural sources (e.g., voices or birds). However, recent research[8] shows that people are also capable of more holistic strategies for processing soundscapes, when individual source identification is difficult in the presence of too many noncharacteristic events ("background noise").

There have been various attempts to simulate human perception of soundscapes with computer algorithms, with methodologies that closely resemble the two alternative cognitive strategies mentioned earlier. A majority of contributions[9–14] take the strategy to identify the constituent sound sources individually. The typical implementation describes sound extracts with generic frame-level features, such as MPEG-7 spectral descriptors,[11] and use hidden Markov models[15] to represent their statistical dynamics. Recent research[14] proposes to enhance this typical scheme by learning problem-specific features, adapted to each sound class, with genetic programming.

However, another trend of works[16–18] proposes to directly recognize soundscapes as a whole, without the prior identification of constituent sound sources. In these works, sound-scapes are modeled as the long-term accumulative distribution of frame-based spectral features. This approach has been nicknamed "bag-of-frames" (BOF), in analogy with the "bag-of-words" treatment of text data as a global distribution of word occurrences without preserving their organization in phrases, traditionally used in text classification and

a)Electronic mail: aucouturier@gmail.com
b)Electronic mail: boris.defreville@orelia.fr
c)Electronic mail: pachet@csl.sony.fr

TABLE I. Number of contributions using the bag-of-frames paradigm in past ISMIR symposiums.

Year	BOF papers	Total papers	Percentage
2000	6	26	23
2001	9	36	25
2002	14	58	24
2003	12	50	24
2004	23	104	22
2005	24	114	21
Total	88	388	23

retrieval.[19] The signal is cut into short overlapping frames (typically 50 ms with a 50% overlap), and for each frame, a feature vector is computed. Features usually consist of a generic, all-purpose spectral representation such as mel frequency cepstrum coefficients[15] (MFCC). The physical source of individual sound samples is not explicitly modeled: All feature vectors are fed to a classifier (based, e.g., on Gaussian mixture models[20]) which models the global distributions of the features of signals corresponding to each class (e.g., pedestrian street or park). Global distributions for each class can then be used to compute decision boundaries between classes. A new, unobserved signal is classified by computing its feature vectors, finding the most probable class for each of them, and taking the overall most represented class for the whole signal.

The BOF approach has proved very effective for soundscapes. Ma *et al.*[18] report 91% classification precision on a database of 80 3 s sound extracts from 10 everyday soundscape classes (street, factory, football game, etc.). Notably, such systems seem to perform better than average human accuracy on the same task (35%), which suggests that 3 s audio data provide enough information for pattern recognition, but not for people. Similarly, Peltonen *et al.*[4] report that the average recognition time for human subjects on a list of 34 soundscapes is 20 s. This supports the cognitive strategy of source identification, which typically imposes longer latencies, depending on the temporal density of discriminative sound events.

B. Music

For the analysis of polyphonic music signals also, the BOF approach has led to some success and is by far the most predominant paradigm. Table I shows an enumeration of paper and poster contributions in the ISMIR conference[21] since its creation in 2000. Each year, about a fourth of all papers, and on the whole 88 papers out of a total 388, use the approach. Each contribution typically instantiates the same basic architecture described earlier, only with different algorithm variants and parameters. Although they use the same underlying rationale of modeling global timbre/sound in order to extract high-level descriptions, the spectrum of the targeted descriptions is rather large: genre,[22] mood,[23] singing language[24] to name but a few.

However, contrary to its application to soundscapes, recent research[25–27] on the issue of polyphonic timbre similarity shows that BOF seems to be bounded to moderate performance, most notably:

(1) Glass ceiling: Surprisingly, thorough exploration of the space of typical algorithms and variants (such as different signal features, static or dynamic models, parametric or nonparametric estimation, etc.) and exhaustive fine-tuning of the corresponding parameters fail to improve the precision above an empirical *glass ceiling*,[25] around 70% precision (although this of course should be defined precisely and depends on tasks, databases, etc.).
(2) Paradox of dynamics: Further, traditional means to model data dynamics, such as delta coefficients, texture windows, or Markov modeling, do not provide any improvement over the best static models for real-world, complex polyphonic textures of several seconds length.[26] This is a paradoxical observation, since static models consider all frame permutations of the same audio signal as identical, while this has a critical influence on their perception. Moreover, psychophysical experiments[28] have established the importance of dynamics, notably the attack time and fluctuations of the spectral envelope, in the perception of individual instrument notes.
(3) Hubs: Finally, recent experiments[27] show that the BOF approach (when used on polyphonic music) tends to create false positives which are mostly always the same songs regardless of the query. In other words, there exist songs, which we have called *hubs*, which are irrelevantly close to all other songs. This phenomenon is reminiscent of other results in different domains, such as speaker recognition[29] or fingerprint identification,[30] which intriguingly also typically rely on the same BOF approach. This suggests that this could be an important phenomenon which generalizes over the specific problem of polyphonic music similarity, and indicates a general structural property of the class of algorithms examined here, at least *for a given class of signals* to be defined.

C. Objectives

This paper proposes to re-evaluate this situation and to explicitly examine the difference between soundscape and polyphonic music signals with respect to their modeling with the BOF approach.

We apply to a data set of urban soundscapes an algorithmic measure of acoustic similarity that we introduced[25] in the context of polyphonic music. The measure is a typical instantiation of the BOF approach, namely comparing the long-term distributions of MFCC vectors, using Kullback-Leibler divergence between Gaussian mixture models. For music, the measure approximates the perception of similar global timbre, e.g., of songs that "sound the same." As already noted, the measure only achieves moderate precision on music and shows notable discrepancies with human perception. We find here that the same measure is nearly optimal for modeling the perceptual similarity of urban soundscapes. This confirms the situation found in the literature that soundscape and polyphonic music signals are not equal with respect to their modeling with the BOF approach. Notably, the application of timbre similarity to soundscapes does not seem to create hubs.

To explain these differences, we report on two experiments in which we apply specially designed *homogeneity* transforms to each data sets:

(1) Temporal homogeneity, which folds an original signal onto itself a number of times, so the resulting signal only contains a fraction of the original data.
(2) Statistical homogeneity, which only keeps frames in the signal which are the most statistically prototypical of the overall distribution.

We study the influence of each transform on the precision of BOF modeling for both sound-scapes and music, and show very different behaviors. This notably establishes that the distribution of frame-based spectral features is very homogeneous for soundscapes, which makes their BOF modeling very robust to data transformations. i.e., soundscapes can be compressed to only a small fraction of their duration without much loss in terms of distribution modeling. Polyphonic music on the contrary seems to require a large quantity of feature information in order to be properly modeled and compared. Furthermore, it appears that, contrary to environmental textures, not all music frames are equally discriminative: minority frames (the 5% less statistically significant ones) are extremely important for music while they can be discarded to notable advantage for soundscapes. Moreover, it appears that there exists, in typical polyphonic music distributions, a population of frames (in the range [60%–90%] of statistical weight) which is detrimental to the modeling of perceptual similarity.

II. ACOUSTIC SIMILARITY OF URBAN SOUNDSCAPES AND POLYPHONIC MUSIC

A. Algorithm

We sum up here the timbre similarity algorithm presented in Aucouturier and Pachet (2004).[25] The signal is first cut into frames. For each frame, we estimate the spectral envelope by computing a set of MFCCs. We then model the distribution of the MFCCs over all frames using a Gaussian mixture model (GMM). GMM estimates a probability density as the weighted sum of $\mathcal{M}$ simpler Gaussian densities, called components or states of the mixture:

$$p(x_t) = \sum_{m=1}^{m=\mathcal{M}} \pi_m \mathcal{N}(x_t, \mu_m, \Sigma_m) \quad (1)$$

where x_t is the feature vector observed at time t, $\mathcal{N}$ is a Gaussian pdf with mean μ_m, covariance matrix Σ_m, and π_m is a mixture coefficient (also called state prior probability). The parameters of the GMM are learned with the classic E-M algorithm.[20]

We then compare the GMM models to match different signals, which gives a similarity measure based on the audio content of the items being compared. We use a Monte Carlo approximation of the Kullback-Leibler (KL) distance between each duple of models A and B. The KL distance between two GMM probability distributions p_A and p_B [as defined in Eq. (1)] is defined by

$$d(A,B) = \int p_A(x) \log \frac{p_B(x)}{p_A(x)} dx. \quad (2)$$

The KL distance can thus be approximated by the empirical mean:

$$\widetilde{d(A,B)} = \frac{1}{n} \sum_{i=1}^{n} \log \frac{p_B(x_i)}{p_A(x_i)} \quad (3)$$

(where n is the number of samples x_i drawn according to p_A) by virtue of the central limit theorem.

In this work, we use the optimal settings determined by previous research in the context of polyphonic music,[25] namely 20 MFCCs appended with zeroth-order coefficient, 50-component GMMs, compared with $n=2000$ Monte Carlo draws.

B. Data sets

1. Urban soundscapes

For this study, we gathered a database of 106 3 min recordings of urban soundscapes, recorded in Paris using an omnidirectional microphone. The recordings are clustered in four "general classes" as follows:

(1) Avenue: Recordings made on relatively busy thoroughfares, with predominant traffic noise, notably buses and car horns.
(2) Neighborhood: Recordings made on calmer neighborhood streets, with more diffuse traffic, notably motorcycles, and pedestrian sounds.
(3) Street market: Recordings made on street markets in activity, with distant traffic noise and predominant pedestrian sounds, conversation, and auction shouts.
(4) Park: Recordings made in urban parks, with lower overall energy level, distant and diffuse traffic noises, and predominant nature sounds, such as water or bird songs.

Recordings are further labeled into 11 "detailed classes," which correspond to the place and date of recording of a given environment. For instance, "Parc Montsouris (Paris 14è)" is a subclass of the general "Park" class. Some detailed classes also discriminate at identical locations and dates, but with some exceptional salient difference. For instance, "Marché Richard Lenoir (Paris 11è)" is a recording made in a street market on Boulevard Richard Lenoir in Paris, and "Marché Richard Lenoir (music)" is a recording made on the same day of the same environment, only with the additional sound of a music band playing in the street. Table II shows the details of the classes used, and the number of recordings available in each class.

2. Polyphonic music

The polyphonic music data set used in this study contains 350 popular music titles, extracted from the Cuidado database.[31] It is organized in 37 clusters of songs by the same artist, encompassing very different genres and instrumentations (from *Beethoven* piano sonata to *The Clash* punk rock and *Musette*-style accordion). Artists and songs were chosen in order to have clusters that are "timbrally" consistent (all

TABLE II. Composition of the urban soundscape database.

Class	Detailed class	Size
Avenue	Boulevard Arago	14
Avenue	Boulevard du Trône	5
Avenue	Boulevard des Maréchaux	8
Street	Rue de la Santé	7
Street	Rue Reille day1	14
Street	Rue Reille day2	7
Market	Marché Glacière	8
Market	Marché R. Lenoir	22
Market	Marché R. Lenoir (music)	9
Park	Parc Montsouris Spring	20
Park	Parc Montsouris Summer	8

songs in each cluster sound the same). Furthermore, we only select songs that are timbrally homogeneous, i.e., there is no big texture change within each song. The test database is constructed so that nearest neighbors of a given song should optimally belong to the same cluster as the seed song. Details on the design and contents of this database can be found in Aucouturier and Pachet (2004).[25]

C. Evaluation metric

The algorithms are compared by computing their precision after 5, 10, and 15 documents are retrieved, and their *R* precision, i.e., their precision after all relevant document are retrieved. Each value measures the ratio of the number of relevant documents to the number of retrieved documents. The set of relevant documents for a given sound sample is the set of all samples of the same category as the seed. This is identical to the methodology used, e.g., in Aucouturier and Pachet (2004).[25]

D. Results

1. Precision

Table III gives the precision of timbre similarity applied to both data sets. It appears that the results are substantially better for urban soundscapes than for polyphonic music signals, nearing perfect precision in the first five nearest neighbors even for detailed classes. High precision using the general classes shows that the algorithm is able to match recordings of different locations on the basis of their sound level (avenues, streets), and sound quality (pedestrian, birds). High precision on detailed classes shows that the algorithm is also able to distinguish recordings of the same environment made at different times (spring or summer), or in different contexts (with and without music band). This result has a natural application to computer-based classification, e.g., using a simple *k*-nearest neighbor strategy, and could prove useful for context-recognition, for instance in the context of wearable computing.[32]

TABLE III. Comparison of similarity measure for urban soundscapes and polyphonic music.

Database		5 Prec.	10 Prec.	15 Prec.	*R* Prec.
Music		0.73	0.70	0.65	0.65
Soundscapes	General	0.94	0.87	0.77	0.66
	Detailed	0.90	0.79	0.75	0.74

TABLE IV. Five most frequent false positives in the music database.

Song	N_{10}
Mitchell, Joni - Don Juan's Reckless Daughter	57
Moore, Gary - Separate Ways	35
Rasta Bigoud - Tchatche est bonne	30
Public Enemy - Cold Lampin With Flavor	27
Gilberto, Joao - Tin tin por tin tin	25

2. Hubs

As mentioned earlier, an intriguing property of the application of the similarity measure to polyphonic music signals is that it tends to create false positives which are mostly always the same songs regardless of the query. In other words, there exist songs, which we call *hubs*, which are irrelevantly close to all other songs. We give a detailed description of this phenomenon in Aucouturier and Pachet (2007).[27]

A natural measure of the hubness of a given song is the number of times the song occurs in the first n nearest neighbors of all the other songs in the database. An important property of the number of n occurrences N_n of a song is that the sum of the values for all songs is constant given a database. Each query only gives the opportunity for n occurrences to the set of all the other songs, such that the total number of n occurrences in a given $\mathcal{N}$-size database is $n*\mathcal{N}$. Therefore, the mean n occurrence of a song is equal to n, independent of the database and the distance measure.

Table IV shows the five biggest hubs in the polyphonic music database ranked by the number of times they occur in the first ten nearest neighbors over all queries (N_{10}). This illustrates the predominance of a few songs that occur very frequently. For instance, the first song, *Mitchell, Joni—Don Juan's Reckless Daughter* is very close to 1 song out of 6 in the database (57 out of 350), which is more than six times more than the theoretical mean value (10). Among these occurrences, many are likely to be false positives.

Figure 1 shows the histogram of the number of 20-occurrences N_{20} obtained with the above-mentioned distance on the database of urban soundscapes, compared with the same measure on the test database of polyphonic music. It appears that the distribution of number of occurrences for soundscapes is more narrow around the mean value of 20, and has a smaller tail than the distribution for polyphonic music. Notably, there are four times as many audio items with more than 40 20-occurrences in the music data set than in the urban soundscape data set. This is also confirmed by the manual examination of the similarity results for the urban soundscapes: none of the (few) false positives reoccur significantly more than random.

This establishes the fact that hubs are not an intrinsic property of the class of algorithm used here, but rather appear only for a certain classes of signals, which includes polyphonic music, but not urban soundscapes.

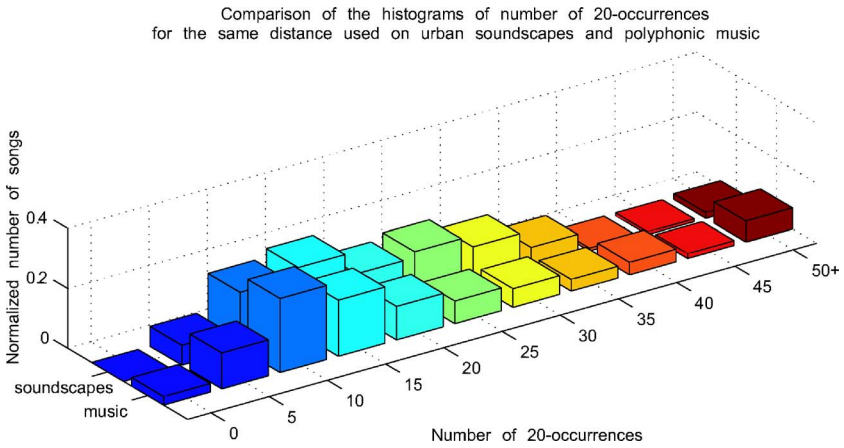


FIG. 1. (Color online) Comparison of the histograms of number of 20-occurrences for the same distance used on urban soundscapes and polyphonic music.

On the whole, these results confirm that urban soundscapes and polyphonic music signals are not equal with respect to their modeling with the BOF approach. To explain these differences, we now report on two experiments in which we apply specially designed *homogeneity* transforms to each data set. We study the influence of each transform on the precision of BOF modeling for both soundscapes and music, and observe very different behaviors.

III. TEMPORAL HOMOGENEITY

A. Transform

We consider a temporal homogeneity transformation of audio data which folds an original signal onto itself a number of times (as seen in Fig. 2). The output of the twofold transform is 50%-sized random extract from the original, repeated twice. Similarly, the threefold transform is a 33%-sized extract of the original repeated three times. All signals processed by n folding from a given signal have the same duration as the original, but contain less "varied" material. Note that since the duration of the fold (an integer division of the total duration) is not a multiple of the frame duration in the general case, n folding does not simply duplicate the MFCC frames of the folded extract, but rather creates some limited jitter. The fact that all n-folded signals have the same number of frames as the original enables one to use the same modeling parameters, notably number of Gaussian components (otherwise we would have had to account for the curse of dimensionality).

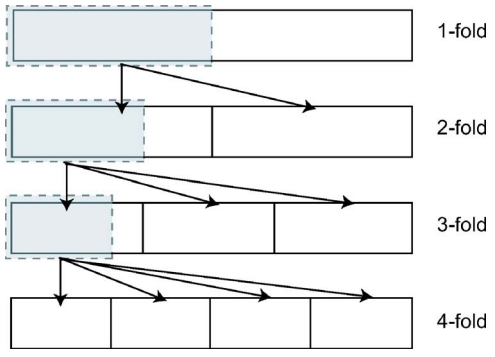


FIG. 2. (Color online) Illustration of applying three successive temporal homogeneity transforms to an audio signal, by folding it twice ("twofold"), three times ("threefold"), and four times ("fourfold"). The transform creates increasingly homogeneous signals by folding a reduced portion of the original signal. Note that the "onefold" transform is the "identity" operator.

We apply nine n-folding transforms for $n \in [1,2,3,4,5,10,20,30,50]$ to the audio signals of each data set (soundscapes and music). Each transformed signal is then processed with the above-described algorithm, namely GMM of MFCCs. This yields nine types of GMM for each original signal in a given data set, and nine similarity measures for each data set.

B. Influence on variance

Figure 3 shows the influence of n folding on the mean variance of the GMM of the transformed signals. The variance of a GMM model can be defined by sampling a large number of points from this model, measuring the variance of these points in each dimension, and summing the deviations together. This is equivalent to measuring the norm of the covariance matrix of a single-component GMM fitted to the distribution of points.[33]

The temporal homogeneity transform has a very different influence on GMM variance when applied to urban soundscapes and music signals. The GMM variance of soundscape signals shows little dependency on temporal homogenization for ratios as low as 10% of the original signal duration. For extreme number of folds (greater than 10), the GMM variance tends to decrease slightly. This shows that the statistics of urban soundscape signals are stationary on time scales of the order of 10 s.

On the contrary, temporal homogenization has a complex influence on the GMM variance polyphonic music signals. Folding audio extracts of the original signal with durations down to 50% of the original signal's tends to reduce GMM variance. However, when the number of folds is greater than 2, the variance exponentially increases. It reaches its original 100% value when folding 15% of the signal's original duration, and increases to more than twice its original value for ratios lower than 5%. This shows that extracts smaller than half of the original duration (i.e., of the order of 100 s) are typically more heterogeneous than the overall signal in the case of polyphonic music. This indicates a rather high density of outlier frames, whose probability is overestimated when considering small extracts.

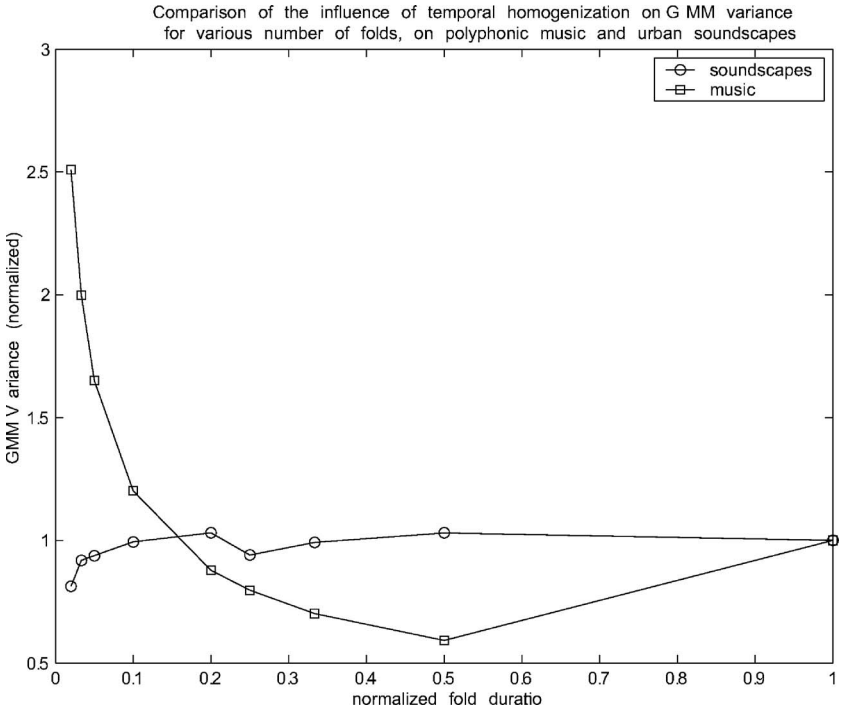


FIG. 3. Influence of temporal homogeneity transform on the mean variance of the GMMs of urban soundscapes and music signals.

C. Influence on precision

Figure 4 shows the influence of folding on the similarity of *R* precision for both classes of signals (where both precision curves are normalized with respect to their maximum). *n* folding is detrimental to the precision for both data sets. However, it appears that urban soundscapes are typically twice more robust to folding than polyphonic signals. Considering only a tenth of the audio signals cuts down precision by 15% for soundscapes, and by more than 35% for polyphonic music. In the extreme case of folding only 3 s out of a 3 min sound extract (50-folding), the precision loss is 20% for soundscapes, but more than 60% for polyphonic music.

This suggests that frame-based feature distributions for urban soundscapes are statistically much more self-similar than polyphonic music, i.e., they can be compressed to only a small fraction of their duration without much loss in terms of distribution modeling. If we authorize a 10% precision loss, soundscapes can be reduced to 10 s extracts. Polyphonic music on the contrary seems to require a large quan-

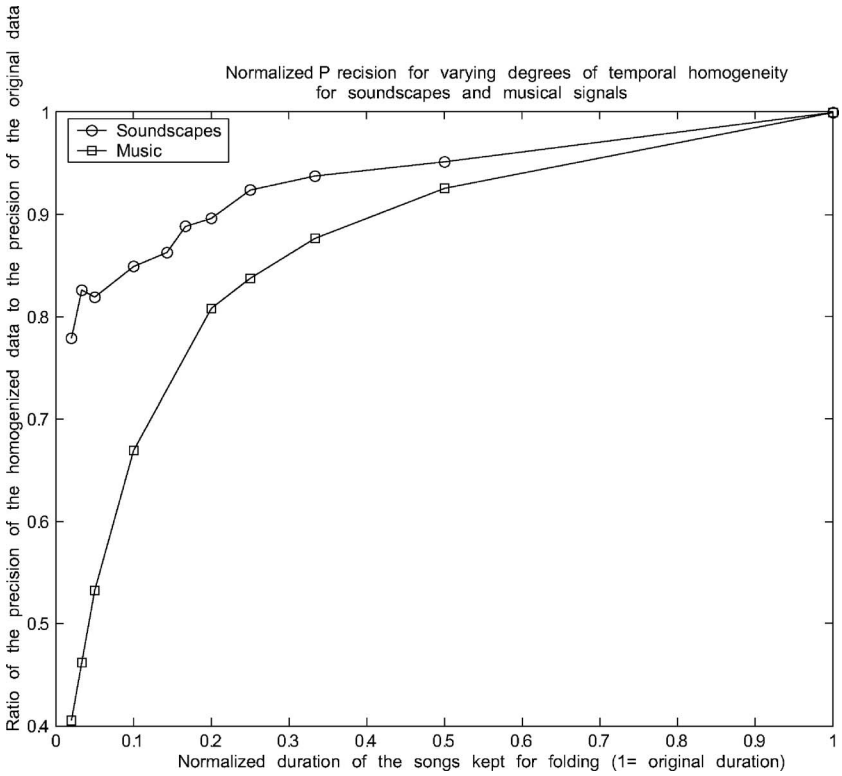


FIG. 4. Influence of temporal homogeneity transform on the precision of the similarity measure for urban soundscapes and music signals.

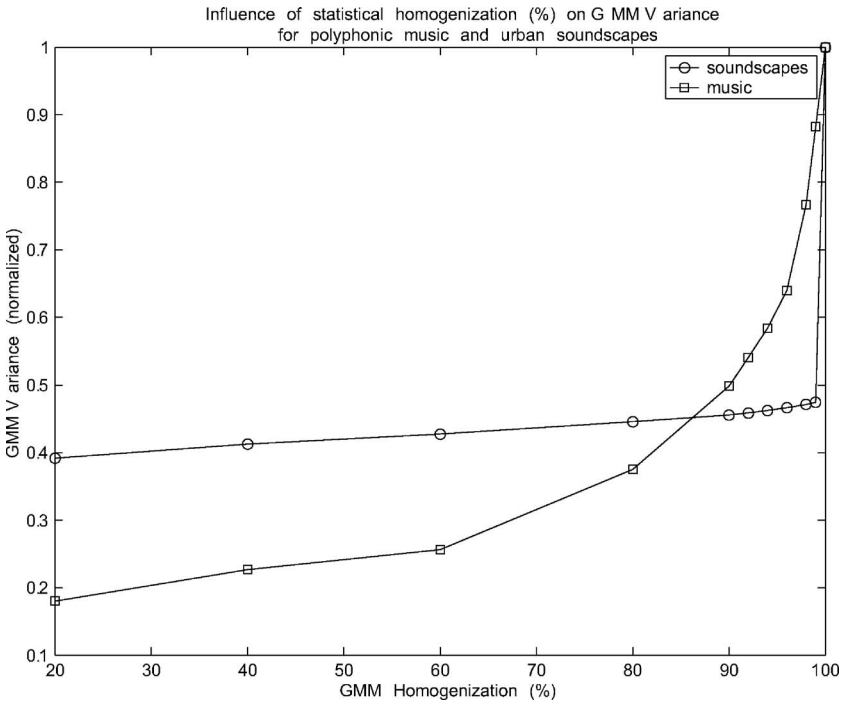


FIG. 5. Influence of statistical homogeneity transform on the variance of the GMMs of urban soundscapes and music signals.

tity of feature information in order to be properly modeled and compared: the same 10% tolerance requires more than 1 min of data.

Note that the former is comparable to the human performance[4] on the task of recognizing everyday auditory scenes (20 s). However, the latter (polyphonic music) is many times less effective than humans, who have been reported able to issue categorical judgments with good precision using as little as 200 ms of audio.[34]

IV. STATISTICAL HOMOGENEITY

A. Transform

We define a statistical homogeneity transform $h_k: \mathcal{G} \circlearrowright \mathcal{G}$ on the space $\mathcal{G}$ of all GMMs, where $k \in [0,1]$ is a percentage value, as:

$g_2 = h_k(g_1)$

$(c_1, \ldots, c_n) \leftarrow \text{sort}(\text{components}(g_1), \text{decreasing } w_c)$

$\text{define } \mathcal{S}(i) = \sum_{j=1}^{i} weight(c_j)$

$i_k \leftarrow \arg\min_{i \in [1,n]} \{\mathcal{S}(i) \geqslant k\}$

$g_2 \leftarrow \text{newGMM}(i_k)$

$\text{define } d_i = \text{component}(g_2, \text{i})$

$d_i \leftarrow c_i, \quad \forall\, i \in [1, i_k]$

$\text{weight}(d_i) \leftarrow \text{weight}(c_i)/\mathcal{S}(i_k), \quad \forall\, i \in [1, i_k]$

return g_2

end h_k

From a GMM g trained on the total amount of frames of a given song, the transform h_k derives an homogenized version of g which only contains its top $k\%$ components. Frames are all the more so likely to be generated by a given Gaussian component c than the weight w_c of the component is high (w_c is also called prior probability of the component). Therefore, the homogenized GMM accounts for only a subset of the original song's frames: those that amount to the $k\%$ most important statistical weight. For instance, $h_{99\%}(g)$ creates a GMM which does not account for the 1% least representative frames in the original song.

We apply 11 transforms h_k for $k \in [20,40,60,80,90,92,94,96,98,99,100]$ to the GMMs used in the above-described similarity measure. Each transform is applied on each data set, thus yielding two sets of 11 similarity measures, the properties of which we study in the following.

B. Influence on variance

Figure 5 shows the influence of the statistical homogenization transform on the variance of the resulting GMM for both data sets. The variance of the model is evaluated with the sampling procedure already described in Sec. III B.

Again, the transformation has a very distinct influence on each type of audio signal. Removing the least important 1% frames from urban soundscape signals drastically reduces the GMM variance by more than 50%. However, further statistical homogenization has little influence on the overall variance. This indicates that soundscape signals are very homogeneous and redundant statistically, except for a very small proportion of outlier frames (the least significant 1%), which account for half of the overall variance, and probably represent very different MFCC frames from the

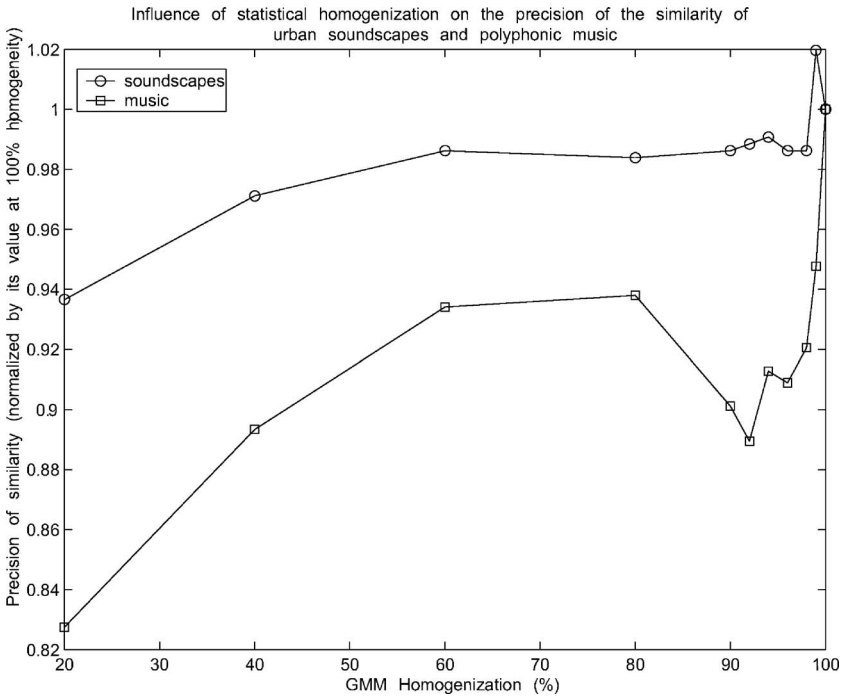


FIG. 6. Comparison of the influence of statistical homogeneity transform on the precision of the similarity measure for urban soundscapes and music signals.

ones composing the main mass of the distribution. Such frames would typically represent very improbable sound events which are not characteristic of a given environment, such as the occasional plane flying over a park.

When applied to polyphonic music signals, it appears that the homogenization transform reduces the variance of the models exponentially. Half of the original variance is explained by the 10% least representative frames, and more than 80% by the 40% least representative frames. This indicates a greater heterogeneity than for soundscape signals, and a more diffuse notion of "outlier" frames.

C. Influence on precision

Figure 6 shows the influence of statistical homogenization on the precision of the resulting similarity measure, for both data sets. The precision for urban soundscapes is measured with the 10 precision using the detailed classes as ground truth, and with the R precision for polyphonic music. For both data sets the precision is measured by reference to the baseline precision corresponding to $k=100\%$, which is different for soundscapes and music, as shown in Table III.

On both data sets, increased homogenization decreases the precision of the similarity measure: homogenization with $k=20\%$ degrades the measure's precision by 6% (relative) for urban soundscapes, and by 17% (relative) for polyphonic music. It seems reasonable to interpret the decrease in precision when k decreases as a consequence of reducing the amount of discriminative information in the GMMs (e.g., from representing a given song, down to a more global style of music, down to the even simpler fact that it *is* music).

Apart from this general trend however, the transform has a very different influence on the measure's precision depending on the class of audio signals.

In the case of urban soundscapes, 99% homogenization is slightly beneficial to the precision. This suggests that the 1% less significant frames, which were found in Fig. 5 to account for half of the overall variance, are spurious frames which are worth smoothing out. Further homogenization down to 60% has a moderate impact on the precision, which is reduced by about 1% (absolute). The decrease in precision from 99% down is monotonic. This suggests that the frame distribution from 99% down is very homogeneous and redundant. Urban soundscapes can be discriminated nearly optimally by considering only the most significant 50% of the frames.

In the case of polyphonic music, the decrease in precision is not monotonic. Figure 6 clearly shows a very important decrease in the precision in the first few percent of homogenization. The severely degraded precision observed for $k=30\%$ is reached as early as $k=95\%$. This is a strong observation: the precision of the measure seems to be controlled by an extremely small amount of critical frames, which represent typically less than 5% of whole distribution. Moreover, these frames are the least statistically significant ones, i.e., are modeled by the least important Gaussian components in the GMMs. This indicates that the majority (more than 90%) of the MFCC frames of a given song are a very poor representation of what discriminates this song from other songs. This is the exact opposite behavior to the one observed for soundscape signals, where these least significant frames can be removed to some advantage.

Moreover, Fig. 6 shows that after the abrupt sink when removing the first 5% frames in typical music distributions, the precision tends to increase when k decreases from 90% to 60%, and then decreases again for k smaller than 60%. The maximum value reached between 60% and 80% is only 6% (relative) lower than the original value at $k=100\%$.

The behavior in Fig. 6 suggests that there is a population of frames in the range $[60\%, 95\%]$ which is mainly responsible for the bad precision of the measure on music signals. While the precision of the measure increases as more frames

are included when k increases from 20% to 60% (such frames are increasingly specific to the song being modeled), it suddenly decreases when k gets higher than 60%, i.e., this new 30% information is detrimental for the modeling and tend to diminish the discrimination between songs. The continuous degradation from 60% to 95% is only eventually compensated by the inclusion of the final 5% critical frames.

V. DISCUSSION

A. Physical specificities in each class of sounds

We observe critical differences in the temporal and statistical structure of the typical frame distribution for soundscapes and polyphonic music signals. The experiments reported here show that frames in polyphonic music signals are not equally discriminative/informative, and that their contribution to the precision of a simulated perceptual similarity task is not proportional to their statistical importance and long-term frequency (i.e., the corresponding component's prior probability w_c):

(1) The very informative frames for the simulation of the perception of polyphonic music (measured by their effect of acoustic similarity) are the least statistically representative (the bottom 1%).
(2) A large population of frames (in the range [60%, 95%]) is detrimental to the modeling. Another study by the authors[27] shows that the inclusion of these frames increase the hubness of a song, i.e., their statistical weight masks important and discriminative details found elsewhere in statistical minority.

Such structure cannot be observed in the frame distribution of typical urban soundscape signals.

B. A possible reason for the failure of BOF

Such differences in homogeneity for each class of signals can be proposed to explain the uneven performance of their respective modeling with the BOF approach. High performance with BOF correlates with high homogeneity: BOF-based techniques are very efficient for soundscapes, with both high precision and absence of perceptive paradoxes like hubs, while they fail for polyphonic music, which is more heterogeneous.

However, we do not give here any formal proof that heterogeneity is the main factor in explaining the failure of BOF modeling for polyphonic music signals. More complete evidence would come, e.g., by synthezing artificial signals spanning a more complete range of homogeneity values, and by comparing algorithmic predictions to human perceptive judgments.

C. Psychological relevance

The BOF approach to simulate the auditory perception of signals such as soundscapes and music makes an implicit assumption about the perceptive relevance of sound events. Distributions are compared (e.g., with the Kullback Leibler distance) on the basis of their most stereotypical frames. Therefore, with BOF algorithms, frames contribute to the simulation of the auditory sensation in proportion of their statistical predominance in the global frame distribution. In other words, the *perceptive saliency*[35] of sound events is modeled as their *statistical typicality*.

BOF is not intended (neither here nor in the pattern recognition literature) as a cognitive model, but rather is an engineering technique to simulate and replicate the outcome of the corresponding human processing. Nevertheless, it is useful to note that the above-mentioned model of auditory saliency would be a very crude cognitive model indeed, both to model preattentive weighting (which has been found a correlate of frequency and temporal contrasts,[36] i.e., arguably the exact opposite of statistical typicality) and higher-level cognitive processes of selective attention (which are partly under voluntary control, hence products of many factors such as context and culture[37]).

The above-presented results establish, as expected, that the mechanism of auditory saliency implicitly assumed by the BOF approach does not hold for polyphonic music signals: For instance, frames in statistical minority have a crucial importance in simulating perceptive judgments. However, surprisingly, the crude saliency hypothesis seems to be an efficient/sufficient representation in the case of soundscapes: Frames are found to contribute to the precision of the simulated perceptive task in degrees correlated with their global statistical typicality, and overall BOF provide near-perfect replication of human judgments.

The fact that such a simple model is sufficient to simulate the perception of sound-scapes could suggest that the cognitive processes involved in their human processing are less "demanding" than for polyphonic music. This finding is only based on algorithmic considerations, and naturally would have to be validated with proper psycho-sociological experimentations. Nevertheless, it seems at odds with a wealth of recent psychological evidence stressing that soundscapes judgment does not result in a low-level immediate perception, but rather high-level cognitive reasoning which accounts for the evidence found in the signal, but also depends on cultural expectations, *a priori* knowledge, or context. For instance, the subjective evaluation of urban soundscapes has been found to depend as much on semantic features than perceptual ones: Soundscapes reflecting activities with higher cultural values (e.g., human versus mechanical) are systematically perceived as more pleasant.[5] Similarly, cognitive categories have been found to be mediated by associated behaviors and interaction with the environment: A given soundscape can be described as, e.g., "too loud to talk," but "quiet enough to sleep."[38]

What our results could indicate is that, while there are indeed important and undisputed high-level cognitive processes in soundscape perception, these may be less critical in shaping the overall perceptive categories than for polyphonic music. Discarding such processes hurts the perception of music more than that of soundscapes.

A possible reason for this is that there are important specificities in the structure of polyphonic music, namely very definite temporal units (e.g., notes) with both internal (transient, steady-state) and external (phrase, rhythm) organization. For instance, a recent study[39] in automatic instru-

ment classification suggests that the transient part of individual notes concentrates very discriminative information for timbre identification, but that its scarsity with respect to longer steady-state information makes it difficult to exploit for machine learning algorithms. This situation of trading too little good information against too much poor-quality information is reminiscent of what we observe here. Human perception, by its higher-level cognitive processing of the structure of musical notes, gives increased saliency to frames that are otherwise in statistical minority.

Such structural specificities in polyphonic music signals may require cognitive processes active on a more *symbolic and analytical* level than what can be accounted for by the BOF approach, which essentially builds an *amorphous and holistic* description of the object being modeled. These computational experiments open the way for more careful psychological investigations of the perceptive paradoxes proper to polyphonic music timbre, in which listeners "hear" things that are not statistically significant in the actual signal, and that the low-level models of timbre similarity studied in this work are intrinsically incapable of capturing.

ACKNOWLEDGMENTS

The authors would like to thank Anthony Beurivé for helping with the implementation of signal processing algorithms and database metadata management. We also thank Laurent Daudet and Pierre Leveau, as well as anonymous reviewers for their valuable comments. This research has been partially funded by the Semantic Hifi IST European project (http://shf.ircam.fr/).

[1]M. Schafer, *The Tuning of the World* (Random House, Rochester, VT, 1977).
[2]P. Lercher and B. Schulte-Forktamp, "The relevance of soundscape research to the assessment of noise annoyance at the community level," in the Eighth International Congress on Noise as Public Health Problem, Rotterdam, The Netherlands, 2003.
[3]J. Ballas, "Common factors in the identification of an assortment of brief everyday sounds," J. Exp. Psychol. Hum. Percept. Perform. **19**, 250–267 (1993).
[4]V. Peltonen, A. Eronen, M. Parviainen, and A. Klapuri, "Recognition of everyday auditory scenes: Potentials, latencies and cues," in Proceedings of the 110th Convention of the Audio Engineering Society, Amsterdam, The Netherlands, 2001.
[5]D. Dubois, C. Guastavino, and M. Raimbault, "A cognitive approach to urban sound-scapes: Using verbal data to access everyday life auditory categories," Acta. Acust. Acust. **92**, 865–874 (2006).
[6]Note that this ("birds, children") is not intended as the definition *in intension* of a sociological representation of what a "park" is (for which we would have to give evidence and cultural context, as in Ref. 40), but only as an arbitrary example of an individual perceptual category (see Ref. 5).
[7]B. Defréville and C. Lavandier, "The contribution of sound source characteristics in the assessment of urban soundscapes," Acta. Acust. Acust. **92**, 912–921 (2006).
[8]C. Guastavino, B. Katz, J. Polack, D. Levitin, and D. Dubois, "Ecological validity of soundscape reproduction," Acta. Acust. Acust. **91**, 333–341 (2005).
[9]D. Dufournet, P. Jouenne, and A. Rozwadowski, "Automatic noise source recognition," J. Acoust. Soc. Am. **103**(5), 2950 (1998).
[10]C. Couvreur, V. Fontaine, P. Gaunard, and C. G. Mubikangiey, "Automatic classification of environmental noise events by hidden Markov models," Appl. Acoust. **54**, 187–206 (1998).
[11]M. Casey, "Mpeg-7 sound recognition tools," IEEE Trans. Circuits Syst. Video Technol. **11**, 737–747 (2001).
[12]M. Cowling and R. Sitte, "Comparison techniques for environmental sound recognition," Pattern Recogn. Lett. **24**, 2895–2907 (2003).
[13]A. Harma, J. Skowronek, and M. McKinney, "Acoustic monitoring of the patterns of activity in the office and the garden," in Proceedings of the fifth International Conference on Methods and Techniques in Behavioral Research, Wageningen, The Netherlands, 2005.
[14]B. Defréville, P. Roy, C. Rosin, and F. Pachet, "Automatic recognition of urban sound sources," in Proceedings of the 120th Audio Engineering Society Convention, Paris, France, 2006.
[15]L. Rabiner and B. Juang, *Fundamentals of Speech Recognition* (Prentice-Hall, Englewood Cliffs, NJ, 1993).
[16]K. El-Maleh, A. Samouelian, and P. Kabal, "Frame level noise classification in mobile environments," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Phoenix, AZ, March 1999.
[17]V. Peltonen, J. Tuomi, A. Klapuri, J. Huopaniemi, and T. Sorsa, "Computational auditory scene recognition," in Proceedings of the International Conference on Acoustic, Speech and Signal Processing (ICASSP), Orlando, FL, 2002.
[18]L. Ma, D. Smith, and B. Milner, "Context awareness using environmental noise classification," in Proceedings of Eighth European Conference on Speech Communication and Technology (Eurospeech), Geneva, Switzerland, 2003.
[19]F. Sebastiani, "Machine learning in automated text categorization," ACM Comput. Surv. **34**, 1–47 (2002).
[20]C. M. Bishop, *Neural Networks for Pattern Recognition* (Oxford University Press, Wulton Street, Oxford, 1995).
[21]ISMIR, International Conference on Music Information Retrieval; http://www.ismir.net
[22]G. Tzanetakis, G. Essl, and P. Cook, "Automatic musical genre classification of audio signals," in The Second International Conference on Music Information Retrieval (ISMIR), Oct. 2001, Bloomington, Indiana.
[23]D. Liu, L. Lu, and H.-J. Zhang, "Automatic mood detection from acoustic music data," in Proceedings of the Fourth International Conference on Music Information Retrieval (ISMIR), Baltimore, MD, 2003.
[24]W.-H. Tsai and H.-M. Wang, "Towards automatic identification of singing language in popular music recordings," in Proceedings of the International Conference on Music Information Retrieval (ISMIR), Barcelona, Spain, 2003.
[25]J.-J. Aucouturier and F. Pachet, "Improving timbre similarity: How high's the sky?," Journal of Negative Results in Speech and Audio Sciences **1**, (2004).
[26]J.-J. Aucouturier and F. Pachet, "The influence of polyphony on the dynamical modelling of musical timbre," Pattern Recogn. Lett. **28**(5), 654–661 (2007).
[27]J.-J. Aucouturier and F. Pachet, "A scale-free distribution of false positives for a large class of audio similarity measures," Pattern Recogn. http://dx.doi.org/10.1016/j.patcog.2007.04.012
[28]S. S. McAdams, S. Winsberg, S. Donnadieu, G. De Soete, and J. Krimphoff, "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes," Psychol. Res. **58**, 177–192 (1995).
[29]G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, "Sheep, goats, lambs and wolves, a statistical analysis of speaker performance," in Proceedings of the Fifth International Conference on Spoken Language Processing (ICSLP), Sydney, Australia, 1998.
[30]A. Hicklin, C. Watson, and B. Ulery, "The myth of goats: How many people have fingerprints that are hard to match?," Patriot Act Report 7271, National Institute of Standards and Technology, Gaithersburg, MD (2005).
[31]F. Pachet, A. LaBurthe, A. Zils, and J.-J. Aucouturier, "Popular music access: The Sony music browser," J. Am. Soc. Inf. Sci. **55**(12), 1037–1044 (2004).
[32]B. Clarkson, N. Sawhney, and A. Pentland, "Context awareness via wearable computing," in Proceedings of the 1998 Workshop on Perceptual User Interfaces (PUIE8), San Francisco, CA, 1998.
[33]Note that a more precise measure of the width of a GMM is nontrivial to compute from the variance of its individual components (e.g., summing them, weighted with each component's prior probability), because this would have to account for the possible overlap between individual components (i.e., computing the volume of the intersection between a set of many ellipsoids in a high dimension space).
[34]D. Perrot and R. O. Gjerdinger, "Scanning the dial: An exploration of factors in the identification of musical style," in Proceedings of the 1999 Society for Music Perception and Cognition, Evanston, IL (1999).
[35]The saliency of a sound event is defined as what makes it attract auditory attention, thereby providing its weight in the representation of our envi-

ronment. Here we consider a rather permissive notion of saliency, which encompasses both preattentive mechanisms and higher-level selective attention effects.
[36]C. Kayser, C. Petkov, M. Lippert, and N. K. Logothetis, "Mechanisms for allocating auditory attention: An auditory saliency map," Curr. Biol. **15**, 1943–1947 (2005).
[37]P. Janata, B. Tillmann, and J. Bharucha, "Listening to polyphonic music recruits domain-General attention and working memory circuits," Cognitive, Affective and ehavioral Neuroscience **2**, 121–140 (2002).
[38]C. Guastavino, "Categorization of environmental sounds," Can. J. Exp. Psychol., **60**(1), 54–63 (2007).
[39]S. Essid, P. Leveau, G. Richard, L. Daudet, and B. David, "On the usefulness of differentiated transient/steady-state processing in machine recognition of musical instruments," in Proceedings of the 118th AES Convention, Barcelona, Spain, 2005.
[40]B. De Coensel and D. Botteldooren, "The quiet rural soundscape and how to characterize it," Acta. Acust. Acust. **92**, 887–897 (2006).

Measurement and modeling of the acoustic field near an underwater vehicle and implications for acoustic source localization

Paul A. Lepper[a] and Gerald L. D'Spain
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla. California 92093

(Received 5 August 2006; revised 17 May 2007; accepted 18 May 2007)

The performance of traditional techniques of passive localization in ocean acoustics such as time-of-arrival (phase differences) and amplitude ratios measured by multiple receivers may be degraded when the receivers are placed on an underwater vehicle due to effects of scattering. However, knowledge of the interference pattern caused by scattering provides a potential enhancement to traditional source localization techniques. Results based on a study using data from a multi-element receiving array mounted on the inner shroud of an autonomous underwater vehicle show that scattering causes the localization ambiguities (side lobes) to decrease in overall level and to move closer to the true source location, thereby improving localization performance, for signals in the frequency band 2–8 kHz. These measurements are compared with numerical modeling results from a two-dimensional time domain finite difference scheme for scattering from two fluid-loaded cylindrical shells. Measured and numerically modeled results are presented for multiple source aspect angles and frequencies. Matched field processing techniques quantify the source localization capabilities for both measurements and numerical modeling output. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2749410]

PACS number(s): 43.60.Kx, 43.60.Fg, 43.60.Jn, 43.30.Wi [EJS] Pages: 892–905

I. INTRODUCTION

Effects such as diffraction and scattering by an object placed in an acoustic field can result in a complex field structure around that object. Diffracted energy from a source may be observed in the geometrical shadow zone behind an object and so cannot be described by simple ray-based modeling. This energy can also interact with the incident signal energy outside the shadow zone, resulting in interference fields that are observed in conventional optics and acoustics (Morse and Ingrad, 1986 and Skelton *et al.*, 1997). The resulting interference patterns often exhibit strong spatial, angular, and spectral dependence for particular source-scatterer-receiver geometries and therefore may provide valuable information regarding the source location.

An investigation of the complex field around two spheres at multiple source aspect angles and signal frequencies has been carried out using both measured data and numerical modeling results. The measurements were made by an eight-element hydrophone array mounted on the Marine Physical Laboratory's Odyssey IIb autonomous underwater vehicle (AUV). The body of this vehicle is free flooding with a majority of the control and instrumentation electronics contained in two 17-in. (43.2 cm)-diameter air-filled glass spheres. The outer shell of the vehicle is a thin walled (3.5 mm) shroud of high-density polyethylene. For the numerical modeling, the shroud is assumed to be acoustically transparent at the frequencies of interest and the glass spheres are considered to be the major acoustic scatterers.

Motivation for this study was provided by the results from the field of human hearing. Studies with humans have shown that the spectral patterns resulting from the interference of the direct path arrival with sound scattered from the head, torso, and pinna provide important cues for source localization (Blauert, 1983). In particular, for sound sources that vary in elevation angle in the median sagittal plane of the human body, a notch occurs in the head-related transfer function that varies in frequency (from about 6 to 10 kHz) in a sensitive way with changes in elevation angle (Butler and Balendiuk, 1997). Therefore, the two instrumentation glass spheres in the AUV used in this study can be viewed as playing the role of two heads.

The aim of this paper is to examine source localization performance in the presence of scattering from the platform making the measurements. In Sec. II, the measurements of the pressure field using an eight-element hydrophone array mounted on the Odyssey IIb are presented. These measurements were made at the Transducer Evaluation Center (TRANSDEC) of the Space and Naval Warfare Systems Center, San Diego, where data for multiple source aspect angles and for frequencies in the range of 70 Hz–8 kHz were collected. Section III outlines a numerical modeling effort using CABRILLO (Gerstoft, 2004), a two-dimensional (2D) time domain finite difference scheme, (Luebbers and Beggs, 1992 and Yee, 1966). A comparison between measured data and the numerical modeling results is given in Sec. IV for the midfrequency (2–8 kHz) tone bursts. The disadvantage of the comparison is that the 2D modeling approach does not account for the three-dimensional (3D) configuration of the AUV or the potential elastic scattering effects of the spheres. However, the results presented do

[a]Author to whom correspondence should be addressed. Electronic mail: p.a.lepper@lboro.ac.uk

 0001-4966/2007/122(2)/892/14/$23.00

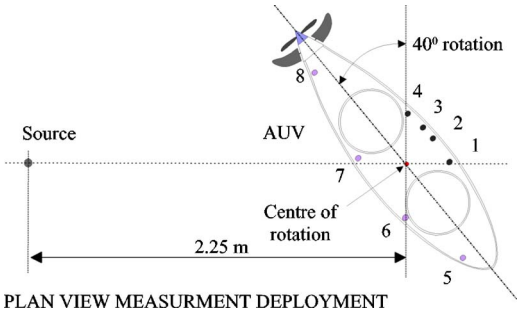


FIG. 1. (Color online) The plan view of the experimental configuration for the measurement effort at the TRANSDEC transducer calibration facility.

illustrate many major physical effects that contribute to the complex acoustic field structure around the AUV. The use of these fields for source localization is then discussed in Sec. V. Results are quantified using matched field processing techniques. Finally, conclusions from this work are given in Sec. VI.

II. DESCRIPTION OF THE MEASUREMENTS

A. Experimental setup

Measurements of the acoustic pressure field surrounding the Marine Physical Laboratory's AUV were made at the TRANSDEC facility, a 22 million l fresh water tank, comprising a 53-m-diam inner pool and a canted, ellipse-shaped outer section. The depth of the inner pool varies from 12 m in the center to less than 1 m at the edge to minimize acoustic reflections from the edge of the pool. The ellipsed-shaped outer section also acts as an acoustic trap.

Figure 1 shows a schematic of the AUV deployment at TRANSDEC. The vehicle was placed in the center of the tank at the mid-water depth of 6 m. An omni-directional acoustic source was placed at a horizontal range of 2.25 m from the center of rotation of the AUV body (displaced 4 cm aft along the main axis of the AUV from the point equidistant between the two glass spheres). The vehicle was suspended from a single rigid shaft linked to a geared stepper motor, allowing precise control of the source / AUV aspect angle in the horizontal plane over a 360° sector. Figure 2 shows a side view of the AUV. All receivers were placed in approximately the same horizontal plane corresponding to the equatorial plane of symmetry of the two glass spheres and within the outer polyethylene shell of the AUV.

The data acquisition system installed in one of the glass spheres was a PC-104+ based digital recording system capable of collecting up to eight channels of data. Eight High

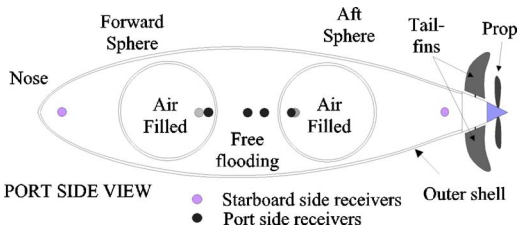


FIG. 2. (Color online) Port-side view of the AUV hydrophone configuration during the measurement effort at TRANSDEC.

Tech, Inc. HTI-94-SSQ hydrophones with built-in 18 dB preamplifiers were used as the array elements. The analog output from each hydrophone/preamp assembly was band-pass filtered between 20 Hz and 10 kHz and amplified by 20 dB for an overall channel sensitivity of −160 dB re 1 V/μPa. The resulting filtered and amplified analog signals were sent to a common analog-to-digital converter, which provided 16 bit data samples from each channel at a 20 ksamples/s sampling rate. These digital data then were written directly to hard disk. The DAQ electronics were housed in the aft sphere in the AUV along with an independent battery power supply allowing continuous recording for up to 4 h. Further details on the hydrophone array and data acquisition system are provided in Zimmerman *et al.*, 2005.

B. Experimental procedure

Pulsed continuous wave (cw) combs were transmitted in the low (less than 1 kHz) and mid (1–10 kHz) frequency ranges using a USRD-J15 moving coil source and a ITC-1007 ceramic transducer, respectively. Each of the pulsed comb signals was composed of a linear sum of pulsed cw tones. For the mid-frequency comb signals, a pulse was 6.4 ms in duration, corresponding to the time difference between the direct path arrival and the arrival of the first reflections from the surface and tank bottom. The low-frequency comb was transmitted as both pulsed with 12.8 ms duration (the longer pulse length was required to achieve the frequency resolution necessary to separate the low-frequency tones) and as a steady-state cw comb.

During each test, the AUV was rotated continuously through a 360° sector at an angular speed of either 0.2 deg s^{-1} or 0.4 deg s^{-1}. This rotation rate provided a 0.2° or 0.4° angular resolution, respectively. At the higher rotation rate, the variation of 2.56×10^{-3} deg during the measurement period of 6.4 ms was considered insignificant. Tests were conducted over a 360° sector for both comb signals. Each test was started with the AUV at broadside (0°) to the source with receivers 5 to 8 (starboard side) closest to the source, as shown in Fig. 1. The AUV was rotated counter-clockwise through positive angles with the propellor of the AUV closest to the source at an angle of +90° and the nose closest at +270°.

C. Signal analysis

Figure 3 shows the time-frequency response for the mid-frequency comb (1900, 2925, 4060, 5100, 6030, 7080, and 7910 Hz) transmission received on hydrophone 1 with the AUV/source orientation in the broadside (0°) position. At broadside, hydrophone 1 is in the geometrical shadow zone of the forward sphere. Therefore, the initial arrivals at 7910, 7080, and 6030 Hz correspond to scattering of the direct path within the AUV. These arrivals are followed 6.4 ms later by stronger multi-path arrivals via surface and tank bottom reflections. This arrival structure can be compared with that recorded on hydrophone 1 when the AUV is rotated 180°, as shown in Fig. 4. The direct path from the source in this case is unimpeded by any part of the AUV.

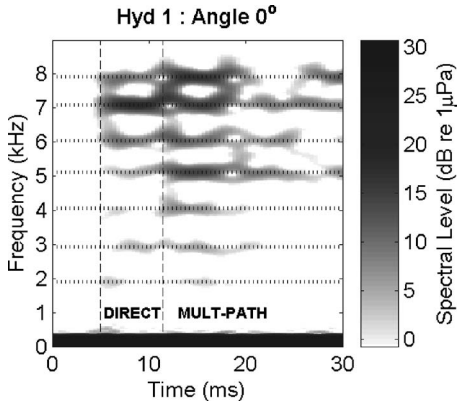


FIG. 3. Spectrogram from hydrophone 1's recording of the mid-frequency comb signal (1900, 2925, 4060, 5100, 6030, 7080, and 7910 Hz) at 0° incidence angle (broadside to the starboard side).

A pulse duration of 6.4 ms will propagate an equivilent distance of 9.6 m in water from beginning to end. This can be compared to the significantly smaller maximum receiver separation in the AUV of 1.55 m. The tank multi-path free pulse duration of 6.4 ms was therefore considered long enough to allow the interaction of multiple reflections between different parts of the AUV and for a steady-state condition to be reached.

Figure 5 shows the spectral levels of the mid-frequency comb signal (1900, 2925, 4060, 5100, 6030, 7080, and 7910 Hz), recorded by hydrophone 1 as a function of incidence angle in azimuth. All the tone levels exhibit an angular interval with high received levels, consistent with direct exposure of the receiver to the source (incidence angles between 110° and 250°). However, the tones also exhibited a degree of variation in received level over this same interval, where the variations have a strong angular, spatial, and frequency dependence. Many of the tones in Fig. 5 show a relatively complex amplitude structure over the complete 360° sector. In the case of the 7910 Hz signal, as much as 30 dB variation in received level can be observed even where the receiver is within the shadow zone of either of the spheres (30°–90° and 280°–330°). In general, the dependence of the spectral levels on incidence angle decreases in complexity with decreasing frequency, although all the tones exhibit some angular structure over the entire sector. The

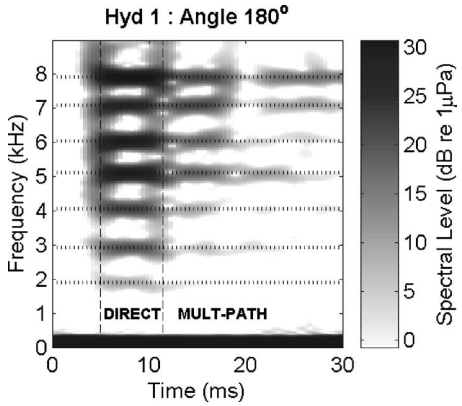


FIG. 4. Spectrogram from hydrophone 1's recording of the mid-frequency comb signal at 180° incidence angle (broadside to the port side).

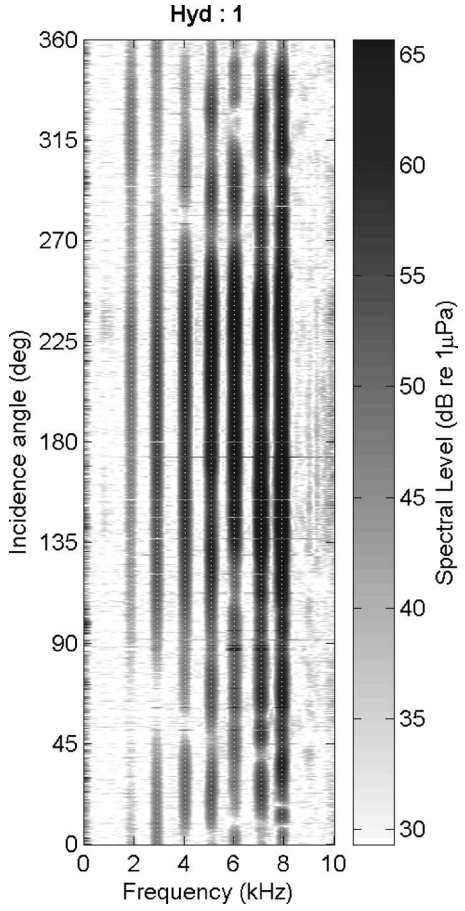


FIG. 5. Mid-frequency comb spectral levels received by hydrophone 1 on the port side of the AUV as a function of angle of incidence over the complete 360° azimuthal interval.

observed reduction in overall signal level with decreasing frequency is consistent with the frequency dependence of the (ITC-1007) transducer transmit voltage response curve.

For comparison with Fig. 5, Fig. 6 shows the received levels versus frequency and incidence angle for receiver 6 situated approximately on the opposite side of the AUV from hydrophone 1. In this case, the location of the angular interval with high received levels for each of the tones is shifted by nearly 180° to that of receiver 1. Again, the incidence angle interval of high received levels for all tones is strongly correlated with the geometric orientation of the source/receiver direct path and the two glass spheres, suggesting that the spheres are the major contributors to the angular variability of the acoustic field. Figures 5 and 6 clearly illustrate the strong frequency and angular dependence of the spectral levels of the individual tones recorded by a single receiver and the significant differences between the received levels of the same tones on two spatially separated receivers.

III. NUMERICAL MODELING

A. Time domain finite difference approach

A numerical modeling effort was carried out using a time-domain computer code based on a staggered grid, pseudo-spectral finite difference scheme implemented in two spatial dimensions CABRILLO (Gerstoft, 2004). The finite dif-

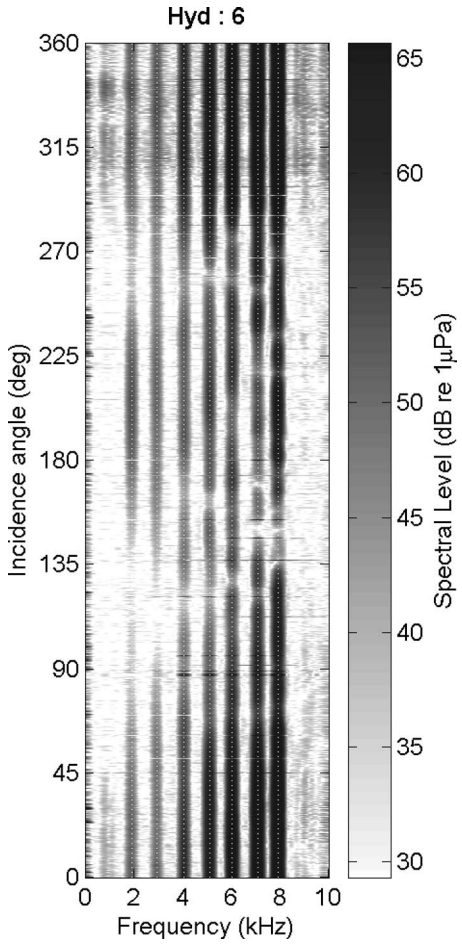


FIG. 6. Mid-frequency comb spectral levels recorded by hydrophone 6 on the AUV's starboard side as a function of angle of incidence in azimuth.

ference method provides the ability to model finite frequency effects in both forward and backward propagation (Fricke, 1993, Tirkas *et al.*, 1993, Stephen, 1996, Wang, 1996, Hastings *et al.*, 1997; and Chen *et al.*, 1998). These full wave field effects are particularly important in describing the complex near-field interference structures observed in the acoustic field near an underwater vehicle.

The CABRILLO code approach to solving the wave equation in the time domain contrasts with most other numerical techniques used in ocean acoustics, such as the Parabolic Equation solution (Collins *et al.*, 1989, 1992, Levy and Zaporozhets, 1998 and Schneider *et al.*, 1998), where the frequency domain Helmholtz wave equation is solved. Calculation of the time-domain solution allows direct comparison with the measured data outlined in Sect. II. CABRILLO is capable of modeling acoustic, elastic, and poro-elastic media. For the purpose of this study, only the surrounding water and the thin shells of the two instrumentation glass spheres were modeled as acoustic media, where the sound speed in glass was set to the value of the glass compressional wave speed. The interior of the glass spheres was modeled as a vacuum with zero sound speed. The 2D modeling approach approximates the two instrumentation glass spheres as fixed cylindrical shells. The use of a staggered grid gives improved numerical accuracy and is better able to handle the large sound speed and density contrasts between the instrumentation spheres of the AUV, the enclosed air, and surrounding water in this problem. In a pseudo-spectral method, the spatial derivatives are solved by multiplication of the wave number in the wave number domain. For a regular staggered grid, the spatial derivative is given by

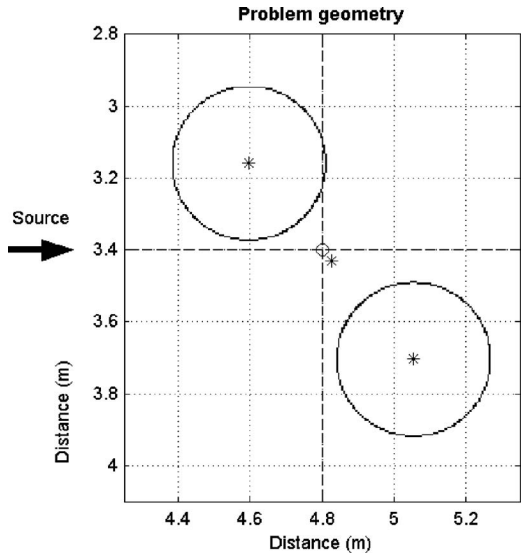


FIG. 7. Geometry for the two-dimensional finite difference modeling of sound scattered from two spherical shells, modeled as cylindrical shells with a vacuum interior, at a 40° incidence angle. Asterisks mark the centers of each of the two spheres and the midway point along the line between these spheres' centers. The small circle in the center of the plot, offset by 4 cm from the midway point between the two spheres, indicates the center of rotation of the AUV during the measurements at TRANSDEC. The upper left large circle in the figure represents the aft instrumentation glass sphere in the AUV and the other circle represents the forward AUV glass sphere.

$$\frac{\partial U(x)}{\partial x} = F^{-1}\left[ik_x e^{\frac{-ik_x h}{2}} F\{U(x)\}\right], \tag{1}$$

where the spatial wave number, k_x, is

$$k_x = 2\pi j/h \text{ for } j = 1, \ldots, N. \tag{2}$$

The quantity h is the grid spacing, N is the number of grid points, i is $\sqrt{-1}$, and F represents the Fourier transform. The pseudo-spectral method provides the highest theoretically possible accuracy for a spatial differentiation (Gerstoft, 2004). Euler integration is then used to solve the time derivative using a conventional finite difference approach. For an acoustic medium

$$\frac{\partial U(x,y,t)}{\partial t} = \frac{1}{2\Delta t}\left[U^{(n+1)}(i,j) - U^{(n-1)}(i,j)\right], \tag{3}$$

where i and j are the spatial indices in directions x and y, respectively, and n is the temporal index.

B. Problem geometry

Figure 7 shows the 2D finite difference grid geometry for the two sphere problem. The outer edges of the two shells are placed 0.280 m apart, corresponding to the separation of the glass instrumentation spheres in the AUV. The TRANSDEC measurement setup was modeled using a single omnidirectional source placed 2.25 m from the center of AUV rotation, marked by the small circle in the center of Fig. 7.

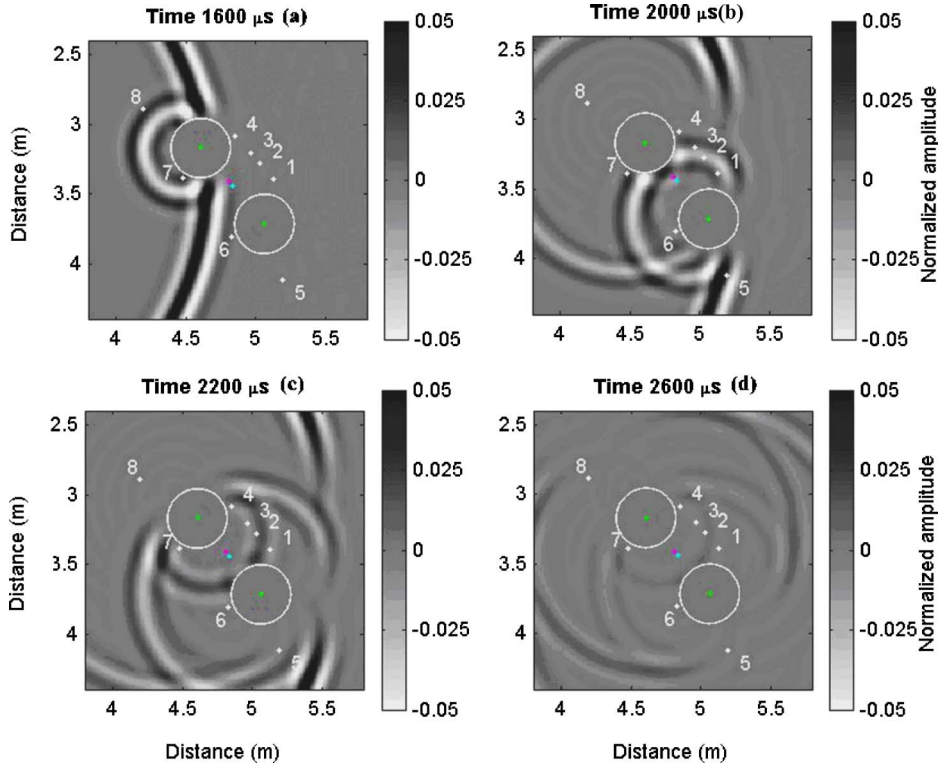


FIG. 8. (Color online) Snapshots of the pressure field amplitude at four different times (listed above each plot) after transmission of a Ricker-type pulse, as calculated by the 2D time-domain finite difference computer code. The point source in each figure is at a range of 2.25 m to the left from the center of each figure and the bearing from the source to the plots' center (incidence angle) is 40°. The amplitudes in all figures have been normalized by the same value.

(The point equidistant between the centers of the two spheres, marked by an asterisk in Fig. 7, is offset 4 cm from the center of rotation). The shell density and P-wave velocity were taken as 2300 kgm^{-3} and 5200 ms^{-1}, respectively, representative of the properties of glass. Standard values for water were used for the surrounding fluid. The interior of the shells was assumed to be a void with zero sound speed. For a shell radius of 0.215 m, the product of the wave number and radius (ka) varies from 0.9 to 7.2 over the frequency band 1–8 kHz. A time step dt of 0.4 μs (equivalent to a data sample frequency of 250 kHz) was used in the modeling and the computations were allowed to advance 15 000 steps corresponding to a 6 ms total propagation time.

C. Results

Figures 8(a)–8(d) show a series of four snapshots of the acoustic pressure field amplitude (both positive and negative values) as a function of the two spatial dimensions for a spherically spreading 8 kHz Ricker (Ricker, 1953) type source pulse incident at 40° to the spheres and at a source range of 2.25 m. At 1600 μs (Fig. 8(a)), the primary wave front interacts with the upper sphere (the aft sphere in the AUV) with a strong reflection back towards the source. As the primary wave front progresses (Fig. 8(b)), diffraction of the primary wave front into the geometrical shadow zone of the upper sphere is evident. In addition, the interaction of the first reflection of the primary wave front from the lower sphere with the first reflection from the upper sphere can be seen around hydrophone 6. Figure 8(c) and 8(d) (the lower two panels) show the development of progressive multiple reflections between the two spheres well after the main wave front has passed. The effects of diffraction also are visible as each reflection passes the opposite sphere. Figure 8(d) shows a fourth-order reflection arrival at the lower sphere just under 1 ms after the initial wave front passes. However, this multiply reflected signal is 40 dB below the primary wave front arrival and therefore is considered insignificant in relation to the earlier arrivals.

The individual pressure field solutions at each tone frequency then were calculated from the time domain solutions for a 1 cm spaced grid over a 2 m $\times$ 2 m area centered on the center of rotation. Figure 9 shows the magnitude of the complex pressure field solution at 7910 Hz for three incidence angles (0°, 50°, and 90°). In the endfire orientation (90°) shown in the lower panel, the interference pattern to the left of both spheres is due to the interaction of the incident signal with the scattered signal from the sphere closest to the source. As the incidence angle is rotated by an increasing amount away from endfire, the effects of scattering from the second sphere become increasingly more significant.

The middle panel (50° incidence) shows the development of a more complex field structure. In particular, the lower left quadrant of this panel shows scattered energy from the left-hand sphere, the backscattered field from the other sphere, and the direct field mutually interfering with one another (Ingenito, 1987, Carrion *et al.*, 1990; and Stephen, 2000). This asymmetric field becomes symmetrical again as the incidence angle decreases to 0° (broadside), as observed in the uppermost panel of Fig. 9.

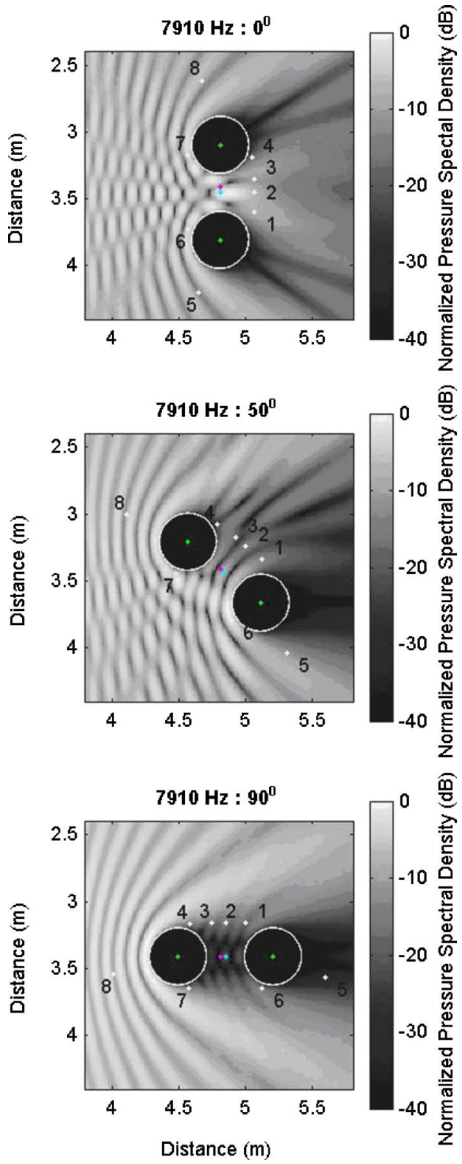


FIG. 9. (Color online) Numerical modeling results for the 2D spatial distribution of the normalized complex pressure field magnitude at 7910 Hz for an incidence angle of 0° (upper plot), 50° (middle plot), and 90° (lower plot).

Diffraction effects result in energy appearing in the geometrical shadow zone of the two spheres. For 90° incidence, diffracted energy in the shadow zone of the left-hand sphere then is scattered from the right-hand sphere, resulting in the weak interference field structure in the region of receivers 1–4 and 6,7. At other geometries, more complex field structures may exist in the spheres' shadow zones due to the interaction of both scattered and diffracted energy. As an example, the shadow zone field structure at an incidence of 50° for the left-hand sphere in Fig. 9 shows considerably more complexity due to scattered energy from the right-hand sphere. A relatively simple shadow zone structure is observed behind the right-hand sphere consistent with standard edge diffraction effects. These results indicate that the multiple-reflected wave fronts, as observed in Fig. 8(d), are weak compared to the diffraction of the primary wave front.

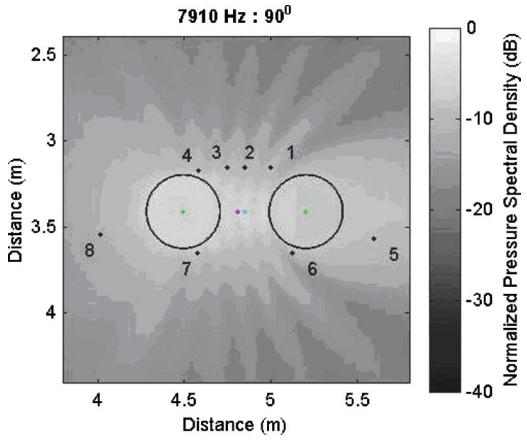


FIG. 10. (Color online) The numerically modeled scattered pressure field magnitude at 7910 Hz and 90° angle of incidence. The scattered component is obtained by subtracting the field calculated without any scatterers present from the total field calculated with scatterers present.

The field structure associated only with the interaction of multiple-reflected and diffracted energy can be seen more clearly in Figs. 10 and 11, where the incident pressure field with no scatterers present is subtracted from the total pressure field with scatterers present. In Figs. 10 and 11, the difference field to the left of the spheres is homogeneous since it is composed only of energy scattered from the spheres. The weaker interference patterns due to interaction between diffraction and multiple reflections between the two spheres however is clearly present, particularly in Fig. 11. At 90° incidence (Fig. 10), the interference structures are quite weak because they arise solely to multiple reflections of energy diffracted from the left-hand sphere. In contrast, a stronger interference field is present in Fig. 11 due the interaction of the primary scattered fields from both spheres.

In summary, Figs. 9–11 illustrate that complex interference field structures can exist in the acoustic field surrounding two glass spheres due to diffraction and multiple scattering. These structures demonstrate a strong spatial and frequency dependence similar to that observed in the measured data from the AUV.

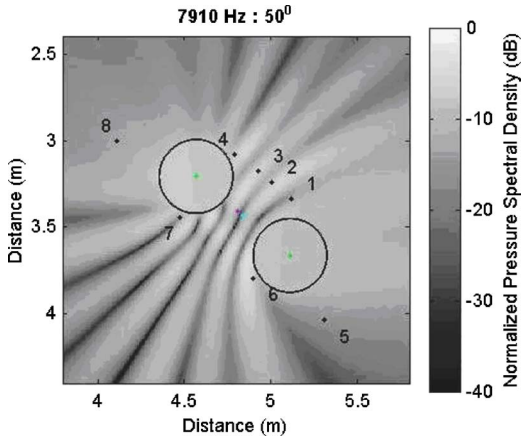


FIG. 11. (Color online) The magnitude of the scattered component of the pressure field at 7910 Hz and 50° angle of incidence.

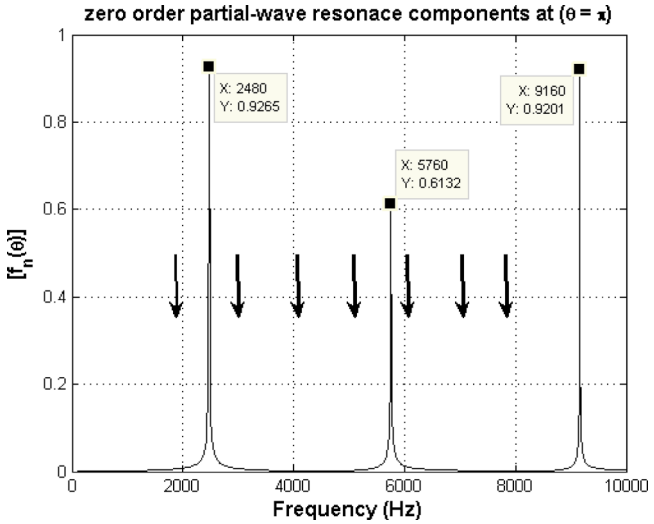


FIG. 12. (Color online) The magnitude of the total field form function (Eq. (4) in the text) for the zeroth mode in the analytical solution for the scattering from an elastic, vacuum-filled, thin-walled spherical shell as a function of frequency. The incident field is a single plane wave, the elastic properties of the shell are those of glass, and the form function is evaluated in the forward-scattered direction (θ=180°).

IV. COMPARISON OF THE MEASUREMENTS AND NUMERICAL MODELING RESULTS

In the 2D finite difference calculations presented in the previous section, the three-dimensional AUV instrumentation glass spheres were modeled as two-dimensional infinite cylinders. The cylinders' shells were modeled as an acoustic medium, with a sound speed equal to the compressional speed in glass. In reality, the small value of the ratio of the shell thickness to outer radius, of order 6.5%, suggests the spheres might be effectively modeled as "acoustically soft bubbles," with no intermediate boundary between the outer fluid and the inner void. On the other hand, if the elastic properties of the shell are considered, the resonance characteristics may be altered appreciably from those of a bubble and so could contribute significantly in a different way to the interference fields surrounding the AUV.

To evaluate the potential effects of the elastic properties of the glass shell on the character of scattering, exact solutions for the scattered pressure from a single thin, elastic, spherical shell enclosing a vacuum were implemented as outlined by (Vesker, 1993). The partial-mode form function as a function of angle θ (the angle with respect to the direction of the incident plane wave signal) is given as

$$f_n(\theta) = \frac{2}{x}(2n+1)\exp(i\delta_n)\sin(\delta_n)P_n(\cos(\theta)), \tag{4}$$

where δ_n is derived from the fifth-order determinants $D_n^{(1)}$ and $D_n^{(2)}$ given by Goodman and Stern, 1962, and $P_n\cos(\theta)$ is the Legendre polynomial of order n. Figure 12 shows the resonance components of the first (zero-order) partial-wave mode for a single AUV instrumentation glass sphere modeled as a 3D thin elastic shell with a vacuum interior. Three resonances exist across the spectrum of interest, at frequencies 2480, 5760, and 9160 Hz. Arrows in Fig. 12 indicate the frequencies of the mid-frequency tones transmitted by the source in this study. These resonances could contribute significantly to the interference field structure around the body of the AUV at source frequencies equal, or close, to the resonance frequencies. Therefore, to achieve a good match between the numerical modeling and data results, the number and frequencies of these resonances must be accurately calculated. The analytical model in Eq. (4) was used to determine the sensitivity of these resonances to changes in the properties of the glass shell. The effect of decreasing the shear wave speed to small values while leaving the compressional speed fixed was to cause a shift in the frequencies of the three resonances downward slightly by a few hundred hertz, suggesting that the elastic properties of the glass play only a minor role. Similarly, setting the compressional and shear wave speeds to that of glass and reducing the thickness of the glass shell to a negligibly small value resulted in a resonance structure identical to that in Fig. 12, but upshifted in frequency by 100–200 Hz. Setting the shear wave speed to that of glass, the glass thickness to its original value, and reducing the compressional wave speed significantly also causes an up-shift in the resonances in the mid-frequency band of interest. In this case, shifts of 1–2 kHz were observed in the mid-frequency band resonance frequencies for a compressional wave speed reduced to around 100 m/s above the shear wave speed. In any case, if a realistic value for the compressional speed is used in the modeling, the properties of the glass shell have only a small effect on the scattering resonances. Note that a depth dependence of the resonance frequencies of these spherical shells has been observed in the ocean, believed to be caused by changes in air temperature (and so sound velocity) inside the spherical shell (D'Spain *et al.*, 1991). Calculations showed that an 8 kHz resonance could be shifted by 50 Hz with an internal temperature variation of 6–2 °C. For this paper, the implementation of a three-dimensional solution for coupled, elastic scattering from two spheres was beyond the scope of the work. However, the effects of elastic scattering may explain some of the observed measured/modeled differences and so should be considered in any future work. Clearly, accounting for the 3D nature of the scattering is important in any future effort. At the least, the results for the two-dimensional case presented in the previous section still are useful in understanding the implications of near-field interference structures in source localization.

Two-dimensional finite difference simulations were run for incidence angles in azimuth over a 360° sector. Comparison then was made between modeled and measured results at all frequencies. Figure 13 shows the normalized level of the 7910 Hz signal for hydrophone 1 as a function of incidence angle. The normalized levels from the TRANSDEC measurements are plotted as small diamonds, the finite difference modeling results are shown as connected circles, and the effective mean noise over all channels during the TRANSDEC tests, normalized by the maximum received level over all azimuths, is represented by the horizontal dashed line. An angular interval with high received levels is observed between 100° and 270°, closely corresponding to the incidence angle interval where hydrophone 1 is exposed directly to the source. Receivers 1–4 on the AUV's port side and 6 and 7 on

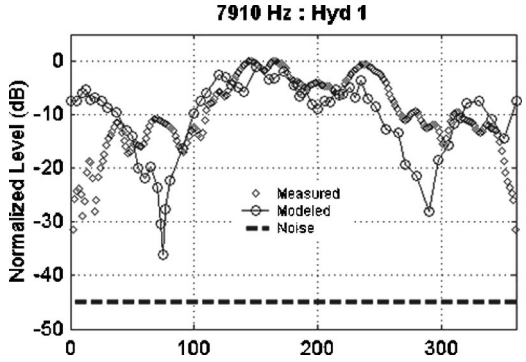


FIG. 13. Comparison of the normalized pressure amplitude at 7910 Hz for hydrophone 1 as a function of incidence angle from the numerical modeling output (connected circles) and from the TRANSDEC measurements (diamonds), The normalized average electronic self-noise level over all hydrophones in the TRANSDEC measurement effort is plotted as a horizontal dashed line.

the starboard side all exhibit similar correlation between the measured data and the model calculations (re Fig. 14 for hydrophone 7) over the angular interval of high received levels. Variations in hydrophone 1's received levels with changes in incidence angle in the angular interval of high received levels are relatively small, with less than a 6 dB variation in level from 110° and 260°. This hydrophone is placed close to the forward sphere in the AUV and so the interference field structure is relatively simple as the incidence angle is varied. A greater degree of complexity with incidence angle, with variations as large as 15 dB, is observed in the modeling results for hydrophone 7 (Fig. 14). This high degree of variability is not present in the measurements. The finite difference model predicts a relatively strong interference pattern between the two spheres (where hydrophone 7 is located) which is a sensitive function of incidence angle. This structure is primarily due to interference of the fields scattered from either sphere and the incident field. The much smoother response observed in the measured data may be due to the much weaker scattering response of a three-dimensional sphere compared to the infinitely long, 2D cylinder used in the modeling (Stanton, 1988 and Stanton *et al.*, 1998). For example, at $ka \gg 1$, a rigid, fixed sphere has an equivalent target strength computed in the back direction (Urick, 1983 and Page *et al.*, 2000) of around −19 dB compared with −6 dB for a rigid, fixed cylinder. For the weaker target, more of the incident energy will propagate (diffract) around the sphere resulting in less well-defined nulls. At higher frequencies (re Fig. 14), the variability in level over the angular interval of high received levels is less in the measurements than in the model output because the field backscattered from the sphere (measured data) is significantly weaker than that predicted by the modeling (Levy and Zaporozhets, 1998). Additional effects contributing to the measurements such as reflections and scattering from the AUV components other than the two spheres also may effectively smooth over the interference structures predicted by the modeling.

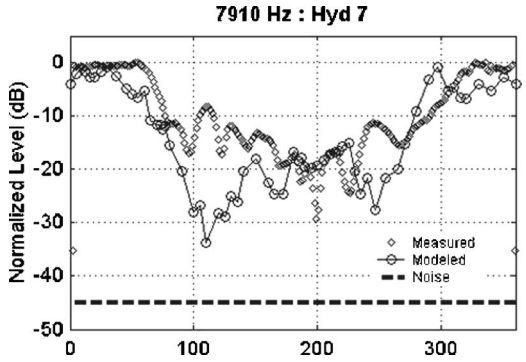


FIG. 14. Comparison of the numerical modeling and measurement results for the normalized amplitude at 7910 Hz as a function of incidence angle as in Fig. 13, but now for hydrophone 7.

As the incidence angle changes so that the direct source/receiver path is blocked by one or both of the instrumentation spheres, both modeling results and measurements for most receivers exhibit a more complex field structure; re Figs. 13 and 14. In a few cases, the measurements in this geometrical shadowing regime show a slightly more complex structure than that predicted by the modeling, possibly because of additional interactions from scattered energy from other parts of the AUV. However, the levels in this regime generally are much lower (peaks are 10–20 dB lower) than the levels in the angular interval where the path between the source and receiver is unimpeded by any scatterers. Therefore, at smaller signal-to-noise ratios, much of the finer detail observed in this angular interval would be lost.

Both measurements and modeling results reveal a reduction in the complexity of the field structure with decreasing frequency, consistent with the increase in the acoustic wavelength. The interference structures become less prominent at the lower frequencies and the match between measurement and prediction improves. These results suggest that at lower mid-frequencies, the two glass spheres are the major contributors to the near-field scattering response at all angles of incidence, and that this scattering is reasonably well modeled by the 2D case. As at higher frequencies, the response with incidence angle at lower frequencies shows strong nulls (sometimes as much as 40 dB below the main peak) in the interference field structure. Again, the nulls observed in the modeling outputs are deeper than those in the measured data due to the differences in scattering response of a 2D infinitely long cylinder and a 3D sphere.

V. SOURCE LOCALIZATION

Often in source localization problems, multiple equally likely source solutions (ambiguities) exist. However, as more independent information is added to the problem (e.g., greater number of receivers, greater spatial diversity, greater number of frequency components), some or all of the ambiguous solutions can be eliminated, leaving a unique (hopefully correct) source position estimate. Matched field processing techniques (Cox *et al.*, 1987 and D'Spain, 1994) were used in this study to compare a series of replica vectors $\boldsymbol{r}(x_i)$ with measured data $\boldsymbol{d}(x_s)$, where x_s represents the desired parameter. In this study, x is the incoming signal incidence angle ($\theta_{i,s}$) in azimuth, assumed to equal source bearing. The replica vectors are a subset of all possible solutions

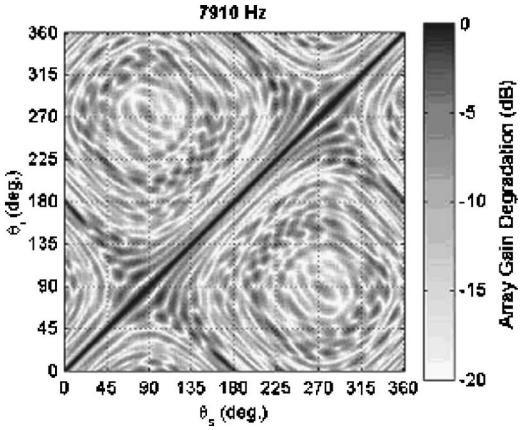


FIG. 15. Array gain degradation (Eq. (5) in the text) for the numerical modeling results at 7910 Hz compared with themselves in the case where no scatterers are present.

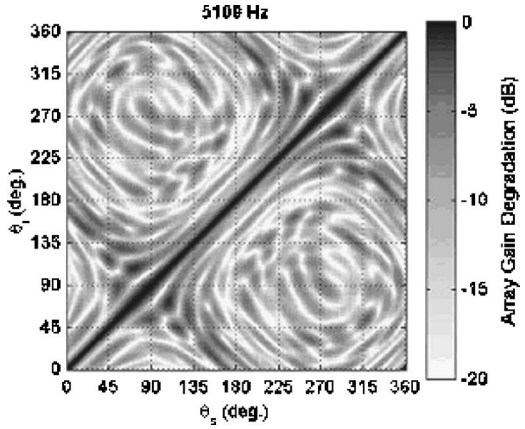


FIG. 16. Array gain degradation for the numerical modeling results compared with themselves in the case with no scatterers present, as in Fig. 15, but now at a frequency of 5100 Hz.

over the full domain of x. For narrowband acoustic signals, both amplitude and phase information on multiple receivers can be used to estimate source bearing. In this section, sets composed of complex vectors (the complex acoustic pressure containing both amplitude and phase information), obtained from both measurements at TRANSDEC and numerical model outputs for all eight elements in the AUV hydrophone array, are quantitatively compared with themselves and with each other (they play the role of both replica and data vectors) to allow assessment of array directivity properties and ambiguity resolution for a particular parameter. In addition to complex vector sets formed from the measurements and numerical modeling described in Secs. II and III, an additional vector set was generated using the output from a numerical model in which no scatterers were present (i.e., the two spheres in Sec. III were removed) to evaluate the case of free-field propagation.

A metric based on the conventional Bartlett processor (Kuperman *et al.*, 1990 and Thode *et al.*, 2000), given in Eq. (5), was used to evaluate the comparisons

$$C(\theta_i,\theta_s)=\frac{|\boldsymbol{r}(\theta_i)\boldsymbol{d}(\theta_s)^*|}{\sqrt{|\boldsymbol{r}(\theta_i)\boldsymbol{d}(\theta_i)^*||\boldsymbol{r}(\theta_s)\boldsymbol{d}(\theta_s)^*|}}, \tag{5}$$

The results then were plotted as ambiguity surfaces on a log scale. The quantity in Eq. (5), which takes on values less than or equal to 0 dB where 0 dB signifies an exact match, is referred to as "array gain degradation" and is a quantitative measure of the difference between the "actual" (data) vector and the calculated (replica) vector. Figure 15 shows the array gain degradation for the case of no scatterers present in the model for a signal frequency of 7910 Hz. These modeled results are matched to themselves so that the 0 dB values along the diagonal from bottom left to top right correspond to matching complex vectors with themselves. However, ambiguities (high side lobes) off this diagonal exist, particularly at incidence angles corresponding to array broadside (near 180° and 360°). Relatively high sidelobe structures are also present around 90° and 270°, corresponding to endfire source/receiver orientations. In particular, for θ_s equal to 80°, a series of narrow sidelobes on either side of the true value (on the diagonal) covering a 55° range from 50° to 105° have peaks within 2 dB of the correct solution. Even worse, the ambiguity ridges at the two broadside positions (180° and 360°) have levels within 0.5 dB at angle offsets of ±180° of the true solution (e.g., at a source bearing of 175°, a sidelobe at 7° is within 0.4 dB of the true result). Sidelobes with less than 5 dB array gain degradation are present across the whole range of incidence angles.

The ambiguity surface corresponding to Fig. 15 (numerical modeling with no scatterers present), but at a signal frequency of 5100 Hz, is presented in Fig. 16. The longer wavelength results in a similar but broader sidelobe structure. High-level sidelobes still are present at both the broadside and endfire orientations. The increase in wavelength eliminates any possible phase ambiguities between receivers since the minimum receiver separation (100 mm) now is smaller than $\lambda/2$, resulting in a simplified sidelobe structure (but broader main lobe). The array gain degradation plots for all seven individual tone frequencies in the no-scatterer modeling case exhibit a similar sidelobe structure, with the major ambiguities at broadside and endfire, and with a gradual shift to a broader main lobe and fewer sidelobes with decreasing frequency. The sidelobe structure around the endfire orientations has a significant frequency dependence (variation in the number and locations of the sidelobes), whereas the ambiguity structure at broadside is relatively insensitive to changes in frequency.

For comparison with the no-scatterer case, Figs. 17 and 18 present the ambiguity surfaces for the 2D numerical model output at 7910 Hz with two cylinders present and for the TRANSDEC measurements, respectively. Both of these ambiguity surfaces have similar overall structures, and also have some significant differences with Fig. 15, suggesting that the effects of scattering and shadowing from the two spheres is a major contributor to the directivity properties of the AUV-mounted array. Particularly noteworthy is the near disappearance of the high-level ambiguity ridges present in Fig. 15 at the broadside positions (180° and 360°). As a specific illustration, in Fig. 17 at 175° a broad sidelobe exists at 255° with an approximate level of −2 dB, and in the measurement case (Fig. 18), broad sidelobes occur at 110° and 230° with −4 dB levels. In cases where some degree of mis-

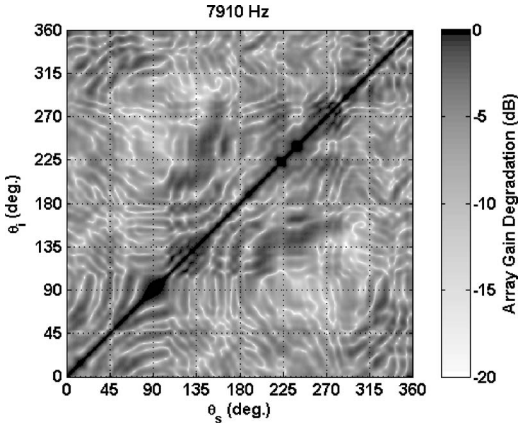


FIG. 17. Array gain degradation for the numerical modeling results at 7910 Hz compared with themselves in the case where two thin-walled cylindrical shells are present.

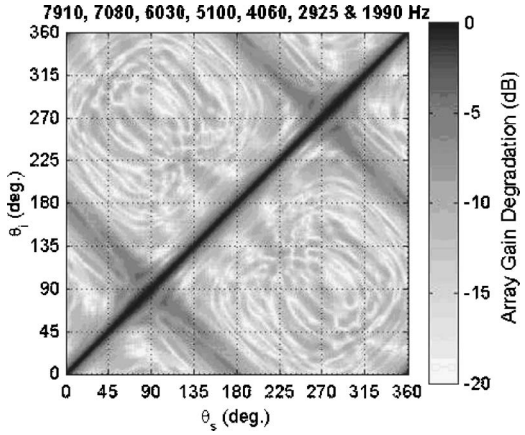


FIG. 19. Array gain degradation averaged over the seven frequencies in the tone comb signal (7910, 7080, 6030, 5100, 4060, 2925, and 1990 Hz) for numerical modeling results compared with themselves in the case of no scatterers present.

match exists, these ambiguities could result in an angular error of 80° corresponding to the −2 dB sidelobe and up to 65° for the −4 dB sidelobe, respectively. In contrast, an error of 180° can occur in the no-scatterer case for incidence angles near broadside (Fig. 15) because of the corresponding −0.5 dB level sidelobes. Similarly, the sidelobe structure around the endfire orientations in the no-scatterer case has a significantly modified appearance in Figs. 17 and 18. In both of these figures, the main-lobe response along the main diagonal broadens. Additional sidelobe structures are present between 0° and 90° and 270° and 360°. These two angular intervals correspond to the orientation of the source/AUV system where the port-side receivers (hydrophones 1–4) are within the geometrical shadow zone of the two glass spheres. This region contains a relatively complex interference field (re Fig. 9) which is primarily due to the interaction of scattered and incident energy and therefore has a strong frequency dependence.

Due to this strong frequency dependence of some of the sidelobe structures, spectral averaging can be used to reduce ambiguities. Figure 19 shows the spectral average for the 7 tones (7910, 7080, 6030, 5100, 4060, 2925, and 1990 Hz) for the no-scatterer case. A general reduction in sidelobe level along with a broadening of the main lobe occurs, resulting in a much clearer definition of the main lobe. For an 82° incidence angle (near endfire), most of the sidelobe levels are decreased to at least 10 dB below the level on the main diagonal. A similar reduction in sidelobe level occurs at 175° (near broadside), although a sidelobe at −5 dB level still is present at 7° which potentially could result in a source bearing estimate error of nearly 180°. This remaining ambiguity appears as the linear gray shading perpendicular to the main diagonal and intersecting it at 90° and 270°.

Spectral averaging of the array gain degradation across the seven tone frequencies also was carried out for both the numerical modeling with two spheres present (Fig. 20) and for the TRANSDEC measurements (Fig. 21). As with the no-scatterer case, a general reduction in sidelobe level occurs because of the strong spectral dependence of the sidelobe structures. In particular, almost all of the sidelobe structures corresponding to incidence angles where hydrophones 1–4 are shadowed by one or both spheres (0°–90° and 270°–360°) and present in the single frequency results (most prominent in Fig. 18) now are reduced to greater than 10 dB below the main-lobe level. In addition, spectral averaging eliminates the ambiguity at endfire orientations that

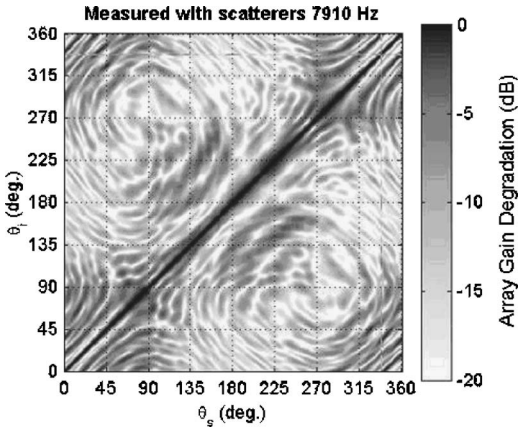


FIG. 18. Array gain degradation for the measurements at TRANSDEC at 7910 Hz compared with themselves.

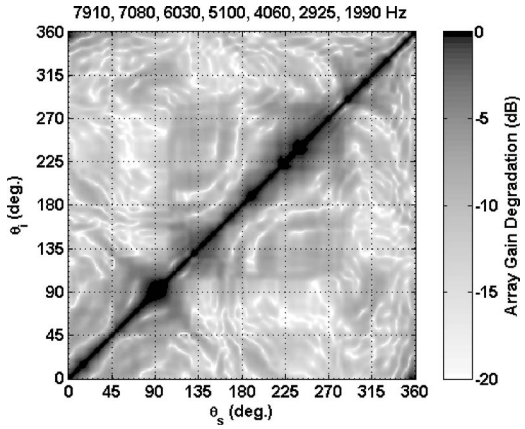


FIG. 20. Array gain degradation averaged over the seven tone frequencies for numerical modeling results compared with themselves in the case of the two scatterers present.

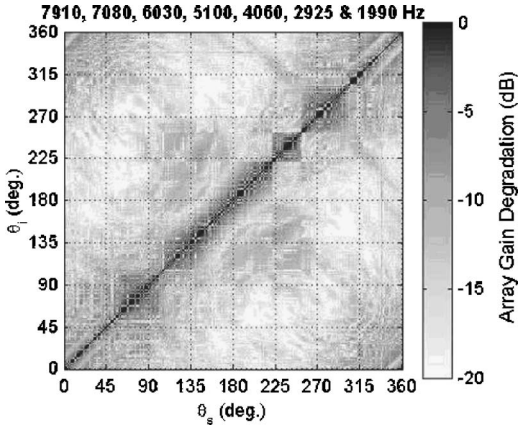


FIG. 21. Array gain degradation averaged over the seven tone frequencies for the measurements at TRANSDEC compared with themselves.

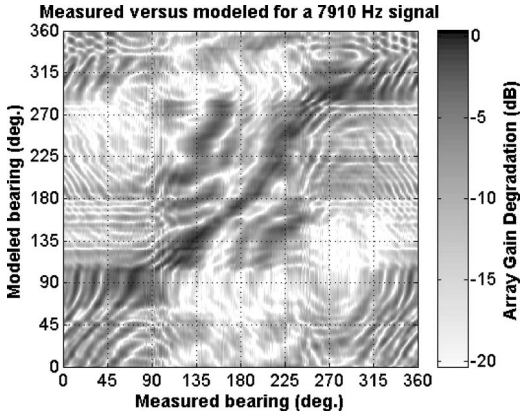


FIG. 22. Array gain degradation at 7910 Hz for the numerical modeling results with scatterers present compared to the TRANSDEC measurements.

exists in the no-scatterer case. In the spectral-averaged no-scatterer case, the maximum sidelobe level is 5 dB below the main-lobe response, which is reduced to around 8 dB below in the modeling-with-scatterers case and 10 dB below in the TRANSDEC measurements. The greater reduction in sidelobe levels for the measurements compared to those for the numerical model output with two spherical shells present is the result of additional frequency-dependent scattering effects by other vehicle components and/or elastic resonance effects of the two glass instrumentation spheres. In general, however, the highest-level sidelobes away from the main diagonal in both Figs. 20 and 21 are approximately 4–5 dB lower than those in the no-scatterer case. Therefore, the presence of the scatterers leads to enhanced array performance for source localization in azimuth.

The focus of this paper is on the effects of scattering from an object such as an AUV on source localization performance of a hydrophone array in the near field of the scattering object. However, to take full advantage of the increase in complexity of the received field due to this scattering, the numerical model must accurately predict this complexity. Figures 13 and 14 and the discussion in Sec. IV suggest that some of the major scattering features observed in the AUV measurements at TRANSDEC can be modeled using a simple two-dimensional numerical model of two spherical shells (actually cylindrical shells). Quantitative comparison of this model and measured data can be made using the array gain degradation in Eq. (5). Figure 22 shows the array gain degradation for the TRANSDEC measurements versus the 2D model results for the 7910 Hz tone. Mismatch now exists as measured by array gain degradation values less than 0 dB along the main diagonal. However, Fig. 22 still demonstrates a relatively strong main-lobe response even with this simplified 2D model. In comparison with the no-scatterer case (Fig. 15), the high-level sidelobe structures around 90° and 270° incidence angles and the high-level ambiguity ridges at the broadside positions (0° and 180°) are significantly reduced. The main lobe in Fig. 22 is broader than in either Fig. 17 or Fig. 18, but its levels still are everywhere higher than those of the associated sidelobes. Some fine sidelobe structure is present between 270°–360°and 0°–90° corresponding with the structures in the measurement-only array gain degradation plot (Fig. 18). The variability of these structures, however, decreases with increasing wavelength as illustrated in Fig. 23 for a measurement versus model comparison at 6030 Hz. The main lobe in Fig. 23 still occurs approximately in the correct location along the main diagonal, indicating a significant degree of correct processor performance. Therefore, enhancement in certain aspects of localization performance still is achieved in this simplified modeling case over that for the no-scatterer case (Figs. 15 and 16) due to suppression of the high-level sidelobe structures.

VI. CONCLUSIONS

The matched field processing results in the previous section, as quantified by array gain degradation, show that scattering from the AUV, if not taken into account, adversely affects the localization performance of the AUV hull-mounted hydrophone array. On the other hand, knowledge of the scattering patterns around the AUV is useful in reducing ambiguities that would exist if the scatterers were not present. Both results from a 2D time domain, finite difference numerical model and measurements from an AUV-mounted eight-element hydrophone array collected in a large calibration tank show complex field structures due to scattering from the AUV. These scattering features are strongly dependent on incidence angle and therefore contain informa-

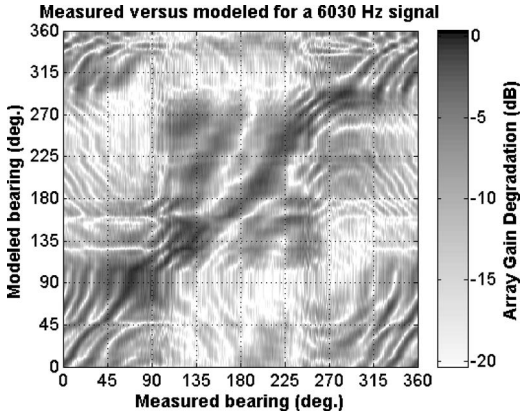


FIG. 23. Array gain degradation at 6030 Hz for the numerical modeling results with scatterers present compared to the TRANSDEC measurements.

tion on the direction to the source. The array gain degradation plots demonstrate that if the scattering features are taken into account, the overall level of the sidelobes decreases and the sidelobes tend to reorient themselves closer to the main lobe, thereby enhancing source localization capabilities. In addition, with the introduction of scatterers, most of the sidelobe structures in the array gain degradation plots become strongly dependent upon frequency so that averaging across frequency further reduces sidelobe levels.

Shadowing effects due to the two instrumentation glass spheres in the AUV appear dominant in generating the structure of the received amplitude as a function of incidence angle. An angular interval with high received levels is observed in the data from all hydrophones and at all frequencies. This angular interval is highly dependent upon the receiver position in relation to the source-scatterer configuration—it occurs when the receiver is located between the source and the two glass spheres—and is relatively frequency independent. Even in low signal-to-noise ratio conditions, this structure is useful in reducing ambiguities in source localization that would exist if the scatterers were not present. Additional received field complexity is evident in the TRANSDEC measurements and is most likely due to additional scattering interactions from other parts of the AUV assembly and elastic scattering effects from the two spheres. This additional complexity shows a strong frequency dependence and therefore any associated sidelobe structure in the array gain degradation plot can be reduced significantly by spectral averaging. In low signal-to-noise ratio situations, much of the information in the field structures at lower level may be lost. However, good resolution on the correct source bearing at all angles of incidence without high-level ambiguities using the received field structure over the angular interval where the receiver was positioned between the source and AUV body still can be obtained at lower signal-to-noise ratio for the band of mid-frequency tones used in this study. A reduction of 5 dB in maximum sidelobe level was observed in comparison with an identical array configuration in an acoustic field without scatterers present.

The numerical modeling results provide valuable insight into the scattering and diffraction effects around the AUV body. One example illustrated by the modeling results is that multiple reflections between the two scatterers are significantly weaker than the diffraction of the incident wave front. Comparison between the numerical modeling results and the measurements shows that a good correlation exists in the locations of the angular intervals with the highest received levels for all hydrophones. This result suggests that much of this structure is dictated by the scattering effects of the two glass spheres. The agreement between the numerical model predictions and the measurements steadily improves with decreasing frequency, probably because the corresponding increase in acoustic wavelength reduces the scattering contributions from AUV components not included in the model. Variations in the levels in the angular interval with high received levels that are present in the modeling results but not in the measured data are most likely due to difference in backscattering strength of the scatterers in the two cases. That is, in the numerical model, each of the two glass spherical shells is modeled as a 2D cylinder with a thin fluid shell and a vacuum interior. As illustration of the impact of modeling a 3D object with a 2D approximation, a rigid, 2D cylinder has a target strength in the backscattered direction 13 dB greater than a rigid, 3D sphere (−6 dB vs −19 dB, respectively) for wave number/radius products much greater than unity ($ka \gg 1$). For ka values around unity, the elastic properties of a spherical glass shell can create scattering resonances that greatly increase the scattering cross section over that of an identically sized rigid sphere. The frequencies of the tones transmitted by the source during the tank measurements do not correspond to the three resonance frequencies of the lowest radial mode of a single vacuum-filled glass shell as calculated by an analytical model; however, they both fall in the same frequency (1–10 kHz) band. Therefore, the resonances must be modeled correctly in order to accurately predict the scattered field. Scattering from additional AUV components that contribute to the measurements but not included in the model also could be a source of discrepancy between the numerical model ouputs and the tank measurements. Future numerical modeling efforts most likely would show significantly improved agreement with the measurements by taking these effects into account. In any case, the modeling results in this paper demonstrate the potential improvement in source localization performance due to increased complexity of the received acoustic field caused by scattering from an object such as an AUV near the receiving array.

ACKNOWLEDGMENTS

Richard Zimmerman and the Deep Tow engineering group at the Marine Physical Laboratory installed the eight-element hydrophone array and digital data acquisition system in the AUV and conducted the measurement effort at TRANSDEC. Dave Ensberg, Marine Physical Lab, and Howard McManus at the Space and Naval Warfare Center, San Diego, also helped make the TRANSDEC measurements. Peter Gerstoft, also at the Marine Physical Lab, provided us with the Cabrillo time domain finite difference code and thorough documentation on how to use it. Thanks to Bryan Woodward, Loughborough University (UK), for assistance in manuscript preparation. This work was supported by the Office of Naval Research, Codes 321(US) and 321(OA).

APPENDIX: EFFECTS OF VARIATION IN ARRAY POSITION

To illustrate the impact of variation in array position, the set of complex pressure vectors obtained from the numerical modeling with the hydrophone array in its original position was matched with the complex pressure vector set obtained after the array position was offset by various amounts and in various directions in the horizontal plane, i.e., in the plane of Fig. 9. (Note that since the shape of the array remains unchanged during an offset, the effect on the phase differences between the elements of the array is minimized, thereby decreasing the role of phase in this sensitivity analysis). Figure 24 shows the array gain degradation for a 6030 Hz signal

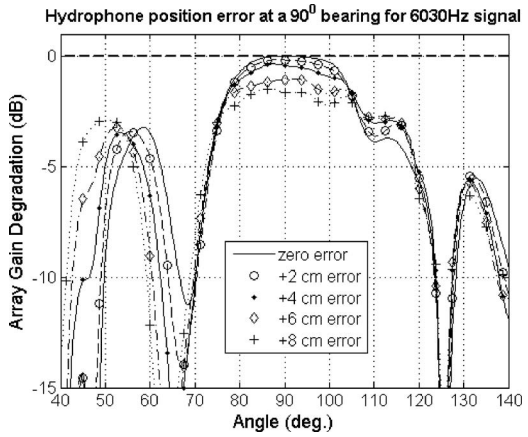


FIG. 24. Array gain degradation as a function of assumed angle of incidence given the true incidence angle is 90° for hydrophone position errors of 0–8 cm in 2 cm increments. The source frequency is 6030 Hz.

arriving at a bearing of 90° (endfire as shown in the lowermost panel in Fig. 9) for array offsets of 0–8 cm in 2 cm increments. All offsets in Fig. 24 were in the direction parallel to the line connecting the centers of the two spheres. The zero-offset plot shows the broad mainlobe structure observed at 90° (re Fig. 17 for 7910 Hz) with no significant sidelobes within 3 dB of the main lobe. For array offsets up to 8 cm, little degradation in the mainlobe structure occurs; the maximum decrease in mainlobe level at 8 cm offset is only 2 dB. A corresponding decrease in the mainlobe to sidelobe ratio occurs, from 3 dB at zero offset to about 1 dB at 8 cm. This significant tolerance to array offset is the result of the fact that very little change in the acoustic field structure occurs in the direction parallel to the line connecting the centers of the two spheres for an incidence angle at endfire, as seen in the lower panel in Fig. 9. For comparison, Fig. 25 shows the array gain degradation for a source incidence angle of 180° (broadside). In this case, significant changes in the acoustic field structure occur for offsets in array location along the direction parallel to the line connecting the two spheres, as illustrated in the upper panel of Fig. 9 (which is for an incidence angle of 0°). The mainlobe to sidelobe ratio in Fig. 25 decreases from around +4 dB for zero offset to around +1 dB at 4 cm offset. At 6 cm, the degradation in the main lobe of nearly 10 dB results in mainlobe levels around 4 dB below the highest sidelobes; at 8 cm offset, a deep null appears at the original location of the main lobe. The acoustic wavelength at 6030 Hz is around 24.8 cm, so that an 8 cm offset corresponds to a third of a wavelength, sufficient to cause significant changes in the field structure recorded by the elements of the array.

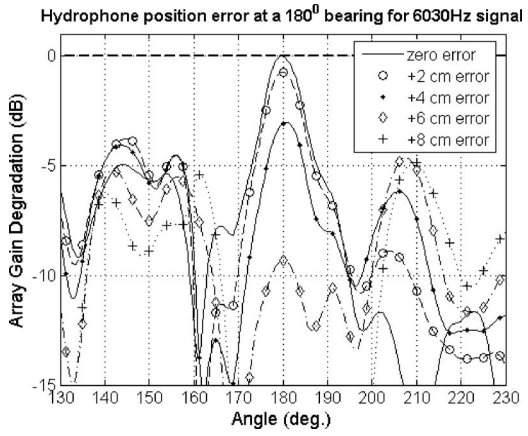


FIG. 25. Array gain degradation as a function of assumed angle of incidence for various element position offsets as in Fig. 24, but now for a true incidence angle of 180°.

These results illustrate that the sensitivity of the matched field processing output to array position offsets is highly dependent upon the variability of the field structure in the direction of the offset.

Blauert, J. P. (**1983**). *Spatial Hearing*, MIT Press, Cambridge, MA.

Butler, R. A., and Balendiuk, K. (**1997**). "Spectral cues utilized in the localization of sound in the median sagittal plane," J. Acoust. Soc. Am. **61**(5), 1264–1269.

Carrion, P. M., and de Brito, M. A. P. (**1990**). "Resonance scattering in waveguides: Acoustic scatterers," J. Acoust. Soc. Am. **87**, 1062–1069.

Chen, Y., Chew, W. C., and Liu, Q. (**1998**). "A three-dimensional finite difference code for the modeling of sonic logging tools," J. Acoust. Soc. Am. **103**, 702–712.

Collins, M. D., and Werby, M. F. (**1989**). "A parabolic equation model for scattering in the ocean," J. Acoust. Soc. Am. **85**, 1895–1902.

Collins, M. D., and Evans, R. B. (**1992**). "A two-way parabolic equation for acoustic backscattering in the ocean," J. Acoust. Soc. Am. **91**, 1357–1368.

Cox, H., Zeskind, R. M., and Owen, M. M. (**1987**). "Robust adaptive beamforming," IEEE Trans. Acoust., Speech, Signal Process. **10**, 1365–1376.

D'Spain, G. L., Hodgkiss, G. L., and Edmonds, G. L. (**1991**). "The simultaneous measurement of imfrasonic acoustic partical velocity and acoustic pressure in the ocean by freely drifting swallow floats," IEEE J. Ocean. Eng. **16**(2), 195–207.

D'Spain, G. L. (**1994**). "Relationship of underwater acoustic intensity measurements to beamforming," Can. Acoust. **22**, 157–158.

Fricke, J. R. (**1993**). "Acoustic scattering from elemental Artic ice features: Numerical modeling results," J. Acoust. Soc. Am. **93**, 1784–1796.

Gerstoft, P. (**2004**). "CABBRILLO 1.0: Acoustic, elastic and poroelastic finite difference modeling," (http://www.mpl.ucsd.edu/people/gerstoft/cabrillo/cabrillo.html), last viewed 17th May (**2007**).

Goodman, R. R., and Stern, R. (**1962**). "Reflection and transmission of sound by elastic spherical shells," J. Acoust. Soc. Am. **34**, 338–244.

Hastings, F. D., Schneider, J. B., and Broschat, S. L. (**1997**). "A finite-difference time-domain solution to scattering from a rough pressure-release surface," J. Acoust. Soc. Am. **102**, 3394–3400.

Ingenito, F. (**1987**). "Scattering from an object in a stratified medium," J. Acoust. Soc. Am. **82**, 2051–2059.

Kuperman, W. A., Collins, M. D., Perkins, J. S., and Davis, N. R. (**1990**). "Optimal time-domain beamforming with simulated annealing including application of *a priori* information," J. Acoust. Soc. Am. **88**, 1802–1810.

Levy, M. F., and Zaporozhets, A. A. (**1998**). "Target scattering calculations with the parabolic equation method," J. Acoust. Soc. Am. **103**, 735–741.

Luebbers, R. J., and Beggs, J. (**1992**). "FDTD analysis and experimental measurements," IEEE Trans. Antennas Propag. **AP-40**, 1403–1407.

Morse, P. M., and Ingrad, K. U., (**1986**). *Theoretical Acoustics*, Princeton University Press, Princeton, NJ, ISBN 0-691-08425-4.

Page, S. J., Brothers, R. J., Murphy, K. M., Elston, G. R., and Bell, J. M. (**2000**). "A hybrid ray-trace/finite difference model for target strength evaluation," in *Proceedings Fifth European Conference on Underwater Acoustics*, edited by M. E. Zakharia, P. Chevret, and P. Dubail, Lyon, France, 15–20.

Ricker, N. (**1953**). "The form and laws of propagation of seismic-wavelets," Geophysics, **18**, 10–40.

Schneider, J. B., Wagner, C. L., and Kruhlak, R. J. (**1998**). "Simple conformal methods for finite-difference time-domain modeling of pressure-release surfaces," J. Acoust. Soc. Am. **104**, 3219–3225.

Skelton, E. A., and James, J. H. (**1997**). *Theoretical Acoustics of Underwater Structures*, Imperial College Press, ISBN 1-86094-085-4.

Stanton, T. K. (**1988**). "Sound scattering by cylinders of finite length fluid cylinders," J. Acoust. Soc. Am. **83**, 55–63.

Stanton, T. K., Wiebe, P. H., and Chu, D. (**1998**). "Differences between sound scattering by weakly scattering spheres and finite-length cylinders

with applications to sound scattering by zooplankton," J. Acoust. Soc. Am. **103**, 254–264.
Stephen, R. A. (**1996**). "Modeling sea surface scattering by the time-domain finite-difference method," J. Acoust. Soc. Am. **100**, 2070–2078.
Stephen, R. A. (**2000**). "Optimum and standard beam widths for numerical modeling of interface scattering problems," J. Acoust. Soc. Am. **107**, 1095–1102.
Tirkas, P. A., Balanis, C. A., Purchine, M. P., and Barber, G. C. (**1993**). "Finite difference time domain method for electromagmetic radiation, interference, and interactions with complex structures," IEEE Trans. Electromagn. Compat. EMC- **35**, 192–203.
Thode, A. M., Kuperman, W. A., D'Spain, G. L., and Hodgkiss, W. S. (**2000**). "Localization using Bartlett matched-field processor sidelobes," J. Acoust. Soc. Am. **107**(1), 278–286.
Urick, R. J. (**1983**). *Principles of Underwater Sound* (McGraw–Hill, New York, **1983**).
Vesker, N. D. (**1993**). *Resonace Acoustic Spectroscopy* (Springer-Verlag, Berlin).
Wang, S. (**1996**). "Finite-difference time-domain approach to underwater acoustic scattering problems," J. Acoust. Soc. Am. **99**, 1924–1931.
Yee, K. S. (**1966**). "Numerical solution of initial boundary value problems involving Maxwell's equation in isotropic media," IEEE Trans. Antennas Propag. **AP-14**, 302–307.
Zimmerman, R., D'Spain, G. L., and Chadwell, C. D. (**2005**). "Decreasing the radiated acoustic and vibration noise of a mid-size AUV," IEEE J. Ocean. Eng. **30**(1), 179–187.

Finite-element analysis of middle-ear pressure effects on static and dynamic behavior of human ear

Xuelin Wang, Tao Cheng, and Rong Z. Gan[a)]
School of Aerospace & Mechanical Engineering and Bioengineering Center, University of Oklahoma, Norman, Oklahoma 73019

(Received 21 December 2006; revised 16 May 2007; accepted 18 May 2007)

A finite-element analysis for static behavior of middle ear under variation of the middle-ear pressure was conducted in a 3D model of human ear by combining the hyperelastic Mooney-Rivlin material model and geometry nonlinearity. An empirical formula was then developed to calculate material parameters of the middle-ear soft tissues as the stress-dependent elastic modulus relative to the middle-ear pressure. Dynamic behavior of the middle ear in response to sound pressure in the ear canal was predicted under various positive and negative middle-ear pressures. The results from static analysis indicate that a positive middle ear pressure produces the static displacements of the tympanic membrane (TM) and footplate more than a negative pressure. The dynamic analysis shows that the reductions of the TM and footplate vibration magnitudes under positive middle-ear pressure are mainly determined by stress dependence of elastic modulus. The reduction of the TM and footplate vibrations under negative pressure was caused by both the geometry changes of middle-ear structures and the stress dependence of elastic modulus. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2749417]

PACS number(s): 43.64.Bt, 43.64.Ha [BLM] Pages: 906–917

I. INTRODUCTION

The middle-ear pressure is affected by the ambient pressure, the gas exchange to and from the middle ear and mastoid cavity, and the partial pressure of the gases present. When the middle-ear pressure is changed relative to atmospheric pressure, the tympanic membrane is deformed and hearing ability is altered.

In order to understand the behavior of the middle ear under static pressure load, the effect of middle-ear pressure on hearing has been one of the research interests in middle-ear mechanics. Hüttenbrink (1988) investigated the behavior of the ossicular chain with its typically gliding ossicular joints at variations of static air pressure. Deformation of the tympanic membrane (TM) induced by positive and negative static pressures in the middle ear was measured by Dirckx and Decraemer (1991,1992) using phase shift moiré topography on human temporal bones. Gaihede (1999) investigated the pressure-volume relationship of the middle ear and found that the middle-ear system exhibited hysteresis and nonlinear mechanical behavior in response to large pressure load.

To detect the effect of middle-ear pressure on mobility of the TM, Lee and Rosowski (2001) and Rosowski and Lee (2002) reported their measurements of the TM vibration at different middle-ear pressure in gerbil ears. Their results showed that acoustic stiffness and inertance of both pars tensa and pars flaccida of the TM were altered by static pressure. Murakami *et al.* (1997) measured vibrations of the umbo and stapes head in human temporal bones using video system and suggested that stapes vibration was affected by the TM stiffness and the stiffness of the annular ligament and ossicular chain, while the umbo vibration was primarily affected by the TM performance. Recently, Gan *et al.* (2006a) reported the movements of the TM and stapes footplate under positive and negative middle-ear pressures in temporal bones, and found that the positive and negative pressures might have different effects on middle-ear transfer function.

In addition to experimental measurements, theoretical methods such as finite-element modeling have been employed to simulate sound transmission through the middle ear (Funnell *et al.*, 1987; Prendergast *et al.*, 1999; Koike *et al.*, 2002; Gan *et al.*, 2004, 2006b). However, most of the work was dedicated to acoustic transmission without pressure difference across the TM. There are only a few studies including static pressure across the TM such as the recent publication by Ladak *et al.* (2006), which reported the effect of geometric nonlinearity on movement of the cat eardrum. Ladak *et al.*'s model did not include middle-ear structures and cochlea, and the manubrium was assumed to be completely immobile along its length.

The finite-element implementation for middle-ear transfer function under various static pressures has not been reported. Although it is empirically known that changes of the TM shape may affect the sound transmission of the middle ear, the degree of geometry effect is not clarified because there are numerous uncertain factors involved in the mechanical process of real ear. Therefore, a numerical or finite-element analysis becomes the choice of tool in assessing the deformations of the TM and other middle-ear components under static pressure.

This paper reports the finite-element (FE) analysis of human middle ear under various static middle-ear pressures. The static behavior of the middle ear in response to pressure variation was first derived by introducing material and geometry nonlinearities. The static deformation field associated

[a)]Electronic e-mail: rgan@ou.edu

with middle-ear pressure then provided nodal displacements of the TM and middle-ear ligaments which were used to update the model mesh for dynamic analysis. Change of material properties induced by static stresses in the middle-ear components were expressed as stress-dependent elastic modulus for dynamic analysis. Therefore, the effects of middle-ear pressure on sound transmission were introduced through the mechanical responses of geometry and material properties changes. The TM and stapes footplate motions under various middle-ear pressures were predicted, and the FE model-derived results were compared with published experimental data measured from human temporal bones.

II. FINITE-ELEMENT MODEL

A. Model construction

A three-dimensional FE model of human left ear was established based on 780 histological sections from a left ear temporal bone by Gan *et al.* (2004). This model consists of the external ear canal, TM, middle-ear ossicles (malleus, incus, and stapes), middle-ear suspensory ligaments/muscle tendons, and middle-ear cavity, and has accurate anatomic structure of the middle ear. To simulate the static deformation of the middle-ear soft tissues, the model used for this study consists of the TM, three ossicular bones, six middle-ear suspensory ligaments and muscles tendons, and the stapedial annular ligament; the ear canal and middle-ear cavity were not taken into consideration. Those middle-ear components were meshed by four-noded tetrahedral solid elements which are the same as those published by Gan *et al.* (2004).

When large middle-ear pressure is applied, the change of the TM shape and orientation of the ossicular chain will affect dynamic behavior of the middle-ear system. In this study, the geometries of the FE model for vibration analysis were updated based on static deformations of the TM and ligaments in response to middle-ear pressure variations.

The effect of cochlear fluid on acoustic-mechanical transmission was modeled as a mass block with ten dashpots attached between the stapes footplate and fixed bony wall, the same as that reported by Gan *et al.* (2004). The damping coefficient of each dashpot was assumed to be 0.02 Ns/m so that the impedance of the cochlea becomes equivalent to 20 GΩ (Gan *et al.*, 2004). The static FE analysis did not include cochlear loading beyond the stapedial annular ligament because the cochlea had no effect on the system response to static middle-ear pressure (Dirckx and Decraemer, 2001).

B. Constitutive model for static deformation

1. Material model for TM and ligaments

Material properties of the TM and other middle-ear tissues such as ligaments and muscle tendons have been measured in our previous studies by using uniaxial tensile tests (Cheng, 2007; Cheng *et al.*, 2007; Cheng and Gan, 2007a, b). Figure 1 shows the stress-stretch curves of the TM, stapedial tendon (C5), tensor tympani tendon (C7), and anterior malleal ligament (C4) obtained from the experimental measurements. All the curves in this figure represent the mean data obtained from 11 specimens of the TM, 12 specimens of stapedial tendon, 11 specimens of tensor tympani tendon, and 10 of the anterior malleal ligament. The uniaxial material properties of the TM and middle-ear tissues provide the basis for further analysis of material properties, especially the TM under static pressure.

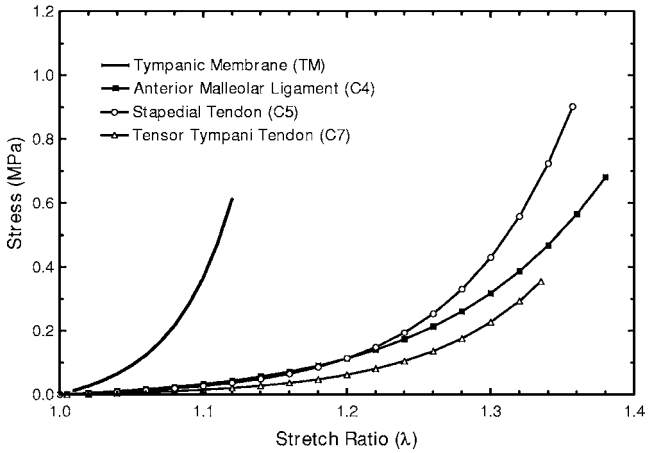


FIG. 1. Stress-stretch curves of tympanic membrane (TM), stapedial tendon (C5), tensor tympani tendon (C7), and anterior malleal ligament (C4) measured from the uniaxial tensile tests in our laboratory.

The TM is a multilayer membrane material and there are no published data on mechanical properties of the TM along the radial and circumferential fiber directions. Thus, the TM, ligaments, and tendons were assumed to be isotropic and homogeneous in this study. A five-parameter hyperelastic Mooney-Rivlin model (Mooney, 1940; Ogden, 1984) for incompressible material was utilized to represent the 3D constitutive relations of the TM and ligaments with the following strain energy function:

$$W = \sum_{i+j=1}^{2} c_{ij}(I_1 - 3)^i (I_2 - 3)^j, \tag{1}$$

where I_1 and I_2 are the first and second invariants of the right Cauchy deformation tensor (Ogden, 1984), $c_{ij}(i+j=1,2)$ are material constants representing five parameters: c_{10}, c_{01}, c_{20}, c_{02}, and c_{11}. Note that the Mooney-Rivlin model is valid only for quasistatic analysis in the present study.

The hyperelastic curve fitting tool in ANSYS (ANSYS Inc., Canonsburg, PA) was used to educe five parameters of the Mooney-Rivlin material model based on the data shown in Fig. 1. The calculated Mooney-Rivlin constants for the TM, C4, C5, and C7 are listed in Table I. The nonlinear material properties of other ligaments such as the superior (C1), lateral mallear (C2), and posterior incudal (C3) ligaments are unknown. In this study, we used the constitutive model of the stapedial tendon (C5) to describe mechanical properties for these ligaments.

There is also a lack of data on the stapedial annular ligament (SAL) in the literature. Price and Kalb (1991) used the spring suspension to fit data from Lynch *et al.* (1982) on static and dynamic stiffness for the cat's stapes, and the model limited maximum displacement to about 20 μm (Pascal and Bourgeade, 1998). Considering the geometry and di-

TABLE I. Hyperelastic material constants of the TM and ligaments.

	c_{10} (MPa)	c_{01} (MPa)	c_{20} (MPa)	c_{11} (MPa)	c_{02} (MPa)
Tympanic membrane (TM)	0.4196	−0.2135	1357.8	−2843.5	1496.7
Anterior mallear ligament (C4)	0.0123	0.0286	12.793	−28.476	16.302
Posterior stapedial tendon (C5)	−0.0524	0.0823	28.033	−62.039	34.864
Tensor tympani tendon (C7)	−0.0071	0.0254	14.059	−30.933	17.297
Stapedial annular ligament (SAL)	−0.1085	0.2111	85.037	−1796.3	953.51

mensions of the stapedial annular ligament in our FE model, we used Price and Kalb's data and derived the Mooney-Rivlin material constant of SAL in Table I.

To date, no experimental measurements of mechanical properties of the pars flaccida, malleoincudal joint (MI-joint), and incudostapedial joint (IS-joint) have been reported. These three components were modeled as linear elastic materials and the pars flaccida was assumed with a Young's modulus of 1 MPa (Ladak *et al.*, 2006). Hüttenbrink (1988) reported that ankylosis of the MI-joint resulted in a reduction of the in- and out movement of the malleus by 50%. In the present model, the static modulus of the MI-joint was determined by a comparison between the malleus displacement obtained from a rigid MI-joint (e.g., the same Young's modulus as the bone) and that from a flexible joint with a variable modulus. When the Young's modulus of the MI-joint was 1.2 MPa, the malleus displacement from the rigid MI-joint was 50% of that from flexible joint under positive pressure, and 46% under negative pressure. Thus, the Young's modulus of the MI-joint was assumed to be 1.2 MPa for static analysis. The Young's modulus of the IS-joint was assumed to be 1.2 MPa as well.

2. Variation of elastic modulus with stress

The stress-stretch curves in Fig. 1 and our previous studies indicated that the TM and middle-ear ligaments/tendons showed nonlinear and viscoelastic properties. To derive the elastic modulus of the tissue for static deformation, the elastic modulus $d\sigma/d\lambda$ was expressed as

$$\frac{d\sigma}{d\lambda} = \alpha(\sigma + \beta), \tag{2}$$

where σ is the normal stress, λ, is the extension or stretch ratio, and α and β are two constants (Fung, 1993). Using this method, the elastic modulus-stress relationships of the TM (pars tensa) over three stress ranges were reported by Cheng *et al.* (2007) as

$$\frac{d\sigma}{d\lambda} = \begin{cases} 32.16\sigma + 0.398 & (0 < \sigma \le 0.1 \text{ MPa}), \\ 29.75\sigma + 0.645 & (0.1 < \sigma \le 0.3 \text{ MPa}), \\ 17.65\sigma + 4.275 & (\sigma > 0.3 \text{ MPa}). \end{cases} \tag{3}$$

Similarly, the elastic modulus-stress relationship of the middle-ear ligaments and tendons such as anterior malleal ligament, stapedial tendon, and tensor tympani tendon was derived from the stress-strain curves in Fig. 1 as follows:

for anterior malleal ligament,

$$\frac{d\sigma}{d\lambda} = \begin{cases} 13.32\sigma + 0.12 & (0 < \sigma \le 0.1 \text{ MPa}), \\ 12.34\sigma + 0.31 & (0.1 < \sigma \le 1.0 \text{ MPa}), \end{cases} \tag{4a}$$

for stapedial tendon,

$$\frac{d\sigma}{d\lambda} = \begin{cases} 23.03\sigma + 0.03 & (0 < \sigma \le 0.1 \text{ MPa}), \\ 10.52\sigma + 1.33 & (0.1 < \sigma \le 1.5 \text{ MPa}), \end{cases} \tag{4b}$$

and for tensor tympani tendon,

$$\frac{d\sigma}{d\lambda} = \begin{cases} 12.89\sigma + 0.14 & (0 < \sigma \le 0.1 \text{ MPa}), \\ 12.00\sigma + 0.31 & (0.1 < \sigma \le 0.7 \text{ MPa}). \end{cases} \tag{4c}$$

Equations (3) and (4) show the variation of elastic modulus of middle-ear tissues with static stress level obtained from experimental curves.

3. Mechanical parameters for dynamic analysis

Rather little is known about dynamic properties of the TM and middle-ear ligaments/tendons. Dynamic properties of the soft tissue vary with the deformation type and the value of amplitude, and can also be frequency-dependent (Bonifasi-Lista, *et al.*, 2005). In most FE models of the middle ear, the Young's moduli of the middle-ear tissues were assumed to be constants and frequency independent (Funnell *et al.*, 1987; Prendergast *et al.*, 1999; Koike *et al.*, 2002; and Gan *et al.*, 2004). Based on the collagen fiber orientations, the TM is usually considered as orthotropic material with different elastic modulus in radial and circumferential directions for dynamic analysis of the FE model. A constant ratio between the radial and circumferential modulus of 35:20 was reported by Gan *et al.* (2004) when zero static pressure was across the TM and the effects of acoustic pressure on mechanical properties were not taken into account.

To define mechanical properties of the middle-ear tissues for dynamic analysis in response to sound stimulus in the ear canal under variable static middle-ear pressures, the TM was considered as an orthotropic material with different modulus in radial and circumferential directions in our FE model. It is also assumed that the nonlinear behavior of static deformation in response to static pressure equally affects the orthotropic moduli in the radial and circumferential directions.

Equations (3) and (4) enable us to evaluate the elastic moduli of middle-ear soft tissues under different static stress levels. Studies on pressure-elastic modulus relationship have shown the increase in elastic modulus with increasing pres-

sure in the aorta (Li, 1987; Buss *et al.*, 1981) and in medial collateral ligament (Bonifasi-Lista *et al.*, 2005). Considering the difference of stress dependence in different soft tissues, and regarding the model of stress-dependent modulus in porous medium (Escuder *et al.*, 2005), an empirical formula converting the static stress-dependent modulus into the elastic modulus E_d for dynamic analysis is derived as

$$E_d = kE_{d0}\left(\frac{E_p}{E_0}\right)^r, \tag{5}$$

where k and r are correction factors, E_{d0} is the elastic modulus for dynamic analysis at zero static pressure across the TM, E_p is the elastic modulus corresponding to a specified stress level, which is relative to static pressure p and calculated from Eq. (3) or (4) as $d\sigma/d\lambda$, and E_0 is the Young's modulus of the TM or ligaments at a reference strain, which was assumed as nominal strain (ε) of 1% in this study. The constants k and r were determined numerically by comparison between the displacement changes at the umbo and stapes footplate obtained from the FE model and that of the temporal bones under different middle-ear pressure. In this study, the calibration of k and r was performed at middle-ear pressure of 0.5 and 1.0 kPa, and the calculated displacements at the umbo and footplate were compared with the experimental data reported by Gan *et al.* (2006a). As a result, the constant k and r were equal to 0.7 and 0.8 for the TM and stapedial annular ligament, and 0.8 and 0.9 for the middle-ear tendons and ligaments, respectively. Regarding the application of Eq. (5), if $E_d < E_{d0}$, the dynamic elastic modulus is E_{d0}; otherwise the modulus is E_d. The E_d induced from Eq. (5) is used as the radial modulus.

In summary, the dynamic elastic moduli of the TM and ligaments were derived as follows: The calculated stress was first extracted at a different pressure level from the results of static nonlinear FE analysis; the stress-dependent modulus was then derived by Eq. (5); and the modulus E_d was finally employed for dynamic analysis in response to sound stimulus in the ear canal and under variable middle-ear static pressure. Note that the completion of dynamic FE analysis to include nonlinear mechanical properties of soft tissues in response to static air pressure in the middle ear will rely on measurements of the TM mechanical properties along radial and circumferential directions as well as the dynamic properties of the TM and middle-ear tissues.

4. Solution procedure

The nonlinear FE analysis was first conducted to determine the static deformation and stress distribution of middle ear components under middle-ear pressure. The overall equilibrium equations for structural static analysis are

$$\boldsymbol{Ku} = \boldsymbol{P}, \tag{6}$$

where $\boldsymbol{K}$ is stiffness matrix, $\boldsymbol{u}$ is nodal displacement vector, and $\boldsymbol{P}$ is nodal load vector. When geometry and material nonlinearities are included, the matrix $\boldsymbol{K}$ is a function of the nodal displacements and their derivatives, and changes with load step. Thus, the iterative process was performed to calculate displacement $\boldsymbol{u}$.

Once the displacement field $\boldsymbol{u}_p$, associated at middle-ear pressure p, was obtained from Eq. (6), the FE model was modified for dynamic analysis. The updated nodal coordinate vector $\boldsymbol{x}$ was described as

$$\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{u}_p, \tag{7}$$

where $\boldsymbol{x}_0$ is the initial nodal coordinate vector used for static analysis, and $\boldsymbol{u}_p$ is nodal displacement vector obtained from Eq. (6) in response to middle-ear pressure p. Therefore, the FE mesh for vibration analysis can be generated based on the nodal coordinate vector represented in Eq. (7).

Based on the updated FE meshes and the induced material parameters by stress evaluation of middle-ear components, the vibrations of the middle-ear structure are described as

$$\boldsymbol{M}_p\ddot{\boldsymbol{v}} + \boldsymbol{C}_p\dot{\boldsymbol{v}} + \boldsymbol{K}_p\boldsymbol{v} = \boldsymbol{F}, \tag{8}$$

where $\boldsymbol{M}_p$, $\boldsymbol{C}_p$, and $\boldsymbol{K}_p$ are the mass, damping, and stiffness matrices at the middle-ear pressure p, respectively, $\boldsymbol{v}$ is the displacement vector for structural vibration, and $\boldsymbol{F}$ is the acoustic pressure load vector applied on the surface of the TM in the ear canal side. If the Rayleigh damping is adopted, the system damping matrix $\boldsymbol{C}_p$ can be expressed by $\boldsymbol{C}_p = \alpha_1\boldsymbol{M}_p + \beta_1\boldsymbol{K}_p$, in which α_1 and β_1 are the damping parameters. Thus, the elastic modulus E_d obtained from Eq. (5) can be introduced into stiffness matrix $\boldsymbol{K}_p$ in Eq. (8).

In this study, we assumed the parameter α_1 was 0; the parameter β_1 was determined using the FE model cross-calibration process. In order to find the best fitting for the damping parameter β_1, we compared the displacement change curves obtained from FE models and those from the measurements (Gan *et al.*, 2006a) under different middle-ear pressure levels. Thus, the damping parameters β_1 associated with each pressure level were determined to be 0.6×10^{-4}, 0.5×10^{-4}, 0.4×10^{-4}, and 0.3×10^{-4} at positive pressure of 0.5, 1.0, 1.5, and 2.0 kPa; and 0.5×10^{-4}, 0.4×10^{-4}, 0.3×10^{-4}, and 0.25×10^{-4} at negative pressure of −0.5, −1.0, −1.5, and −2.0 kPa, respectively.

III. RESULTS

A. Static analysis

The stress distributions in the TM and middle-ear ligaments/tendons were calculated when middle-ear pressure was varied from −2.0 to 2.0 kPa. von Mises stress was used as a scalar measure of the stress state; as an example, Fig. 2 shows the von Mises stress distribution in the TM at the middle-ear pressure of 2.0 kPa [Fig. 2(a)] and −2.0 kPa [Fig. 2(b)], respectively. At positive middle-ear pressure of 2.0 kPa, von Mises stress of the TM in the ear cavity side was mostly less than 0.125 MPa, at negative pressure of −2.0 kPa, the stress distribution for most part of the TM ranges from 0.065 to 0.185 MPa. Comparing stress distribution in the TM under different pressure directions, we found that the negative pressures produced higher stress than that of the positive pressures. However, all the ligaments and tendons had a higher stress at a positive pressure than that of a negative pressure of the same value. Particularly, the

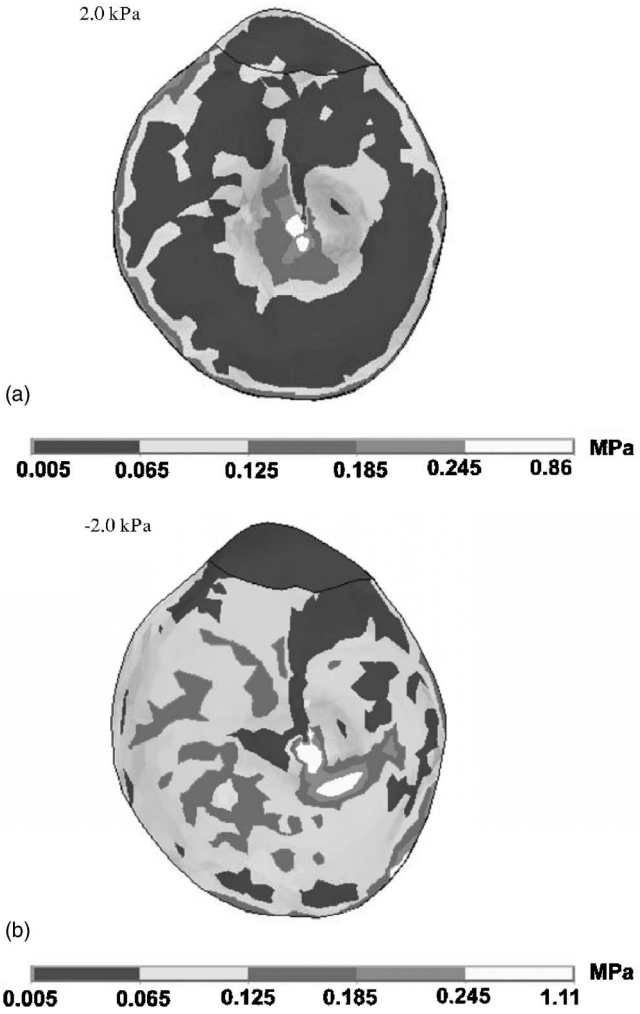


FIG. 2. Distribution of von Mises stress of the TM in the ear canal side. (a) The TM stress distribution for middle-ear pressure of 2.0 kPa; (b) the TM stress distribution for middle-ear pressure of −2.0 kPa.

stresses in the superior ligament and tensor tympani tendon under positive pressures had a larger increase than other ligaments.

Figure 3 shows stress-dependent dynamic elastic modulus E_d of the TM and ligaments obtained from Eq. (5) when middle-ear pressure was varied from −2.0 to 2.0 kPa. It can be seen in Fig. 3 that the values of E_d for the TM and superior ligaments (C1) have a large increase as the pressure varied from 0 to −2.0 or +2.0 kPa, compared with other ligaments. The change of E_d with pressure variation is not symmetric along positive and negative directions. The E_d of the TM at −2.0 kPa is 14% larger than that at 2.0 kPa. However, the modulus E_d of C1 at 2.0 kPa is 100% larger than that at −2.0 kPa. The E_d of C3 and C7 is 54% and 92% larger than that at −2.0 kPa, respectively. This indicates that the increase of E_d with pressure increasing is faster in positive pressure than that in negative pressure for the TM, C1, C3, and C7.

Figure 4 shows static displacements of the umbo and stapes footplate obtained from the FE model when middle-ear pressure was varied from −2.0 to 2.0 kPa in comparison with the published data by Hüttenbrink (1988) and Murakami *et al.* (1997). When positive middle-ear pressures were applied, the outward umbo displacement was 209 μm

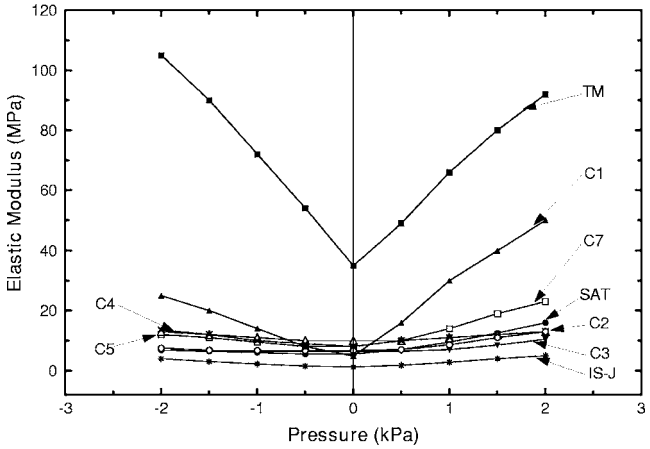


FIG. 3. Variation of elastic modulus with middle-ear pressure for dynamic analysis. Elastic moduli were calculated from Eq. (5). The lines represent tympanic membrane (TM), superior malleal ligament (C1), lateral mallear ligament (C2), posterior incudal ligament (C3), anterior malleal ligament (C4), stapedial tendon (C5), tensor tympani tendon (C7), stapedial annular ligament (SAL), and incudostapedial joint (IS-J).

at 1.0 kPa of middle-ear pressure and 330 μm at 2.0 kPa. Under negative middle-ear pressures, the inward umbo displacement was 140 μm at −1.0 kPa, 202 μm at −2.0 kPa. The displacement of stapes footplate was 18.6 and 26.0 μm at pressure of 1.0 and 2.0 kPa; 13.7 and 20.0 μm at pressure of −1.0 kPa and −2.0 kPa, respectively. The displacement of the umbo was 10 to 14 times greater than that of the stapes footplate. The results in Fig. 4 reveal the umbo displacement for a given positive middle-ear pressure was much larger that for a negative pressure of the same value.

Figure 5 displays the FE model-predicted contours of the TM displacement at static pressure of 2.0 kPa [Fig. 5(a)] and −2.0 kPa [Fig. 5(b)] in medial view. Figure 5(a) shows that the displacement in the posterior and anterior part reached maximum values of 765 and 449 μm, respectively. The displacement in the region of the manubrium was the smaller. Under the pressure of −2.0 kPa [Fig. 5(b)], two displacement maxima were observed in the anterior region: one

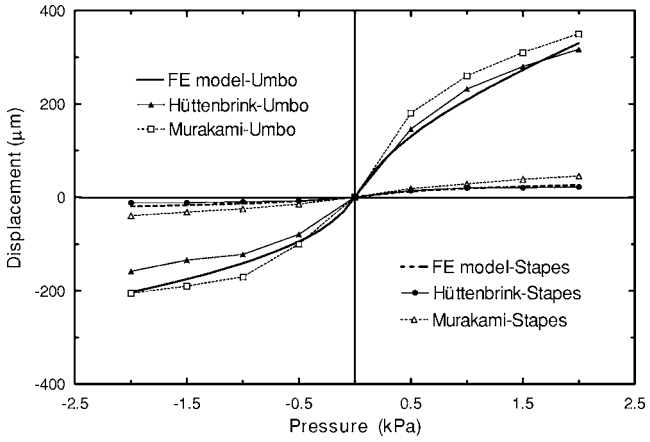


FIG. 4. Comparison of FE model-predicted static displacement (magnitude in unit μm) at the umbo and stapes footplate in response to variation of middle-ear pressure from −2.0 to 2.0 kPa with the measurements reported by Hüttenbrink (1988) and Murakami *et al.* (1997) in human temporal bones.

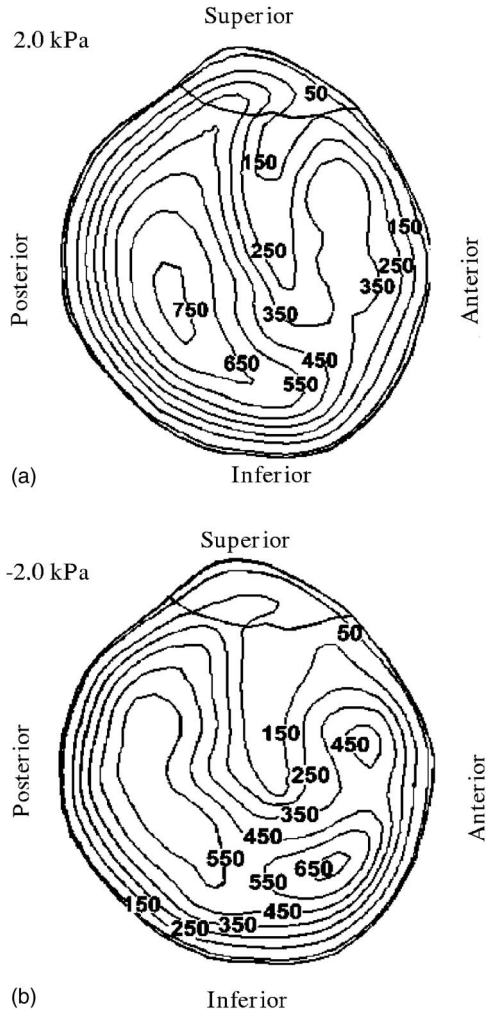


FIG. 5. FE model-predicted displacement contours of the TM (magnitude in unit μm) in medial view. (a) The outward displacement contours of the TM at middle-ear pressure of 2.0 kPa; (b) the inward displacement contours of the TM at middle-ear pressure of −2.0 kPa.

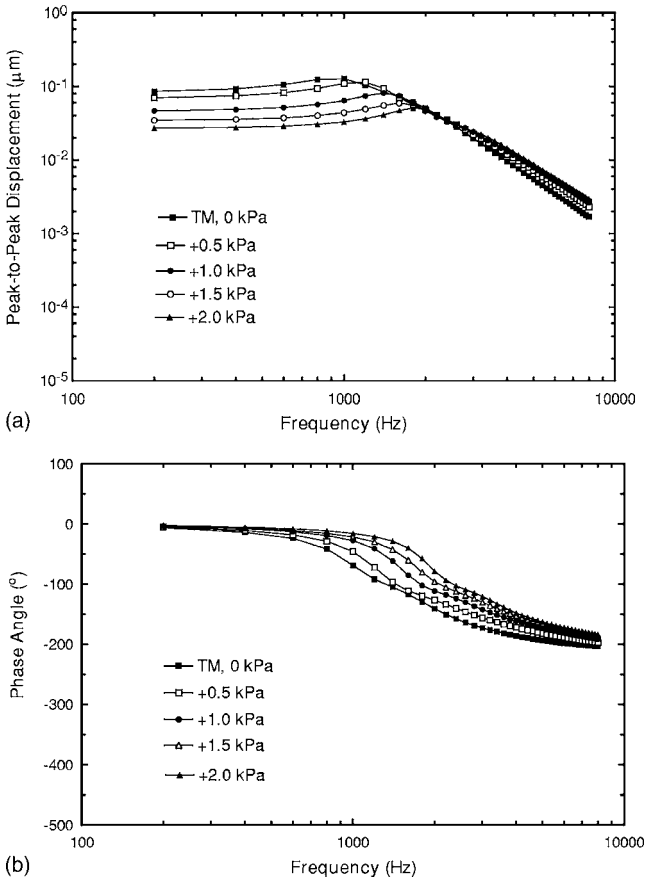


FIG. 6. FE model-derived peak-to-peak displacements of the TM at the umbo across the frequency range of 200–8000 Hz, when middle-ear pressure was varied from 0 to 2.0 kPa. (a) Magnitude; (b) phase angle.

of 490 μm in the supero-anterior quadrant and another of 668 μm in the infero-anterior quadrant. In the posterior part, the maximum of inward displacement was 623 μm.

B. Dynamic analysis

The FE model was conducted on dynamic analysis when 90-dB sound pressure was applied in the lateral side of the TM and the middle-ear pressure was varied from 0 to +2.0 kPa or from 0 to −2.0 kPa. The displacements at the TM and stapes footplate were calculated over the auditory frequency range of 200–8000 Hz based on mechanical properties defined in Sec. II B 3.

Figures 6 and 7 show the FE model-derived frequency response curves of the TM and footplate displacement in response to middle-ear pressure variation from 0 to 2.0 kPa with pressure step of 0.5 kPa. Figures 6(a) and 7(a) show the peak-to-peak displacements of the TM (at umbo) and footplate, respectively. Figures 6(b) and 7(b) show the corresponding phase angles of the umbo and footplate. The curves shown in Figs. 6(a) and 7(a) display that the positive middle-ear pressure caused reduction in the umbo and footplate displacements at frequencies less than 2000 Hz. The phase delay was decreased with the increase of middle-ear pressure as shown in Figs. 6(b) and 7(b).

Figures 8 and 9 show the FE model-derived frequency response curves of the TM and FP displacements in response to middle-ear pressure variation from 0 to −2.0 kPa. The negative middle-ear pressure resulted in more reduction of the umbo and footplate displacement than the positive pressure did. Figures 8(b) and 9(b) show the corresponding phase angles of the umbo and the footplate, respectively. The phase delay was decreased with the increase of absolute value of middle ear pressure.

Figure 10 shows the umbo [Fig. 10(a)] and footplate [Fig. 10(b)] displacement changes (relative to zero middle-ear pressure) under positive middle-ear pressure of 0.5, 1.0, and 2.0 kPa in comparison with the published results by Gan *et al.* (2006a) and Murakami *et al.* (1997). The maximum loss of the umbo displacement [Fig. 10(a)] occurred at a frequency of 800–1000 Hz and the loss values were 2.4, 6.8, and 12.2 dB at middle-ear pressure of 0.5, 1.0, and 2.0 kPa, respectively. The change of footplate displacement induced by the same pressures showed a similar pattern to that of the umbo motion. The maximum loss of footplate displacement was 2.1, 6.9, and 13.6 dB due to the middle-ear pressure of 0.5, 1.0, and 2.0 kPa at a frequency of 800–1000 Hz. The decrease in footplate dynamic displacement at lower fre-

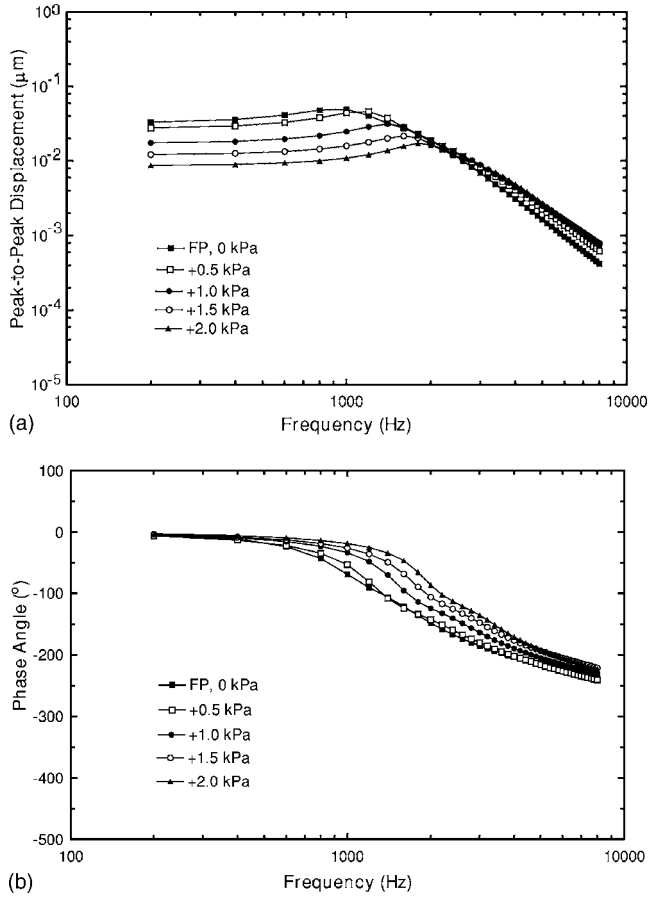


FIG. 7. FE model-derived peak-to-peak displacements of the stapes footplate (FP) across the frequency range of 200–8000 Hz, when middle-ear pressure was varied from 0 to 2.0 kPa. (a) Magnitude; (b) phase angle.

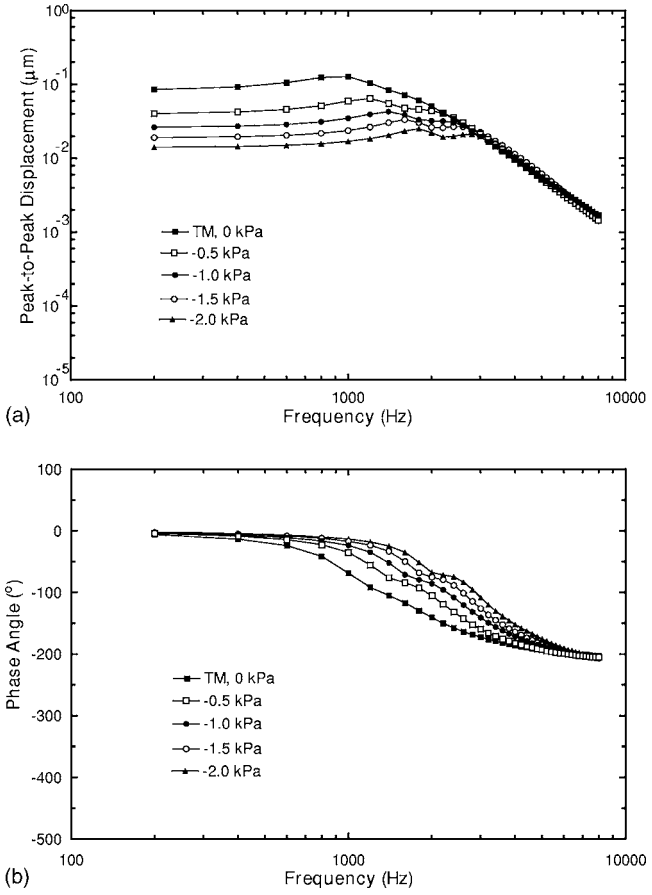


FIG. 8. FE model-derived peak-to-peak displacements of the TM at the umbo across the frequency range of 200–8000 Hz, when middle-ear pressure was varied from 0 to −2.0 kPa. (a) Magnitude; (b) phase angle.

quencies was slightly greater than the decrease of the umbo displacement at a pressure of 1.0 and 2.0 kPa.

Similarly, Fig. 11 shows the FE model-predicted displacement changes at the umbo [Fig. 11(a)] and footplate [Fig. 11(b)] when negative middle-ear pressure was applied. The maximum losses of the umbo displacement occurred at 800–1000 Hz and the reduction values of 7.6, 12.0, and 17.9 dB were observed at middle-ear pressure of −0.5, −1.0, and −2.0 kPa, respectively. The maximum losses of footplate displacement were similar to the values obtained at the umbo (7.6, 11.6, and 17.1 dB caused at −0.5, −1.0, and −2.0 kPa). Compared with the published data by Gan *et al.* (2006a) and Murakami *et al.* (1997), our model-predicted TM and footplate displacement curves show similar frequency-dependent patterns, but the absolute value of the displacement changes are different.

IV. DISCUSSION

A. Static behavior of middle ear in response to middle ear pressure

1. Comparison of model-predicted results with published data

The FE model-derived displacements of the umbo and stapes footplate induced by variations of middle-ear pressure from −2.0 to 2.0 kPa were compared with the data measured in human temporal bones by Hüttenbrink (1988) and Murakami *et al.* (1997) in Fig. 4. As can be seen in this figure, the model-predicted umbo displacement curve agrees reasonably well with the measured data by Hüttenbrink over positive middle-ear pressure range (0–2.0 kPa). Compared with Murakami *et al.*'s data, the predicted umbo displacement was about 16% and 8% smaller than that of Murakami's at 1.0 and 2.0 kPa, respectively. Under negative pressure the umbo displacement curve was located between the curves measured by Hüttenbrink and Murakami *et al.* in temporal bones. Moreover, the model-predicted ratio of the outward to inward TM displacement was 1.38 and 1.63 when the pressure across the TM was 0.5 and 2.0 kPa, respectively. Hüttenbrink's (1988) average outward to inward ratio was 1.85 and 2.0 at pressure of 0.5 and 2.0 kPa, respectively. The model-predicted value of the ratio was smaller than Hüttenbrink's mean data, but the predicted ratio was within Hüttenbrink's results of the individual variation, ranging from 1.1 to 2.9. The derived pressure-displacement curve at stapes footplate was within the data over middle-ear pressure range of −2.0–2.0 kPa reported by Hüttenbrink and Murakami *et al.*

Comparing the TM displacement patterns in Fig. 5 with that measured by Dirckx and Decraemer (1991) in one temporal bone, we found that the local large displacement zones were similar to Dirckx and Decraemer's measurements, but a minor difference existed in the inferior part. Under negative pressure, the model-predicted maximum displacement occurred in the infero-anterior part, while Dirckx and Decrae-

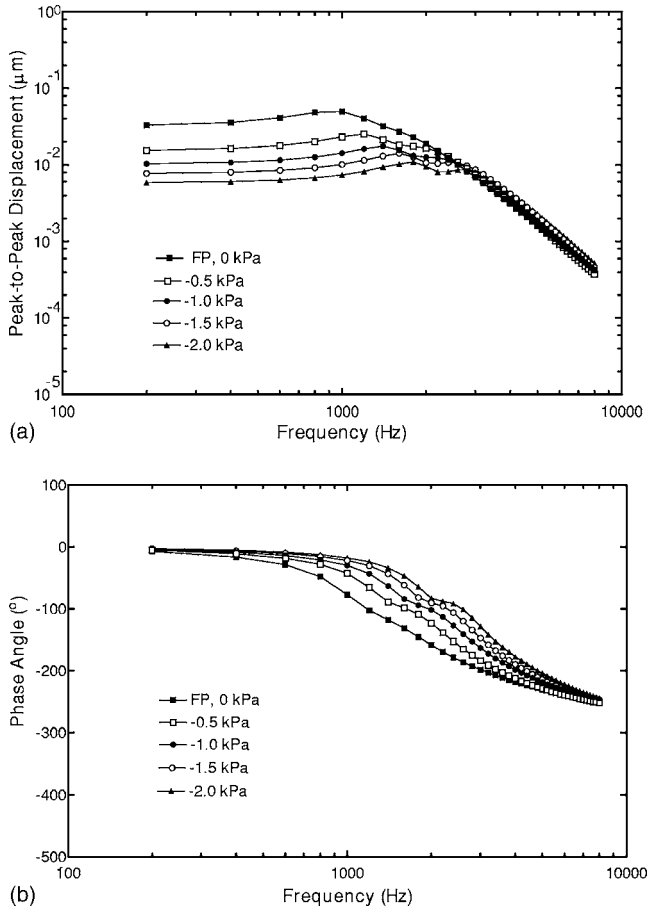


FIG. 9. FE model-derived peak-to-peak displacements of the stapes footplate (FP) across the frequency range of 200–8000 Hz, when middle-ear pressure was varied from 0 to −2.0 kPa. (a) Magnitude; (b) phase angle.

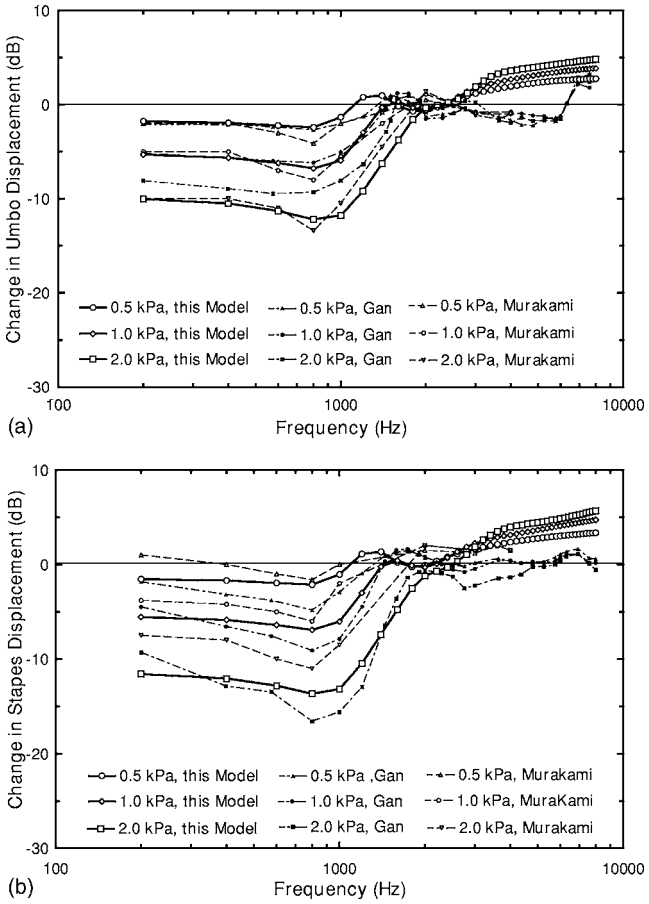


FIG. 10. Comparison of FE model-derived changes in displacement magnitude at positive middle-ear pressure (relative to zero pressure) with the measurements reported by Murakami *et al.* (1997) and Gan *et al.* (2006a) in human temporal bones. (a) Umbo; (b) stapes footplate.

mer's measurements occurred near the middle of the anterior part. Under positive pressure, two displacement maxima were observed by Dirckx and Decraemer in the posterior region: one in the supero-posterior and another in the infero-posterior part, but the model predicted one displacement maximum in the posterior region. The difference between the FE model and experimental observation may arise from the assumption of the TM with a uniform thickness.

The agreement between our model-derived static displacements at the umbo and stapes footplate with the data measured from the temporal bones indicates that the FE analysis with nonlinearities of geometry and material properties provided a fairly accurate prediction for static behavior of the middle ear.

2. Benefits and limitations of the material parameter model for dynamic analysis

It has been observed that the stress level of the TM at negative pressure is higher than that at positive pressure. Thus, the static modulus E_p of the TM is larger at negative pressure than that at positive pressure. This differential sensitivity to pressure indicates the static inward deformation of the TM depends more on the elasticity of the TM itself. This result is generally consistent with the observation by Dirckx and Decraemer (2001) under static pressure.

Dirckx and Decraemer (1991) fitted a curve to describe the relationship of the umbo displacement versus static pressures and found that the static displacement was nearly a factor of 30 higher than the sound-induced motion at the pressure of 0.031 kPa. Fay *et al.* (2005) reported their estimations of elastic modulus of the TM by combining dynamic measurements with composite shell model. Their results were 5–13 times larger than the elastic modulus at the large strains measured by Decraemer *et al.* (1980) from a uniaxial tension test of the human TM. Comparing the dynamic elastic modulus E_d of the TM induced from Eq. (5) with the static elastic modulus E_p, the value of E_d in the radial direction is 35 times and 18 times larger than E_p at the middle-ear pressure of 0.5 and 2.0 kPa, respectively. The value of E_d is 31 times and 11 times larger than E_p at −0.5 and −2.0 kPa, respectively. Therefore, the values of modulus E_d used for dynamic analysis in the present study are generally appropriate.

The present approach is certainly a very simple one, and further improvement of the model can be reached by taking into account both stress and frequency dependence of material properties and the deformation history. On the other hand, we realize that more sophisticated models require a significantly larger effort for their calibration and numerical implementation. As shown in Eq. (5), the present approach

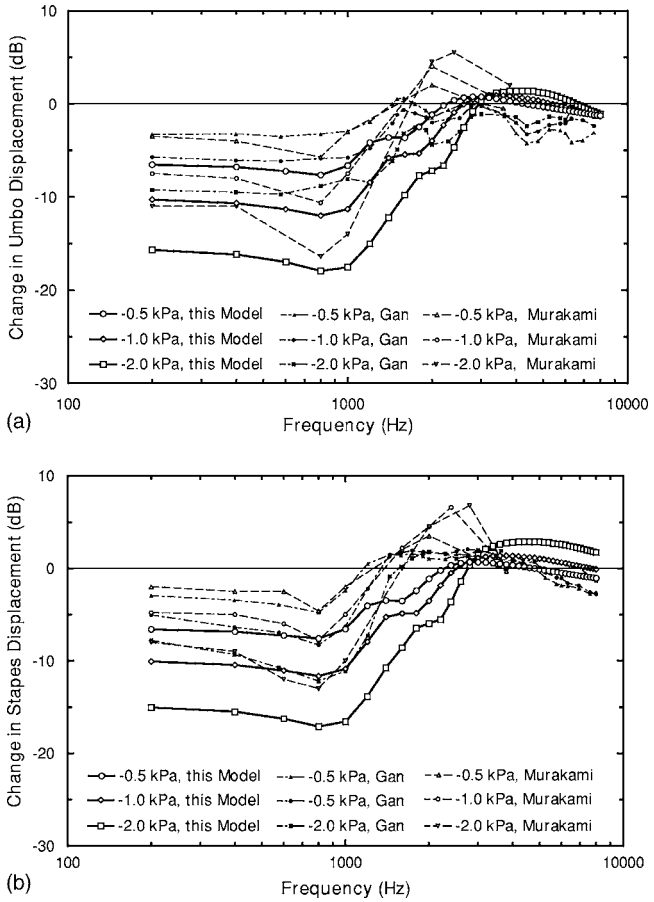


FIG. 11. Comparison of FE model-derived changes in displacement magnitude at negative middle-ear pressure (relative to zero pressure) with the measurements reported by Murakami *et al.* (1997) and Gan *et al.* (2006a) in human temporal bones. (a) Umbo; (b) stapes footplate.

provided a practical model with a minimal number of parameters (k and r), but this method may have some limitations for application in high frequency. However, the proposed model can be used as a basis for further modifications.

B. Dynamic behavior of middle ear in response to middle ear pressure

1. Theoretical considerations

Differentiation of Eq. (8) with respect to pressure p yields

$$\boldsymbol{M}_p\frac{\partial \ddot{\boldsymbol{v}}}{\partial p}+\boldsymbol{C}_p\frac{\partial \dot{\boldsymbol{v}}}{\partial p}+\boldsymbol{K}_p\frac{\partial \boldsymbol{v}}{\partial p}=-\frac{\partial \boldsymbol{C}_p}{\partial p}\dot{\boldsymbol{v}}-\frac{\partial \boldsymbol{K}_p}{\partial p}\boldsymbol{v}. \tag{9}$$

The solution of Eq. (8) or (9) under harmonic analysis can be assumed as

$$\boldsymbol{v}=\boldsymbol{V}e^{i\omega t}, \tag{10}$$

where $\boldsymbol{V}$ denotes displacement amplitude vector, $i=\sqrt{-1}$, ω is circular frequency, and t is time.

Substituting Eq. (10) into Eq. (9) gives

$$(-\omega^2\boldsymbol{M}_p+i\omega\boldsymbol{C}_p+\boldsymbol{K}_p)\frac{\partial \boldsymbol{V}}{\partial p}=-\left(i\omega\frac{\partial \boldsymbol{C}_p}{\partial p}+\frac{\partial \boldsymbol{K}_p}{\partial p}\right)\boldsymbol{V}. \tag{11}$$

Since the middle ear is stiffness dominated at low frequencies ($f<1$ kHz), and assuming the effect of damping can be neglected, Eq. (11) becomes

$$(-\omega^2\boldsymbol{M}_p+\boldsymbol{K}_p)\frac{\partial \boldsymbol{V}}{\partial p}=-\frac{\partial \boldsymbol{K}_p}{\partial p}\boldsymbol{V}. \tag{12}$$

Equation (12) implies that the change of displacement amplitude in response to pressure p is proportional to the variation of stiffness and current displacement. This indicates that the decrease in umbo and footplate vibration at lower frequencies is determined by the increase of system stiffness. The results shown in Figs. 6–9 are consistent with the theoretical analysis.

When the system response is damping dominated ($f>1$ kHz), we neglect the effect of stiffness change and Eq. (9) can be written as

$$(-\omega^2\boldsymbol{M}_p+i\omega\boldsymbol{C}_p+\boldsymbol{K}_p)\frac{\partial \boldsymbol{V}}{\partial p}=-\left(i\omega\frac{\partial \boldsymbol{C}_p}{\partial p}\right)\boldsymbol{V}=-\frac{\partial \boldsymbol{C}_p}{\partial p}\dot{\boldsymbol{V}}, \tag{13}$$

where $\dot{\boldsymbol{V}}$ is velocity amplitude vector.

Further, consider $\boldsymbol{C}_p=\alpha_1\boldsymbol{M}_p+\beta_1\boldsymbol{K}_p$ and assume that $\alpha_1=0$, which yields

$$(-\omega^2\boldsymbol{M}_p+i\omega\boldsymbol{C}_p+\boldsymbol{K}_p)\frac{\partial \boldsymbol{V}}{\partial p}=-\frac{\partial \beta}{\partial p}\boldsymbol{K}_p\dot{\boldsymbol{V}}. \tag{14}$$

Equation (14) suggests that the displacement change in response to pressure p is proportional to current velocity and the variations of damping parameter, as well as system stiffness. From Eqs. (13) and (14) we can see that, if the damping is not varied with the middle-ear pressure, the system will have an identical displacement beyond the stiffness-dominated frequency range.

In dynamic analysis of the model, the effect of prestress or geometric stiffness induced by static pressure on vibration of the middle-ear system was not included because of the limitation of FE analysis codes in ANSYS: the stress field resulted from geometrical nonlinear analysis could not be coupled into harmonic analysis. To estimate the effect of not including prestress in static pressure-induced middle-ear geometry on TM and stapes footplate dynamic movements, we conducted modeling tests by adding a stress field into the original middle ear and deformed middle ear under a certain static pressure loading. The tests results show that there was almost no effect on the TM displacement across the auditory frequency range, and the maximum effect on the footplate displacement was less than 7% of the magnitude at 1000 Hz.

2. Comparison of model-predicted results with published data

Comparisons between the FE model-predicted results and published data on the TM and footplate displacement changes relative to zero middle-ear pressure are shown in Figs. 10 and 11. Murakami *et al.*'s (1997) data were measured at the umbo and stapes head from cadaver temporal bones. Gan *et al.*'s (2006a) data of the umbo displacement

were measured from bones with intact cochlea, and the data of the stapes footplate were measured from bones with opened cochlea.

Figure 10(a) shows that the model-predicted umbo displacement change curves agree reasonably well with the measured data by Murakami *et al.* (1997) and Gan *et al.* (2006a) at frequencies less than 3000 Hz in response to positive middle-ear pressures. Considering the opened cochlea could result in underestimated footplate displacement change (Gan *et al.* 2006a), the footplate displacement curves predicted by the model [Fig. 10(b)] were located between the curves measured by Gan *et al.* and Murakami *et al.* in temporal bones under positive middle-ear pressure.

The effects of negative middle-ear pressure on the umbo and footplate displacements displayed in Fig. 11 suggest that model-predicted umbo and footplate displacement change curves are lower than both Gan *et al.*'s and Murakami *et al.*'s data. We noticed that the model-derived static deformation under negative middle-car pressure was greater than Hüttenbrink's data (Fig. 4). Therefore, the discrepancy between the FE predicted results and the data measured from experiments was likely due to initial geometry of the dynamic model which resulted from a larger static deformation. We will comment on the effect of static deformation of the middle ear on the sound-induced TM and footplate movements in the next section of this paper.

Comparison between Fig. 10(a) and Fig. 11(a) or between Figs. 10(b) and 11(b) indicates that the model-derived displacements show an asymmetry in response to large middle-ear positive and negative pressures. Negative pressure caused more reductions of the sound-induced umbo and footplate displacements than positive pressure did. Our results show a factor ranged from 1.36 to 1.70 of the umbo displacements ratio between the positive and negative pressures. The asymmetry of umbo displacement in response to positive and negative static pressure was similar to the observations by Lee and Rosowski (2001) in gerbil ears. However, the TM movements in response to positive and negative pressure measured in temporal bones reported by Gan *et al.* (2006a) are shown not strongly asymmetric.

It is generally accepted that middle-ear pressure plays an important role in the development of middle-ear diseases, and otitis media with effusion is often associated with negative pressure. A reliable clinical measurement of middle-ear pressure is impractical at this time; tympanometry, an indirect pressure measurement tool, has become the method of choice of many clinicians. A recent report by Dai *et al.* (2007) on comparison of tympanometry and laser interferometry measurements on otitis media with effusion model in human temporal bones indicates that the tympanometric measurements of static compliance and tympanometric width were not affected by the middle-ear air pressure changes. In other words, the pattern or shape of the tympanogram did not reflect the change of middle-ear function change caused by pressure variation. However, both negative and positive middle-ear pressure decrease the mobility of the TM, which was detected by laser vibrometer. Our FE model results are consistent with those published data measured on temporal bones by laser vibrometer.

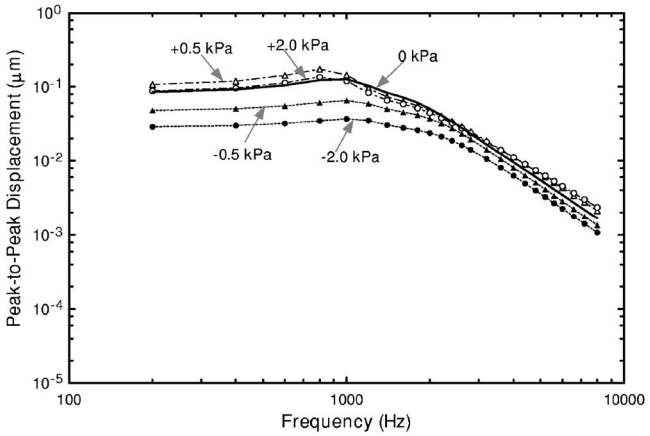


FIG. 12. FE model-derived peak-to-peak displacements of the TM at the umbo in response to middle-ear pressure of ±0.5 and ±2.0 kPa. The calculation was under the assumption that geometry of the TM and middle-ear ligaments was changed with the middle-ear pressure variation, but the mechanical properties of those tissues were maintained as the values at zero middle-ear pressure.

3. Effect of deformed geometry on TM and footplate vibration

The results presented in Figs. 6–9 reflect the effect of the changes of mechanical properties of middle-ear tissues as well as geometry variation of those tissues on the TM and footplate vibration in response to middle-ear static pressure.

To verify the effect of geometry variation of the middle-ear components on TM and stapes footplate vibrations, we selected two levels (0.5 and 2.0 kPa) of positive and negative pressures to investigate the dynamic displacements of the TM and footplate without the change of elastic modulus. The material properties adopted for calculation were those used at zero pressure across the TM. Therefore, the isolated effect of shape change due to middle-ear pressure on the TM and footplate vibration was calculated through the model. Figure 12 shows the TM displacements obtained from the FE analysis in response to ±0.5 and ±2.0 kPa of middle-ear pressure.

Comparing the curves at positive pressure in Fig. 12 with Fig. 6(a), we found that the TM displacements at low frequencies were enhanced at 2.0 kPa, and there was a large increase of the TM displacement at 0.5 kPa, ever higher than the control curve in Fig. 12. The maximum gain below 1000 Hz was 3.0 dB at 0.5 kPa of middle-ear pressure. The model geometry change associated with the pressure of 2.0 kPa resulted in a maximum gain of 0.8 dB. The effect of model geometry on footplate vibration was similar to that of the TM. The maximum gain of footplate below 1000 Hz was 3.0 dB at 0.5 kPa. These results suggested that a small positive pressure in the middle ear could benefit the TM and footplate vibration. This finding may help to explain the phenomenon that a small positive pressure in the middle ear or negative pressure in the external auditory canal produced improvement in stapes vibration observed by Rasmussen (1946) and Murakami *et al.* (1997). It is also demonstrated that the stiffness increase of middle-ear component corresponding to positive middle-ear pressure played an important role in reducing the umbo and footplate movement [Fig. 6(a)]. In other words, the reduction of the umbo and foot-

plate displacement under positive middle-ear pressure was mainly induced by the stress dependence of elastic modulus of the TM and ossicular chain.

The effect of model geometry change induced by negative middle-ear pressure on the vibration amplitude of the TM is also illustrated in Fig. 12. The TM displacement decreased as the pressure varied from −0.5 to −2.0 kPa. The maximum loss of the TM displacement was 6.1 dB at −0.5 kPa, and 11.0 dB at −2.0 kPa. The loss of the footplate vibration was similar to that of the TM. The maximum loss of the footplate vibration was 6.8 dB at −0.5 kPa, and 12.2 dB at −2.0 kPa. Comparing these data with Fig. 8(a), we found that the decrease of the TM and footplate displacements under negative middle-ear pressure was caused by both mechanical properties and geometry, especially the geometry variations.

V. CONCLUSIONS

Combining the hyperelastic Mooney-Rivlin material model and the data of mechanical properties of middle-ear soft tissues measured in our lab, a finite-element analysis for static behavior of human middle ear under various middle-ear air pressures was conducted by taking into account the large deformation and nonlinear behavior of tissues. An empirical formula was developed to calculate material parameters of the ear model for dynamic analysis as the stress-dependent modulus relative to the middle-ear pressure. Dynamic behavior of the middle ear in response to sound stimulus in the ear canal was predicted from the model under various middle-ear pressures. The satisfactory agreements between the model and experimental data in the literature indicate that the effect of middle-ear pressure on static behavior and dynamic functions of the middle ear were well simulated by the present model.

Further, the results from static analysis indicate that a positive middle-ear pressure produces the static displacements of the TM and footplate more than a negative pressure. The dynamic analysis shows that the reductions of the TM and footplate vibration magnitudes under positive middle-ear pressure are mainly determined by stress dependence of elastic modulus. The reduction of the TM and footplate vibrations under negative pressure was caused by both the geometry change of middle-ear structures and the stress dependence of elastic modulus.

The model will be improved in several aspects in future studies such as the structural damping changes with the middle-ear pressure, the acoustic influences of the middle-ear cavity, and the stress relaxation effect on sound transmission through the middle ear.

ACKNOWLEDGMENTS

This work was supported by NIH/NIDCD R01DC006632 and NSF/CMS 0510563 grants.

Bonifasi-Lista, C., Lake, S. P., Small, M. S., and Weiss, J. A. (**2005**). "Viscoelastic properties of the human medial collateral ligament under longitudinal, transverse and shear loading," J. Orthop. Res. **23**, 67–76.

Buss, R., Bauer, R. D., Scattler, T., and Schabert, A. (**1981**). "Dependence of elastic and viscous properties of elastic arteries on circumferential wall stress at two different smooth muscle tones," Pfluegers Arch. **390**, 113–119.

Cheng, T. (2007). "Mechanical properties of human middle tissues," Ph.D. thesis, University of Oklahoma.

Cheng, T., Dai, C., and Gan, R. Z. (**2007**). "Visoelastic properties of human tympanic membrane," Ann. Biomed. Eng. **35**, 305–314.

Cheng, T., and Gan, R. Z. (**2007a**). "Experimental measurement and modeling analysis on mechanical properties of tensor tympani tendon," Med. Eng. Phy. (in press).

Cheng, T., and Gan, R. Z. (**2007b**). "Mechanical properties of stapedial tendon in human middle ear," Transactions of the ASME, J. Biomech. Eng. (in press).

Dai, C., Wood, M. W., and Gan, R. Z. (**2007**). "Tympanometry and laser Doppler interferometry measurement on otitis media with effusion model in human temporal bones," Otol. Neurotol. **28**, 551–558.

Decraemer, W. F., Maes, M. A., and Vanhuyse, V. J. (**1980**). "An elastic stress-strain relation for soft biological tissues based on a structural model," J. Biomech. **13**, 463–468.

Dirckx, J. J. J., and Decraemer, W. F. (**1991**). "Human tympanic membrane deformation under static pressure," Hear. Res. **51**, 93–106.

Dirckx, J. J. J., and Decraemer, W. F. (**1992**). "Area change and volume displacement of the human tympanic membrane under static pressure," Hear. Res. **62**, 99–104.

Dirckx, J. J. J., and Decraemer, W. F. (**2001**). "Effect of middle ear components on eardrum quasi-static deformation," Hear. Res. **157**, 124–137.

Escuder, I., Andred, J., and Rechea, M. (**2005**). "An analysis of stress-strain behavior and wetting effects on quarried rock shells," Can. Geotech. J. **42**, 51–60.

Fay, J., Puria, S., Decraemer, W. F., and Steele, C. (**2005**). "Three approaches for estimating the elastic modulus of the tympanic membrane," J. Biomech. **38**, 1807–1815.

Fung, Y. C. (**1993**). *Biomechanics: Mechanical Properties of Living Tissues* (Springer, New York).

Funnell, W. R. J., Decraemer, W. F., and Khanna, S. M. (**1987**). "On the damped frequency response of a finite-element model of the cat eardrum," J. Acoust. Soc. Am. **81**, 1851–1859.

Gan, R. Z., Feng, B., and Sun, Q. (**2004**). "Three-dimensional finite element modeling of human ear for sound transmission," Ann. Biomed. Eng. **32**, 847–859.

Gan, R. Z., Dai, C., and Wood, M. W. (**2006a**). "Laser interferometry measurements of middle ear fluid and pressure effects on sound transmission," J. Acoust. Soc. Am. **120**, 3799–3810.

Gan, R. Z., Sun, Q., Feng, B., and Wood, M. W. (**2006b**). "Acoustic-structural coupled finite element analysis for sound transmission in human ear—Pressure distributions," Med. Eng. Phys. **28**, 395–404.

Gaihede, M. (**1999**). "Mechanics of the middle ear system: Computerized measurements of its pressure-volume relationship," Auris Nasus Larynx **26**, 383–399.

Hüttenbrink, K. N. (**1988**). "The mechanics of the middle ear at static air pressures," Acta Oto-Laryngol., Suppl. **451**, 1–35.

Koike, T., Wada, H., and Kobayashi, T. (**2002**). "Modeling of the human middle ear using the finite-element method," J. Acoust. Soc. Am. **111**, 1306–1317.

Ladak, H. M., Funnell, W. R. J., Decraemer, W. F., and Dirckx, J. J. J. (**2006**). "A geometrically nonlinear finite-element model of the cat eardrum," J. Acoust. Soc. Am. **119**, 2859–2868.

Lee, C-Y., and Rosowski, J. J. (**2001**). "Effects of middle-ear static pressure on pars tensa and par flaccida of gerbil ears," Hear. Res. **153**, 146–163.

Li, J.K-J. (**1987**). *Arterial System Dynamics* (New York University Press, New York).

Lynch, T. J., Nedzelnitzky, V., and Peake, W. T. (**1982**). "Input impedance of the cochlea in cat," J. Acoust. Soc. Am. **72**, 108–130.

Mooney, M. (**1940**). "A theory of large elastic deformation," J. Appl. Phys. **11**, 582–592.

Murakami, S., Gyo, K., and Goode, R. L. (**1997**). "Effect of middle ear pressure change on middle ear mechanics," Acta Oto-Laryngol **117**, 390–395.

Ogden, R. W. (**1984**). *Non-linear Elastic Deformations* (Wiley, New York).

Pascal, J., and Bourgeade, A. (**1998**). "Linear and nonlinear model of the human middle ear," J. Acoust. Soc. Am. **104**, 1509–1516.

Prendergast, P. J., Ferris, P., Rice, H. J., and Blayney, A. W. (**1999**). "Vibro-acoustic modeling the outer and middle ear using the finite-element method," Audiol. Neuro-Otol. **4**, 185–191.

Price, G. R., and Kalb, J. T. (**1991**). "Insights into hazards from intense

impulses from a mathematical model of the ear," J. Acoust. Soc. Am. **90**, 219–227.
Rasmussen, H. (**1946**). "Studies on the effects upon the hearing through air conduction brought about by variations of the pressure in the auditory canal," Acta Oto-Laryngol. **34**, 415–424.
Rosowski, J. J., and Lee, C-Y. (**2002**). "The effect of immobilizing the gerbil's pars flaccida on the middle-ear's response to static pressure," Hear. Res. **174**, 183–195.

Wave model of the cat tympanic membrane

Pierre Parent[a)]
Mimosa Acoustics, Inc., 129, avenue du Général Leclerc 75014 Paris, France

Jont B. Allen
University of Illinois at Urbana-Champaign, Dept. of Electrical and Computer Engineering, Beckman Institute, Room 2061, 405 North Mathews, Urbana, Illinois 61801

(Received 22 August 2006; revised 20 April 2007; accepted 6 May 2007)

In order to better understand signal propagation in the ear, a time-domain model of the tympanic membrane (TM) and of the ossicular chain (OC) is derived for the cat. Ossicles are represented by a two-port network and the TM is discretized into a series of transmission lines, each one characterized by its own delay and reflection coefficient. Volume velocity samples are distributed along the ear canal, the eardrum, and the middle ear, and are updated periodically to simulate wave propagation. The interest of the study resides in its time-domain implementation—while most previous related works remain in the frequency domain—which provides not only a direct observation of the propagating wave at each location, but also insight about how the wave behaves at the ear canal/TM interface. The model is designed to match a typical impedance behavior and is compared to previously published measurements of the middle ear (the canal, the TM, the ossicles and the annular ligament). The model matches the experimental data up to 15 kHz. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2747156]

PACS number(s): 43.64.Bt, 43.64.Ha, 43.64.Kc [WPS] Pages: 918–931

I. INTRODUCTION

Understanding sound propagation in the ear is critical to our understanding of both the middle ear and the cochlea, and can have a significant impact on the diagnosis of hearing loss. Various diagnosis methods have been derived to isolate the different factors playing a role in the ear's response, both in the middle ear, such as otitis media with effusion (Allen *et al.*, 2005), and in the cochlea, such as damaged outer hair cells (Allen, 2001, 2003). Such models of middle ear and cochlear wave propagation may be roughly classified into two broad categories: distributed and lumped circuit models.

Lumped-parameter circuit representations are usually implemented in the frequency domain using electrical circuit analogies, including the first quantitative model of the cochlea (Wegel and Lane, 1924). In such models, elements of fluid or tissue are represented by inductors representing the element mass, and capacitors representing the stiffness. The analogy with the well-known electrical circuit theory makes this method quite intuitive to use. A key work in the field is the model of the middle ear by Zwislocki (1957,1962), which is based on impedance measurements performed on patients with normal and pathological ears. Due to the tympanic membrane's (TM's) complex geometry and nonrigid construction, and due to its distributed nature, this and other lumped-parameter models are not accurate above a few kHz (Puria and Allen, 1998). Furthermore, modeling details of cochlear and middle ear structures using lumped-parameter methods may require Herculean efforts. For example, modeling a delay in the TM would require a cascade of inductors and shunt capacitors; a second example is the two-piston TM model of Shaw (1977). On the other hand, in their favor, such works have been intuitive and promising starting points for many other models (Shaw and Stinson, 1981; Lynch *et al.*, 1982; Goode and Killion, 1987; Rosowski *et al.*, 1990; Puria and Allen, 1998).

Distributed models may be used when a precise physical model of the ear anatomy and geometry is required. Typically, those models would rely on a finite element analysis or an asymptotic approach, and make sense when it comes to studying complex anatomical structures, such as the eardrum (Funnel and Laszlo, 1978; Rabbitt and Holmes, 1986; Funnel *et al.*, 1987; Fay, 2001; Fay *et al.*, 2002). A clear strength of these models is that they can account for complex mechanical and physical constraints by accurate (but complex) representations of the eardrum behavior. Their main drawbacks reside in the complexity to generate the mesh representing the three-dimensional (3D) structure to be analyzed, their computational time, and that they, like the lumped models, are usually (but not necessarily) implemented in the frequency domain.

Shaw's early representation of the TM as a double-piston source (Shaw, 1977; Shaw and Stinson, 1981) enabled his model to produce a higher-modes response, improving its utility up to 6 kHz. Even better results could be obtained with more sophisticated models, but the number of parameters required to be accurate over an extensive range of frequencies may be unacceptable. All of these models ignore the simple physical source of higher order modes, namely delay. An alternative approach was suggested by Puria and Allen (1998), who represented the TM by a simple distributed transmission line to account for an observed delay, which they estimated from measurements. Using a parameter optimization algorithm, they found excellent agreement with cat impedance data from Allen (1986), over the entire fre-

[a)]Author to whom correspondence should be addressed. Electronic mail: pierre@mimosaacoustics.com

quency range, up to 30 kHz. A major weakness of this TM model is the lack of any impedance transformation, as required by an actual TM. The impedance transformation ratio between the ear canal and the middle ear is known to be around 30 (Bekesy and Rosenblith, 1951; Zwislocki, 1957). In fact, Puria and Allen (1998, page 3475) suggest a more subtle distributed model of the TM, formed by discretizing it into a set of concentric annuli of different impedances, ranging from the canal impedance at the membrane periphery, to the malleus impedance at its center, with the change in impedance along the TM radius being mainly due to its increasing stiffness. In the model presented here, these several ideas are implemented, using a time-domain reflectance model of the middle ear.

The interest of the present study resides in two main points. First, it aims at a full development of the conceptual TM model from Puria and Allen (1998), using a spatially dependent description of the impedance. Second, it uses a time-domain implementation. When a system is described by a lumped-parameter model, it is usually quite easy to derive its frequency response and inverse-Fourier transform it as a convolution in the time domain.

Distributed systems—such as the TM—are infinite order and require a high order approximation to be dealt with properly. A time-domain description only requires interactions of neighboring elements at each observation point and is therefore a computationally sparse representation since only nearest neighbor elements are involved with each time step update. Furthermore, nonlinear systems need to be studied in the time domain; in fact, they usually have to be described by a series of differential equations which need to be solved to represent the system's current state. Conversion from the frequency domain to such a family of equations can be difficult, especially when the order of the differential equations changes (e.g., at a horn's cutoff frequency). A direct time-domain approach is ideal in such cases (Parent, 2005; Parent and Allen, 2006) and has been used to model nonlinear phenomena such as those occurring in the cochlea (Sen and Allen, 2006). Finally, previous works (Allen, 1986; Puria and Allen, 1998) have concluded that the canal and the TM have frequency-independent delays, not always nicely represented by a cascade of mass and stiffness elements. Our approach is to implement this delay in the time domain using transmission lines.

This study is largely about the TM and its dynamics and response, as a variable impedance delay line and impedance matching device. The model is introduced and its results are compared to a range of experimental data. Since impedance is known to be a reliable measurement of the middle ear status (Allen *et al.*, 2005), the model parameters are adjusted to fit experimental impedance-related data from Allen (1986). Finally, the model's behavior is compared to ossicles displacements measurements by Guinan and Peake (1967).

II. BASIC ASSUMPTIONS

This section reviews our underlying assumptions regarding acoustics and transmission lines, and provides our notation.

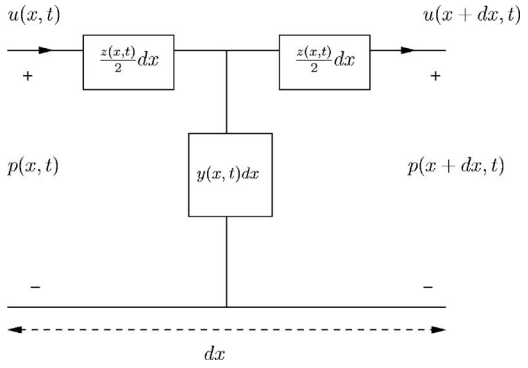


FIG. 1. Circuit representation for an element of transmission line. The series impedance is noted $z(x,t)$ and the shunt admittance is noted $y(x,t)$, in terms of their per unit length distributions (i.e., an impedance will be of the form $z(x,t)dx$).

Most transmission line elements can be approximated by the circuit in Fig. 1 (Brillouin, 1953; Beranek, 1954; Kinsler *et al.*, 2000). In the limit of small dx, this circuit represents a distributed medium. This one-dimensional approach to the middle ear is widely adopted (Rabbitt and Holmes, 1988; Stinson and Khanna, 1994). The pressure is denoted $p(x,t)$, and the volume velocity $u(x,t)$, both variables depending on their position along the propagation axis and on time. Let us define the Laplace variable, $s=i\omega$, where ω is the angular frequency. In the frequency domain, state variables are denoted $P(x,s)$ and $U(x,s)$ and, assuming a one-dimensional (1D) approximation, they are related by the impedance $z(x,t)\leftrightarrow Z(x,s)$ and admittance $y(x,t)\leftrightarrow Y(x,s)$ (Brillouin, 1953):

$$\frac{\partial}{\partial x}\begin{bmatrix} P(x,s) \\ U(x,s) \end{bmatrix} = -\begin{bmatrix} 0 & Z(x,s) \\ Y(x,s) & 0 \end{bmatrix}\begin{bmatrix} P(x,s) \\ U(x,s) \end{bmatrix}. \tag{1}$$

Assuming that no dispersion occurs, that the propagation is plane and lossless, and that impedances are constant along the small length dx, Eq. (1) leads to the classical d'Alembert solution in the time domain (Kinsler *et al.*, 2000):

$$p(x,t) = e^{st}(\mathbf{A}e^{-\gamma x} + \mathbf{B}e^{\gamma x}), \tag{2}$$

$$u(x,t) = e^{st}(\mathbf{C}e^{-\gamma x} + \mathbf{D}e^{\gamma x}), \tag{3}$$

where **A**, **B**, **C**, and **D** are four complex constants determined by the boundary conditions of the propagation. The complex wave *propagation factor* $\gamma(x,s)$ is defined via the impedance and admittance and given in the frequency domain by

$$\gamma(x,s) = \sqrt{Z(x,s)Y(x,s)}. \tag{4}$$

The pressure and velocity can then be decomposed into a positive (factor $e^{-\gamma x}$), and a retrograde component (factor $e^{\gamma x}$), indicated by superscripts ±.

Based on theoretical arguments (Stinson and Khanna, 1994; Lynch, 1981, pp. 146–148), Puria and Allen (1998) have underlined that the vibration propagation in the ear can be assumed to be plane below 25–30 kHz. In this case, p and u components are related by the medium *characteristic impedance*, $z_0(t)$:

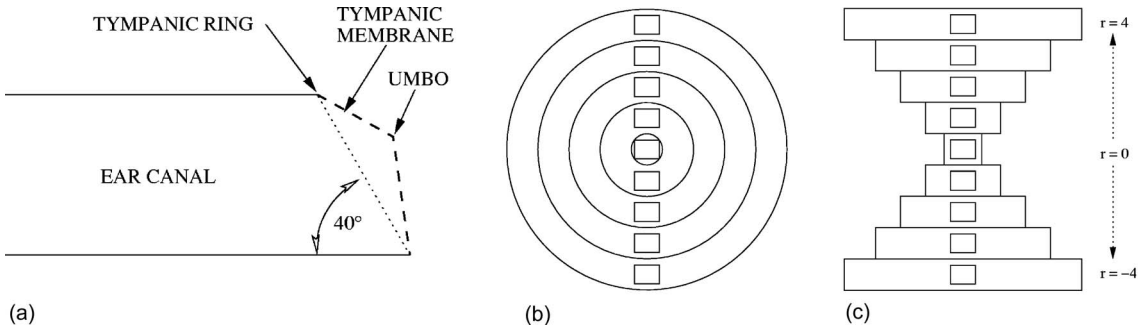


FIG. 2. Discretized tympanic membrane model for $N=5$. (a) Position of the TM in the ear canal. (b) The decomposition in annuli, respecting the circular geometry of the membrane. (c) The one-dimensional model that we actually use, derived from the previous one by applying mass conservation, indicated by the different widths of the stripes. The number of samples on TM is $2N-1$, i.e., nine in this example.

$$p^{\pm}(x,t) = z_0(t) \star u^{\pm}(x,t), \tag{5}$$

where $\star$ represents convolution. After some algebra $z_0(t) \leftrightarrow Z_0(s)$ can be expressed in terms of the circuit element impedances in the frequency domain

$$Z_0(s) = \frac{Z(s)}{\gamma(s)} = \sqrt{\frac{Z(s)}{Y(s)}}. \tag{6}$$

In nondispersive fluids, frequency dependencies in $Z(s) = sM$ and $Y(s) = sC$ cancel out so that $Z_0 = \sqrt{M/C}$ is independent of frequency. If the duct is presumed to be uniform, then Eq. (6) is the expression of the transmission line's characteristic impedance. The solution of the D'Alembert equation leads to the signal representation used in the model: two sets of transmission lines, one representing the forward-going wave and one representing the backward-going wave; both lines actually model the sound wave volume velocity. The use of this approach was first suggested by Kelly and Lochbaum (1963) and is a well-known method in vocal tract (speech) simulations.

In this work the "Kelly-Lochbaum" model is first applied to the ear canal. Since the canal is not perfectly straight, some minor reflections may occur during propagation (Stinson *et al.*, 1982; Stinson and Khanna, 1989; Stinson, 1990; Stinson and Khanna, 1994; Stinson and Daigle, 2005), but shall be ignored. In addition, in an intact ear, the canal opens at the pinna which has previously been modeled by a horn radiation impedance (Rosowski *et al.*, 1988), which for the human pinna has a cutoff that is presently undetermined. However, our study uses the same conditions as the measurements by Allen (1986), i.e., with the stimulus launched very close to the TM. It is therefore reasonable to approximate the remaining part of the canal by a lossless straight tube. Canal samples are then simply distributed along its length, Δx_{ec} apart, where Δx_{ec} is defined by

$$\Delta x_{ec} = \frac{\mathcal{C}_{ec}}{f_s}, \tag{7}$$

where the wave speed in the medium is $\mathcal{C}_{ec}$ and the sampling frequency f_s takes into account spatial sampling constraints of the ossicular chain (OC). The canal length is approximated by the closest multiple of Δx_{ec}, which for our choice of parameters (discussed below) represents a relative error of about 3%.

III. METHODS

This work uses two modeling approaches. The TM, as well as the ear canal, is represented by a distributed model which takes its space-varying properties into account. It is then attached into a classic lumped-parameter model of the OC (Zwislocki, 1962; Puria and Allen, 1998), as is explained in the following sections.

A. Tympanic membrane

The model of the TM is the gist of this study. This section describes basic anatomical aspects of this organ, then explains how it is modeled and interfaced with the model of the ear canal.

1. Anatomical description

We assume that the main role of the eardrum is to ensure energy is efficiently transmitted from the ear canal to the OC. The impedance of the OC is significantly higher than in the canal (our study assumes a factor of 30 (Bekesy and Rosenblith, 1951; Zwislocki, 1957)), thus a direct interface would result in a near-total reflection and large standing waves. The TM has a conical funnel shape, its mouth toward the ear canal and its throat toward the OC (umbo). It is set at an angle with respect to the ear canal axis which varies significantly between species; for the cat it is roughly 40° (Fay, 2001, p. 17), as shown in Fig. 2(a). Lim (1968a,b); Funnel and Decraemer (1996) and others have detailed the geometry of the TM, its microstructure, and its different layers of fibers. The general idea is that these layers, along with the double curvature of the funnel, are responsible for the membrane stiffness (and impedance). To ensure a proper impedance matching, the TM characteristic impedance increases continuously, starting from values close to the canal impedance at its periphery, to much higher values at its center. Propagation of the vibration is realized by transverse waves on the TM surface, coupled to the compressional airborne waves in the canal. Previous works have highlighted that the TM brings an important delay; estimates from Olson (1998) and from Puria and Allen (1998) have shown it is on the order of 30–40 μs: thus, the TM can be seen as a delay line, with a space-varying characteristic impedance to match the air and OC waves. Given this view, the TM is similar to an acoustic horn. Note that this TM horn-like propagation does

not occur in air, but rather as a transverse wave, which is a significant difference when compared to previous theoretical works on traditional acoustic horns (Beranek, 1954; Salmon, 1946a,b).

2. Distributed model

The essence of this contribution is inspired primarily by previous research on high-frequency middle ear models (Shaw, 1977; Goode and Killion, 1987; Puria and Allen, 1998). The *first* goal is to identify the key factors in the interaction of the airborne canal compressional wave and the membrane-borne TM transverse wave. Our *second* goal is to implement a time-domain, reflectance-based (Kelly and Lochbaum, 1963) model simulation of this "canal ⇔ TM" wave interaction. Since this is the first attempt at such a detailed interaction model, many shortcuts and approximations are necessary. It is hoped that any shortcomings and limitations can be the work of future models.

Although it is neither circular nor symmetrical, the TM can roughly be seen as a very shallow horn, and to a further extent as a plane circular membrane if we neglect its depth with respect to its diameter. This assumption is valid for the range of frequencies we shall consider (0.3–15 kHz), where the wavelength is much larger than the TM dimensions; in Fay (2001, p. 21)) the cat's TM depth is estimated to be around 1 mm while our study focuses on wavelength ranging from 7 to 350 mm.

The TM is discretized into N concentric annuli, each having a *characteristic impedance* that gradually increases, from the periphery to the center. Figure 2(b) shows an example of this first decomposition, for $N=5$; the figure shows only a small number of annuli for the sake of simplicity: actual simulations were obtained with $N=71$ (i.e., 14 times more). This representation, a significant approximation of the TM, requires two-dimensional processing of the wave; in fact, due to the tilt of the TM in the canal, different locations on a given annulus will not contact the canal wave at the same time. Properly modeling the synchronization of the different locations seems a difficult issue and so the model is further simplified, with this circular model a conceptual stepping stone to the model shown in Fig. 2(c).

In order to work in one dimension, the circular model (Fig. 2(b)) is replaced by the rectangular model (Fig. 2(c)). As the airborne canal compressional wave touches the TM, the compressional transverse membrane-born wave is impressed into the TM membrane. It is assumed that the impedance match between the airborne sound and the transverse elastic membrane-born sound is such that most (but not all) of the energy is coupled into the membrane. If this were not the case, all energy would be reflected back into the canal, which is not the experimental observation: in actuality, some energy is reflected, but most of it is propagated to the OC. This coupling of energy requires the conservation of mass and momentum (Kirchoff's laws). Thus the volume velocity is scaled in the spatial domain to ensure mass conservation, by taking into account the relative area of the annuli, so that larger annuli (at the periphery) are given more weight than central annuli. This results in the representation of Fig. 2(c), where one stripe corresponds to one semiannulus in Fig. 2(b). If A_i is the area of the stripe at index i, and r_i its radial position, referenced from the TM center, then

$$A_i = \frac{\pi(r_i^2 - r_{i-1}^2)}{2} \simeq \pi r_i \Delta r_{tm}, \tag{8}$$

where Δr_{tm} is the annulus width. Note that the central stripe has the same area as the central disk in the annular discretization. Thus, N annuli are associated with $2N-1$ stripes on the TM. Also, symmetrical positions with respect to the umbo represent the same single annulus, hence they are characterized by the same impedance. Note that it is an even coarser representation than the circular discretization of Fig. 2(b). Following sections of this paper show, however, that it is relevant. In actuality, the impedance varies continuously from the periphery to the umbo and so we assume that reflections occurring during transverse propagation on the TM are negligible (this is the second of the limitations mentioned above, that could easily be repaired, given the motivation). As a consequence, in this model, once the wave has been transmitted onto the TM from the canal, it is *not* reflected anymore and propagates unaltered to the umbo. On stripe i, the TM is then modeled by a pure delay corresponding to the distance to the umbo. Each stripe is then modeled by: 1) a reflection coefficient from the interface with the canal, 2) a pure delay, represented by a double (forward/backward) radial transmission line. With this representation, the stripes are totally independent from each other and they do not interact.

3. Reflection coefficient function

The main issue is to derive an impedance function for the TM in order to associate each TM transmission line with a reflection coefficient from its interface with the ear canal. This derivation is obviously not trivial and is one of the key contributions of this analysis. From Sec. II, each semiannulus (stripe) of the membrane can be represented by a series $Z(s)$ and shunt $Y(s)$ association, where $Z(s)=\rho s$ (ρ is the annulus density, which we use here because we consider an infinitely small volume) and $Y(s)=Cs$ (C is the annulus compliance brought by the membrane's curvature). The basic hypothesis of the model is to assume impedances are invariant on the annulus, i.e., there is no assumed space dependency other than that of the natural annulus area variation (a third significant simplification). The characteristic impedance of the semiannulus number i is then, from Eq. (6)

$$z_0^i = \sqrt{\frac{\rho_i}{C_i}}. \tag{9}$$

We have not found empirical estimates of this compliance in the literature, however another equivalent variable can be more intuitively considered: the speed of sound, $\mathcal{C}$, defined as

$$\mathcal{C} = \frac{s}{\gamma(s)} = \frac{s}{\sqrt{Y(s)Z(s)}} = \frac{1}{\sqrt{\rho C}}. \tag{10}$$

Given an estimate of the density and speed profiles along the membrane, one may then compute the TM impedance at each location. Such profiles, however, are still not readily available to the authors, and would be complex to implement

(a good exercise for the future). The impedance function used in the model then relies on two more hypothesis: (1) the wave speed is constant over the entire TM surface and (2) the impedance profile is exponential. The constant speed hypothesis may well be wrong: previous works strongly suggest that it is not verified on the whole surface (Funnel and Laszlo, 1978; Rabbitt and Holmes, 1986; Funnel *et al.*, 1987; Rosowski *et al.*, 2006). However, our model does not aim at describing the subtleties of the TM 3D motion. Thus we have assumed a constant speed derived from simple delay estimates, the rationale being that such an approximation would be nearly transparent from the input impedance point of view, on which the whole work is based. As for the impedance exponential profile assumption, we have previously suggested that it is probable that some analogies do exist between the traditional acoustic horn theory and our TM model: we have then assumed a simple, classical profile to start our derivation. Thus, the impedance of annulus i assumes the form

$$z_0^i = z_0 e^{-2\mathcal{M}r_i}, \tag{11}$$

where z_0 is the impedance at the umbo, r_i is the radial position on the membrane, with the origin at the center, and $\mathcal{M}$ the flair constant. It is defined from the radius and the impedance transformer ratio of the TM:

$$\mathcal{M} = -\frac{1}{2r_{tm}} \log\left(\frac{1}{\text{Ratio}_{tm}}\right). \tag{12}$$

Our study assumes an impedance transformation ratio of 30 for the global system including the TM and the OC. It is commonly assumed that the OC performs a transformation due to its lever ratio, $N_{\ell r} \approx 2$ (Puria and Allen, 1998), probably due to rotation about the incudo-malleolar (IM) joint (Guinan and Peake, 1967). The impedance transformation realized by the TM alone is then

$$\text{Ratio}_{tm} = \frac{30}{N_{\ell r}^2}. \tag{13}$$

Note that $N_{\ell r}$ represents the ratio of malleus to incus *displacement*, which is why it is squared in the *impedance* ratio computation. The TM reflection coefficient function $\mathcal{R}_{tm/ec}$ can then be computed from the canal impedance, being aware that it depends on its position along the axis; in fact, the canal cross-section area gets smaller and smaller toward its termination, due to the TM inclination. This model leads to very large impedances at the canal termination, resulting in reflection coefficients being close to −1 and too much reflection at the interface. Thus, we have decreased the negative reflection coefficients (by 70%) to eventually obtain the reflection function shown in Fig. 3. Note that it is asymmetric, due to the TM inclination and the resulting impedance variation in the canal.

Reflection coefficients are more intuitive to deal with than impedances. That is why our approach was

1. to derive an impedance function on the TM,
2. then, to compute the corresponding reflection function due to the interface with the canal, and finally
3. to adjust the negative reflection coefficients via experimentation.

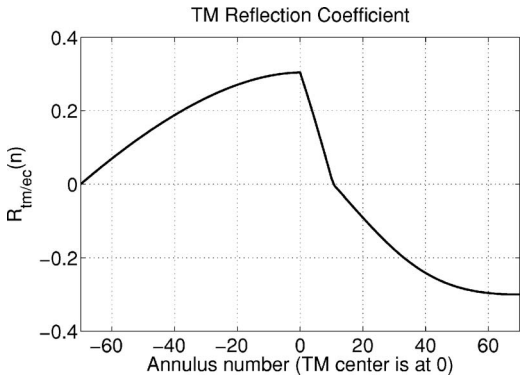


FIG. 3. Tympanic membrane reflection coefficient for a discretized membrane with 71 annuli, i.e., 141 samples. Note the asymmetry in $\mathcal{R}_{tm/ec}$, and its maximum, at the TM center. At the canal termination, the canal impedance is actually greater than the TM impedance, which results in negative reflection coefficients.

It is difficult to justify our final adjustment in terms of TM impedance, but the resulting reflection function seems physical, as is shown from simulations. In our analysis, the previous TM impedance description may not be not an accurate model, but rather viewed as a pedagogical stepping stone to obtain a relevant reflection function.

4. Interface with the ear canal

As in the ear canal model, the number of annuli N is based on the sampling rate f_s and from the wave speed on the eardrum. Both theoretical considerations as well as delay estimates (Olson, 1998; Puria and Allen, 1998), require that the speed of sound on the membrane, $\mathcal{C}_{tm}$, is slower than in the canal by a factor we shall call q, provided in the table of constants. Defining Δr_{tm} as the annulus width, i.e., the spacing between positions on the TM, we then have

$$\mathcal{C}_{ec} = q\mathcal{C}_{tm}, \tag{14}$$

$$\Delta x_{ec} = q\Delta r_{tm}. \tag{15}$$

As a consequence, it is impossible to line up every TM transmission line input with a canal sample. Our solution to this problem is to up sample to improve the density of samples on the TM. The classic up-sampling method consists in padding with zeros and low-pass filtering (Oppenheim *et al.*, 1999). This is not possible here because q is not integral. To solve this problem of fractional delay, we perform two operations. First, we up sample to reduce the size of a sample delay. Second, we model the delay from each TM annulus to the OC independently of the other sections, as summarized in Fig. 4. At location i_0 on the TM, the canal forward wave is scaled by

$$w_{i_0} = \frac{A_{i_0}}{A_{tot}}, \tag{16}$$

where A_{i_0} is the area of stripe i_0, computed from Eq. (8) and $A_{tot} = \Sigma_{i=1}^{2N-1} A_i$ is the TM total area. The wave propagates in air under the TM (first transmission line) and then hits the

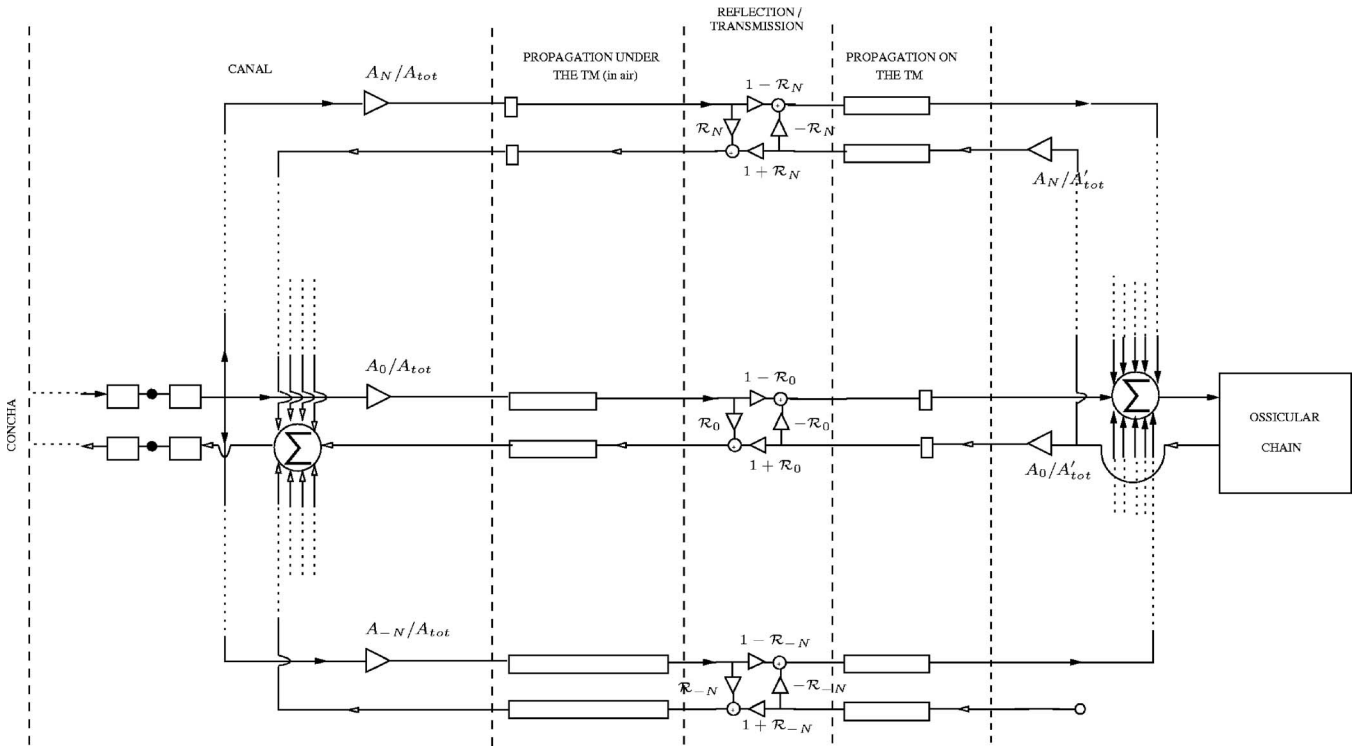


FIG. 4. Interface between the ear canal and the tympanic membrane. Forward and backward propagation path are represented by solid and hollow arrow heads, respectively. Rectangles represent transmission line delays with length proportional to delay. In the first section from the left, at the canal termination, the last forward sample is multiplexed into the interface transmission lines and scaled according to the ratio of the stripe area A_i over the total TM surface area A_{tot}. The second section from the left represents transmission lines showing the wave in air under the TM; due to the inclination of the drum, they bring different delays. At the interface between the canal (air) and the membrane (the middle or third section), the wave is split into a transmitted and a reflected part, computed from the knowledge of the reflection coefficient on the TM (see Fig. 3). In the fourth section the transmitted part propagates on the TM toward the umbo (right-most series of transmission lines) where finally, in the fifth section, all contributions are added and then feed the OC. The reflected part at the canal/TM interface propagates in air back to the canal input. In the backward propagation, the wave coming from the OC is only input into the TM superior region, where it is in contact with the manubrium, and is multiplexed and scaled according to the ratio of the stripes areas over the superior region area, A'_{tot}. It propagates to the canal input using the same path than the forward-going wave. All contributions are then added before being input into the canal transmission line. Note that our actual implementation takes into account the future possibility of adding an input from the middle ear cavity space into the inferior region, but these (velocity) inputs are presently zeroed (far lower-right corner).

TM. At that point, it is split into two contributions: one transmitted on the TM, and the other reflected into the canal. The reflection coefficient, $\mathcal{R}_{i_0}$, is different at each location on the TM and has been derived in Sec. III. Since $\mathcal{R}_{i_0}$ is frequency independent, transmitted and reflected contributions are simply computed by a multiplication:

$$\begin{bmatrix} u^-(x) \\ u^+(x+dx) \end{bmatrix} = \begin{bmatrix} \mathcal{R}_{i_0} & 1+\mathcal{R}_{i_0} \\ 1-\mathcal{R}_{i_0} & -\mathcal{R}_{i_0} \end{bmatrix} \begin{bmatrix} u^+(x) \\ u^-(x+dx) \end{bmatrix}. \quad (17)$$

Both transmitted and reflected parts propagate in their respective medium, toward the umbo or the canal termination where they are summed up and fed to the single-transmission line representation used in the OC and in the canal, respectively. At the umbo, the impedance is matched with the OC and we assume that waves are entirely output to the OC, and that they do not propagate to the other side of the TM. This is actually quite intuitive given the conical shape of the TM.

The backward-going wave, reflected from the OC, uses a similar path in the opposite sense. Along the OC, we assume a transverse propagation of the wave. As a consequence, the OC applies a force on the TM superior region, along the manubrium. Note that it is different from the interface between the TM and the canal: in the canal, we assume the sound wave to be compressional in air and in contact with the TM over its entire surface. Due to the independence of the annuli transmission lines, the inferior region of the TM does not play a role in the backward propagation of the wave. Note that the middle ear cavity space has a compressional wave in air which will apply a pressure over the entire TM. We have presently chosen not to model this phenomenon; however, our current implementation is ready for such an extension. As indicated in Fig. 4, all inputs of the inferior region for the backward-going wave are set to zero but could receive a different input in a future development of the model.

5. Power conservation

It is important to discuss power conservation at this point, as it is necessary for the validity of the model. Two operations are involved in the process of propagation: the spreading of the wave from the canal termination (and from the OC input) to the TM and the reflection junction at the interface between the canal and TM transmission lines. The spreading of the wave is designed to conserve volume velocity (the sum of the scaling factors is 1) and so the operation is equivalent to connecting a duct into a series of smaller

ducts which cross section areas sum up to be equal to the bigger duct cross section area, which does conserve power. As shown by Bilbao (2001), the reflection junction also conserves volume velocity. Thus the middle ear model, as implemented, is lossless.

B. Ossicular chain

1. Anatomical description

The OC is an association of three bones coupling the TM output and the cochlea oval window. Complex mechanical interactions between them and their ligaments act as a lever from the input of the chain to its output: as a consequence, the impedance is increased with minimal sound reflection. Note that this is also an impedance matching process, but lumped rather than distributed. This description as a lever, while valid at low frequencies, breaks down at higher frequencies due to the OC mass. A more refined high-frequency description is beyond the scope of this study and we have assumed, as in (Puria and Allen, 1998), that the lever ratio was constant over the entire frequency range of interest.

After having propagated along the TM, the wave reaches the umbo, where it is connected directly to the malleus manubrium tip. In actuality, the manubrium is connected to the TM along the entire length of its superior region. The wave then propagates through the OC, and arrives at the stapes footplate, a flat piece of bone embedded in the cochlea oval window and fixed by the annular ligament. The footplate moves in the oval window, transmitting the vibration to the cochlea fluid, and in reverse for the retrograde wave. The various lumped elements make up a transmission line, having series mass and shunt stiffness, with a characteristic impedance given by Eq. (6).

2. Lumped-parameter OC model

The middle ear model shown in Fig. 5 is largely inspired by the circuit presented by Puria and Allen (1998), which in turn is based on the work by Zwislocki (1957). The lumped-parameter circuit for the OC representation has proved to be a modeling method which is easy to implement and accurate (Brillouin, 1953; Zwislocki, 1957, 1962; Lynch *et al.*, 1982; Puria and Allen, 1998). This method has been implemented in the time domain, to be used with the TM model.

Generally, bones are represented by mass (inductance) and ossicular joints by springs (compliance): the OC is then a sequence of mass/compliance shunt associations. The reason for this is suggested by Puria and Allen (1998): to increase the bandwidth of lumped-parameter circuits, each mass is coupled with a shunt compliance; in such an association, the two-port characteristic impedance does not depend on frequency (the dependencies in s cancel out) and compliance can be adjusted given the mass, so that the impedance is matched. Each two-port is defined in the time domain by the four frequency-dependent reflectance filters described in the caption of Fig. 5, from which the outputs are computed. A detailed description of these reflectance filters' computation is not provided here, other than to say that the "ABCD matrix" method (expression of output velocities from input velocities) and the bilinear transform were used. Note how the OC lever ratio is represented by a transformer between the malleus and the incus. Its ratio is denoted $N_{\ell r}$ and the malleus and incus impedances are related by $Z_i = N_{\ell r}^2 Z_m$. The OC implementation details are provided in an appendix.

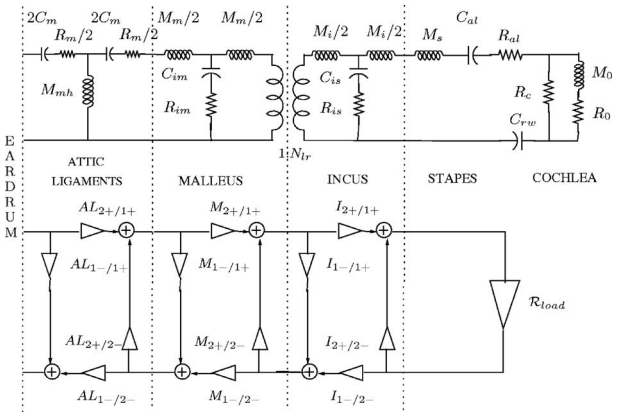


FIG. 5. Ossicular chain circuit representation. Each element is modeled by a two port with four filters given by the matrix of Eq. (17). The multipliers are frequency-dependent, and thus must be implemented as convolutions in the time domain. This is done using a bilinear transformation of the reflectance function in the frequency domain. this operation converts the Laplace-domain formula into its digital domain (Z—transform) equivalent. It replaces the Laplace variables s by $2fs(1—z)/(1+z)$, where fs is the sampling rate. $X_{2+/2-}$ represents the filter computing the forward output from the backward input, $X_{1-/1+}$ the backward output from the forward input, etc. The stapes/cochlea association is modeled by only one filter (Lynch *et al.*, 1982). The OC lever ratio is represented by the transformer between the malleus and the incus, its ratio is denoted $N_{\ell r}$. Malleus impedance Z_m is matched with the TM central impedance, and incus impedance is equal to $N_{\ell r}^2 Z_m$.

IV. RESULTS

Various simulations are run with the model described previously, and the results are compared to experimental conditions, to check for consistency against normal as well as pathological conditions. These comparisons are then used to refine the model parameters, starting from (Puria and Allen, 1998). Once the parameters are established for each section, no further changes are made to that section. At each stage the model is compared to experimental data. Finally the entire model is compared to the ossicles displacements ratios of Guinan and Peake (1967).

A. Impedance-related measurements

Impedance measurements by Allen (1986) are used to determine the model parameters. Following the approach by Zwislocki (1962), simpler cases are studied first, such as the blocked TM, and the disarticulated stapes, in order to reduce the number of unknowns. The complexity of the model is then incrementally increased, with the previously adjusted parameters fixed.

1. Case I: Blocked tympanic membrane

The simplest case (Fig. 6) is the input impedance of a blocked umbo, i.e., loaded by an infinite impedance at its output ($\mathcal{R}_{\text{umbo}}=1$). This condition is necessary as it is used to adjust the TM reflection coefficients. The results of Fig. 6

show the expected characteristics of a pure delay transmission line, corresponding to the residual canal and TM delay. As a sanity check, this delay can be estimated from the output of the TM. In Fig. 6(b), the main pulse occurs at 40.43 μs. Subtracting the canal delay of 4.32 μs (length of 0.15 cm), the TM delay estimate is 36.11 μs (estimate by Puria and Allen (1998) is 35.7 μs for the ear on which this model is based (p. 3476). Their estimates for the other two ears are 34 and 41 μs (p. 3472)).

The general shape of the output signal (Fig. 6(b)) can be roughly approximated by two consecutive broad, dispersed pulses: the first one (the greatest) corresponds to the propagation of the wave coming from the superior region of the TM, and is followed by the wave propagating on the inferior region. Some noticeable "perturbations" occur periodically in the fine "structure" of the signal, especially for the inferior region wave. This is due to the discrete TM implementation that was used: for each location on the TM, the corresponding canal+TM delay is rounded to be a multiple of the sampling period, which leads to the signal not being perfectly continuous at the umbo. Those perturbations do not have any influence over the general behavior of the system, as is seen from the impedance plots. Actually, given a physical measurement of the blocked TM response, one would not see the discrete pulses, which are an artifact of the discrete nature of the TM model, as any low-pass filter effect would remove them. The time reflectance shows three broad consecutive pulses. The first two pulses (at 18.19 and 41.44 μs, respectively) show the same periodical perturbation that we previously discussed: the first one corresponds to the primary reflection of the wave on the superior region of the TM where reflection coefficients are positive, the second one corresponds to negative reflections in the inferior region. Eventually, the propagated signal is reflected by the blocked condition and appears as the third broad main pulse (at 71.26 μs).

Figure 6(c) presents the reflectance magnitude: it reveals a slight decrease of the high frequencies, above 15 kHz, where it crosses 0.8. Since the line is lossless, this phenomenon is nonphysical. Our hypothesis is that the spreading of the wave and the various delays at the canal/TM interface lead to the reflected signal at canal input being similar to a sequence of pulses (see Fig. 6(a)), which Fourier transform rolls off at high frequencies. We believe that, in actuality, a more complex phenomenon (or a combination of such) takes place along the path of propagation which compensate for this high-frequency loss. A probable candidate for such a phenomenon is the interaction between the TM transmission lines. We have assumed no reflection occurred on the TM, which is an approximation: it is possible that such reflections do exist and have a noticeable influence at high frequencies. The model as implemented is therefore not accurate above 15 kHz. The impedance plot of Fig. 6(d) is characteristic of a blocked transmission line with acute poles and zeros, since no damping is assumed. The previously mentioned high-frequency reflectance decrease is also visible in the impedance magnitude's less acute pole around 15 kHz, although less obvious.

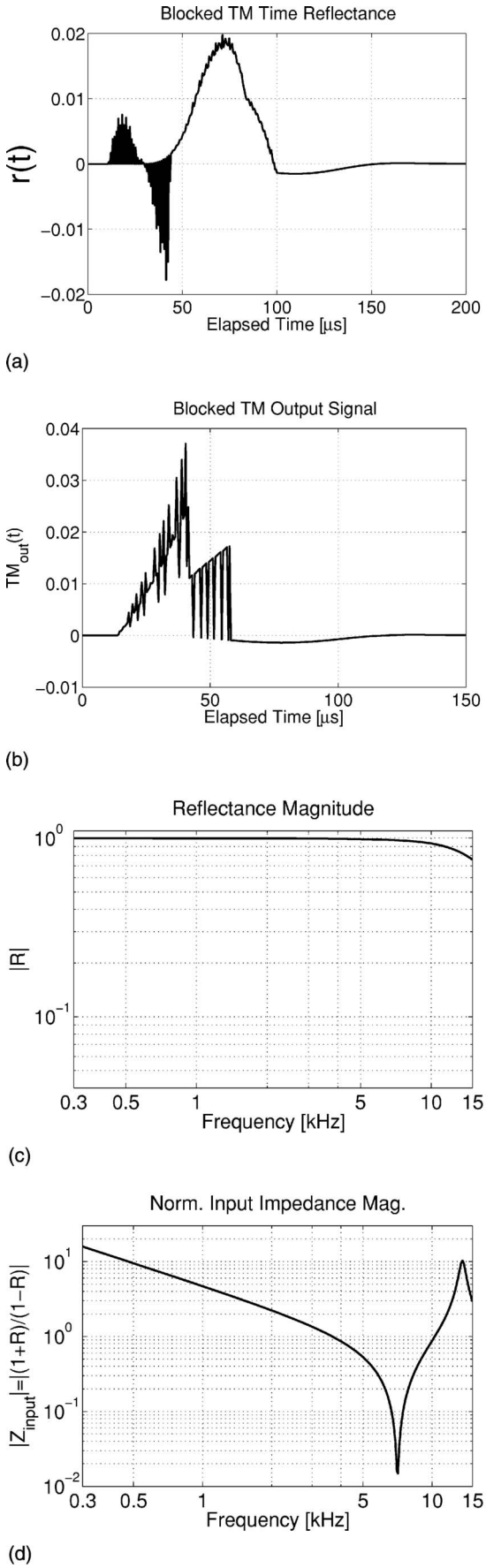


FIG. 6. Tympanic membrane for the *blocked-TM* (clamped-umbo) condition: time-domain reflectance and TM output, reflectance and impedance magnitudes. (a) Reflectance (time signal), (b) TM output, (c) reflectance magnitude, (d) impedance magnitude.

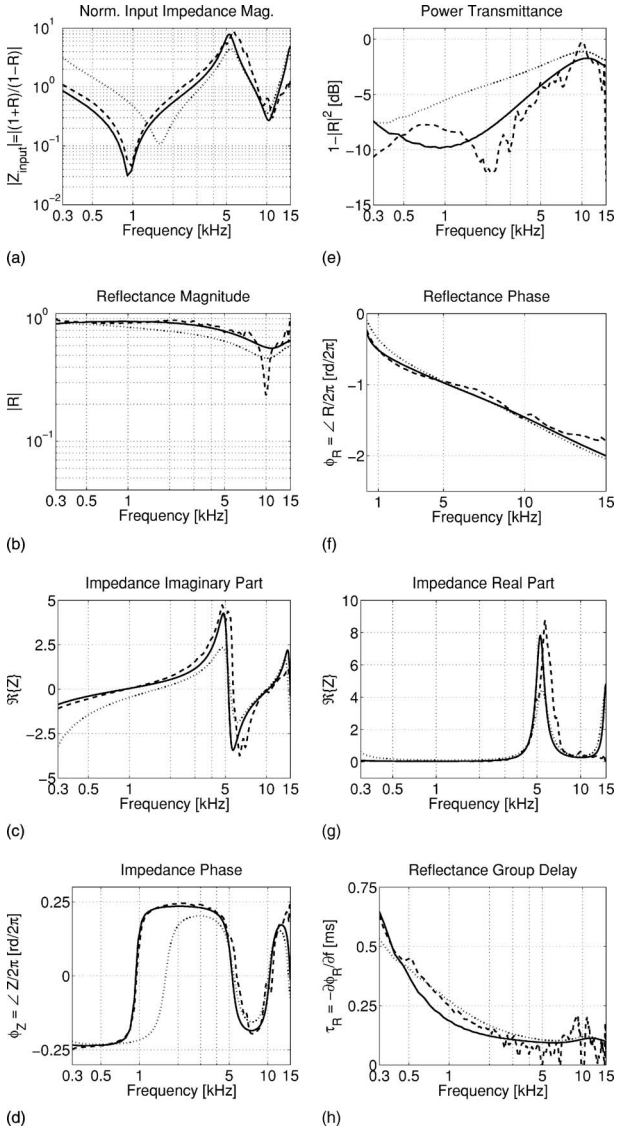


FIG. 7. In this figure we compare the experimental data for the disarticulated stapes (DS) [dashed] with two model simulations, the disarticulated stapes (DS) [solid] and the drained cochlea (DC) [dotted]. In the left column are four input impedance measures: (a) the impedance magnitude, (b) the reflectance magnitude, (c) the impedance imaginary part, and (d) the impedance phase. In the right column are: (e) the power transmittance, (f) the reflectance phase, (g) the impedance real part, and (h) the reflectance group delay. As discussed by Allen (1986), removing the stapes reduces the stiffness below 1 kHz by about a factor of 3, as shown in (a) for the model calculations, but otherwise has only a small effect, especially above 5 kHz.

2. Case II: Pathological ears

Figure 7 displays the results obtained with the model, compared with experimental data from Allen (1986), for the disarticulated stapes (DS) experiment. In this case, the ossicular chain is cut free just before the incudo-stapedial (IS) joint, corresponding to a short circuit (the load impedance is zero), resulting in $\mathcal{R}_{\text{incus}}=-1$. The DS experiment is important for adjusting the attic ligaments parameters, C_m and R_m, as well as the malleus parameters, M_m, C_{im}, and R_{im}. The intuitive behavior of the system is confirmed by the experimental data (dashed line) and the results of the model (solid line). The impedance magnitude has acute poles and zeros, characteristic of low-loss standing waves. A significant difference between the DS data (a) with respect to a normal short-circuit transmission line, is the low-frequency stiffness response below 1 kHz, confirmed by the phase response (d). Namely, the short circuit line has a zero at $f=0$, whereas the middle ear, in short circuit, has a pole. Such a stiffness most likely results from the attic ligaments. In the model this stiffness results from C_m. The irregularity in the experimental resistance in (g) is due to the inherent difficulty in measuring a relatively small resistance in the presence of a large stiffness (i.e., the impedance angle is very close to −90°). Also are shown the reactance (c) and reflectance phase (f).

The reflectance is shown (b). It is less than 1 because of the OC losses. The power reflectance part (e) shows the relative low power transmitted into the disarticulated middle ear. Below 2 kHz, the experimental data show a slight increase of the transmittance from −10.5 to −8 dB between 300 and 700 Hz, then a plateau up to 1.5 kHz, and a sharp decrease down to −12 dB at 2 kHz. This behavior is not captured by the model which first decreases from −7.5 to −10 dB between 300 Hz and 1 kHz, and then increases up to 10 kHz. However, the model remains within ±3 dB of the experimental data, except perhaps around the sharp minimum at 2 kHz; (h) gives the latency of the reflectance, a measure of how long the energy remains in the middle ear. The low-frequency slope of the model is slightly smaller than for the experimental data, highlighting that the model stiffness is also slightly smaller, which is confirmed by the impedance magnitude in this region. Around 10 kHz, the experimental group delay shows a local maximum which is not present in the model simulation: this discrepancy is probably due to the least inaccuracy of the model at high frequencies, as suggested from the blocked-TM results.

In the next stage of analysis we add back the stapes and annular ligament, as well as the cochlea model. In Fig. 7 (dotted lines), are shown model results from the case of the drained cochlea (DC). In this case the experimental data are not shown, to reduce the clutter, and because the difference is easily described. In the DC experiment, the entire OC is left intact but the cochlear fluid is drained, resulting in a great reduction of the cochlear load (stapes volume velocity is basically unloaded). From the data of Allen (1986), the canal impedance is very similar to the DS experiment, except for an increase in the stiffness below 1 kHz, due to the inclusion of the annular ligament. This change can also easily be seen in the real and imaginary parts of the impedance, in Figs. 7(c) and 7(g). The difference is mainly visible at low frequencies, and not at all above about 4 kHz, except for a general decrease of the impedance magnitude because the middle ear is better matched to the canal in this case. Note how the resistance increases over most of the frequency range. At 5 kHz this is difficult to see due to the large peak in the resistance, associated with the pole in the impedance.

3. Case III: Intact ear

With the intact ear, the change from the two previous cases is dramatic, although the main difference in the model is the cochlear resistance, previously equivalent to an air load. The consequence of this highly resistive final load, matched to the middle ear output, is a dramatic damping of

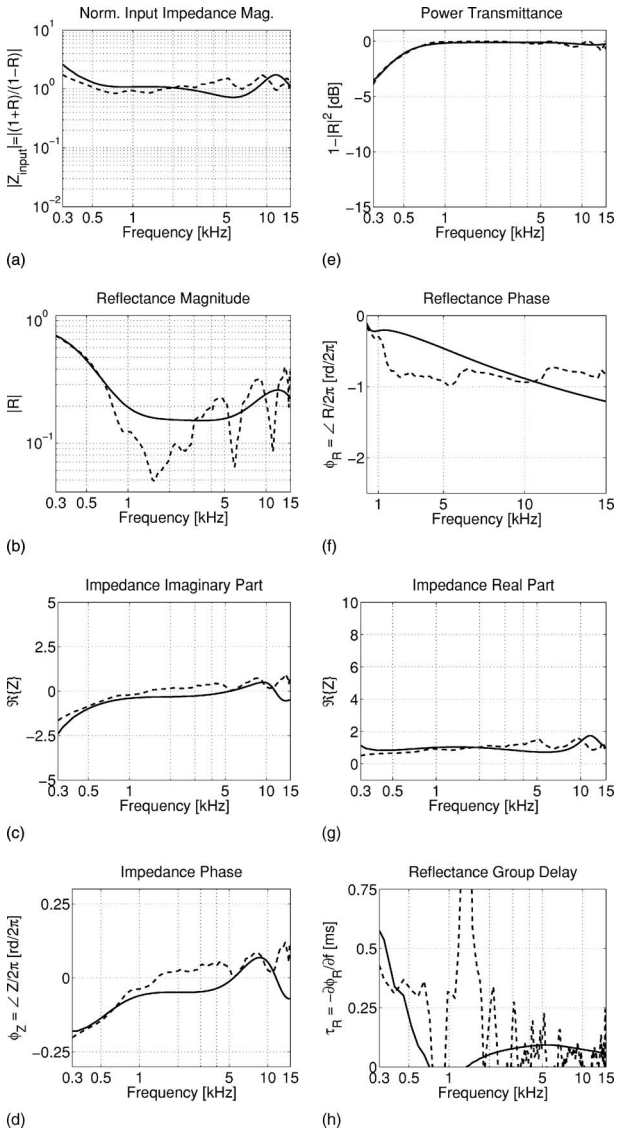


FIG. 8. In this figure we compare the experimental data for the intact ear [dashed] with the model simulation [solid]. In the left column are four input impedance measures: (a) the magnitude impedance, (b) reflectance magnitude, (c) the impedance imaginary part, and (d) the impedance phase. In the right column are: (e) the power transmittance, (f) the reflectance phase, (g) the impedance real part, and and (h) the reflectance group delay. The intact ear brings a major change with respect to the pathological ears: due to the increased load impedance, standing waves are damped out and reflectance is much lower, overall.

the standing waves, and the reflectance being globally much lower, as shown in Fig. 8. From the impedance plot, the damping is obvious since the phase of the impedance is close to zero, above 1 kHz.

The overall match to the experimental data is generally excellent. A more detailed study of the impedance real and imaginary parts, Figs. 8(g) and 8(c) shows some small discrepancies. The low-frequency slope of the impedance is due to the stiffness of the system: it can be seen from Fig. 8(a) that the model is slightly stiffer, which results in too high an impedance and group delay. Above 1 kHz, experimental data show that the resistance slightly increases, while the model is less resistive (Fig. 8(g)). Also, the ear becomes slightly mass dominated, as can be seen from the phase plot in Fig. 8(d) and the imaginary part in Fig. 8(c). This has obvious consequences on the impedance and reflectance between 1 and 5 kHz: Figs. 8(a) and 8(b) clearly show the model is not as well matched as the actual ear. Between 5 and 10 kHz, experimental data show resonances (standing waves) which are not captured here. Above 10 kHz the ear is still slightly mass dominated, while the model has a slightly negative imaginary part for the impedance. Another serious discrepancy can be seen in the reflectance phase in Fig. 8(f) around 1.2 kHz, the measured phase suddenly drops from -0.3 to -1 rad$/(2\pi)$ which can also be seen as the notch in the reflectance plot. More generally, the complex reflectance is characteristic of a series of delayed pulses with different phases and magnitudes, which results in the several notches in the magnitude and the sawtooth-like behavior in the phase. The model, however, remains smooth. This is even more obvious in the group delay plot (Fig. 8(h)) which shows a huge peak at 1.2 kHz, due to the phase discontinuity. Also, the low-frequency group delay is nearly constant while our simulation is progressively decreasing. This behavior probably shows a shift of mode between the stiffness-dominated low-frequency region and the mid-frequency range which is more resistance dominated: the change is very sharp in the actual ear while the model seems to smooth out the transition. Modes transitions are quite difficult to appreciate and our model is probably too simplistic to deal with them accurately (Fay *et al.*, 2006).

Despite those numerous discrepancies, the model matches the data fairly well, and the general behavior of the ear is nicely captured. As the agreement is clearly less good than with the pathological ears, a probable source for the discrepancies is the lack of a cochlea model. Simulations run with the stapes/cochlea filter on its own, using Lynch's values give excellent agreement with his results. However, clear differences appear when it is used with the global TM-OC system since it requires significant adjustments to match experimental data, especially for the helicotrema parameters. Our hypothesis is that the model by Lynch has been derived using measurements made by direct stimulation at the stapes footplate, hence bypassing the entire TM and OC. We have tried to compensate such a major difference by multiplying the impedances by Ratio_{TM}, which seems to be relevant for most parameters except the helicotrema. We suspect that impedance transformation is not the only process involved in the interaction between the OC parameters and the cochlea and it is possible that the current circuit is too coarse to deal with such processes accurately.

B. Ossicles displacement ratios

An ultimate validation of the model is carried out by comparing its results to data which have not been used at all in its derivation. Guinan and Peake (1967) measured ratios of displacements in the OC (slippage). These are taken to be classical references of the ear typical behavior. This comparison is shown in Fig. 9. In general, the model is in good agreement with the experimental data, except for the high-frequency phase of the incus to malleus displacement. This discrepancy is actually the exact same as the one reported by

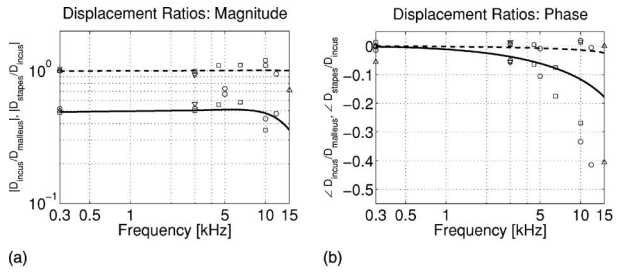


FIG. 9. Ossicles displacement ratios: magnitude and phase. The model incus to malleus displacement ratio is the solid line, the model stapes to incus displacement ratio is the dashed line, symbols are cats 58, 65, 68, and 69 from Guinan and Peake (1967). (a) Ratio magnitude, (b) ratio phase.

Puria and Allen (1998): since the model uses their parameters values, it is not too surprising and actually underlines the consistency of our time-domain approach. This phase difference merely represents a delay difference. It is possible that an additional slippage factor is involved at the IM joint which could have an influence on the phase. Such a factor was used in the work by Puria and Allen (1998, p. 3471) but its influence is not obvious and, in any case, it did not bring any improvement here.

V. DISCUSSION

The model gives a robust match across all conditions. However, some discrepancies in the middle ear model need to be refined, especially at high frequencies, above 10 kHz. Further refinements will require an improved cochlear load simulation. This study leads to two significant results. The first one is the relevance of the time-domain implementation, a very intuitive and convenient approach, and, in our view, certainly useful to model distributed and/or nonlinear systems, such as the TM (and the cochlea). The second is the added insight that the time domain provides into the basic operating of the TM and its impedance-matching properties; in our view this approach provides a different way to look at the TM, and to appreciate how wave propagation occurs and how the different delays interact.

A. Attic ligaments

Experimental data for the disarticulated stapes experiment show the presence of a stiffness element in the ear, before the annular ligament. It can be assumed that the IM- and IS-joint ligaments do not add stiffness to the input impedance, as the joints are compensated by the ossicles mass, associated with them (matched-impedance transmission line condition). It is probable that this stiffness results either from the TM, or the attic ligaments. Puria's model—and consequently this study—assumes the attic ligaments are responsible for this low-frequency stiffness: such an assumption is relevant and both models show reasonable agreement with this hypothesis. It is open to question, however, until it is experimentally verified; in fact, from the input impedance point of view both approaches are quite equivalent and our model is not able to partition between them. At the junction between the umbo and the malleus and at the IM joint, the bone movements are mainly rotational. Thus, it can be assumed that the ligaments are not significantly stretched: they mainly have a "ball-bearing"-like behavior, bonding the ossicles to the attic, rather than being stretched. Alternatively, it seems reasonable to think the TM, with its many layers of microfibers, brings stiffness to the system. This is entirely an experimental issue.

B. Tympanic membrane cutoff

The impedance-matching role of the TM is widely recognized. Various hypothesis have been suggested to explain this role, such as the TM/footplate area ratio, but the hypothesis here is that this match is realized thanks to a radially varying impedance. From theoretical studies, nonhomogeneous media lead to evanescent modes (Salmon, 1946a,b; Leach, 1996): below a certain cutoff frequency, no energy is propagated because the propagation factor γ is real (see Sec. II). To our knowledge, such evanescent waves propagation in the middle ear has never been investigated. We feel this problem needs further experimental and theoretical investigation, with some priority.

VI. SUMMARY

This research has been focused on implementing a time-domain model of the middle ear, and is a significant extension of the frequency domain model of Puria and Allen (1998) in that the ear canal and the tympanic membrane are represented by a parallel complex of delay lines; their interface is complex due to the inclination of the membrane in the canal, due to different speeds of sound in both media, and due to the impedance mismatch between the two. Using a particular spatial distribution, the transmission from one to the other is simulated and the reflection is computed using an exponential TM impedance distribution. Simulations show good consistency with the model.

The TM system is then coupled to a time-domain version of the previously published lumped-parameter model of the middle ear (Zwislocki, 1962; Puria and Allen, 1998). Ossicles parameters are adjusted manually from published values in order to match experimental data in three cases: disarticulated stapes, drained cochlea, and intact ear. The final adjusted model shows relevant similarities with impedance-based and ossicle motion measures.

The model does an excellent job of capturing the essence of the middle ear responses. Some discrepancies still exist at high frequencies, especially in our highly simplified cochlear load impedance, and underline the need for some refinements.

Relative to previous studies, the new model offers two main differences. While previous works have modeled the middle ear in the time domain (Funnel and Laszlo, 1978; Rabbitt and Holmes, 1986; Funnel *et al.*, 1987), as far as we know, no such "wave" models have been published. Our wave formulation allows for significantly more detail in the TM model and for the simulation of wave propagation on its surface. With respect to previous detailed models, the main interest is to be able to directly observe the spreading of the wave on the TM surface by modeling its time response. Second, the TM distributed model enables better simulation of its complex interface with the ear canal and its space-varying

impedance distribution. We show that it interfaces to existing OC models easily, and offers interesting agreement with experimental data. The apparent relevance of the space-varying impedance hypothesis leads to the need for investigation of a TM cutoff and any resulting evanescent modes.

APPENDIX: OSSICULAR CHAIN MODEL

Following are details about each section, as required to implement the OC model.

1. Malleus and incus

Those two ports are in the form of Fig. 1, with

$$\text{Malleus: } Y = \frac{R_{im}C_{im}s}{1 + R_{im}C_{im}s} \quad \text{and} \quad Z = M_m,$$

$$\text{Incus: } Y = \frac{R_{is}C_{is}s}{1 + R_{is}C_{is}s} \quad \text{and} \quad Z = M_i.$$

The filters being of order greater than 1, final states of the filters need to be stored from one time step to the other to ensure proper signal processing. Note that this study has approximated the two ossicles characteristic impedance by $\sqrt{M/C}$, i.e., neglecting the joint resistance in the computation, which is a good approximation at low frequencies (in the long wavelength limit). From Campbell (1922), both two ports are low-pass filters having a cutoff frequency of $f_0 = 1/2\pi\sqrt{MC}$. For the chosen parameters of our simulations

$$\text{Malleus: } f_0 = 17.71 \text{ kHz},$$

$$\text{Incus: } f_0 = 238.3 \text{ kHz}.$$

We can then conclude that the incus does not have much influence over the propagating wave in the considered frequency range. However, the malleus filter has probably a slight influence over very high frequencies (over 15 kHz), which the authors have actually observed when playing with the parameters: increasing the malleus mass does lower the cutoff frequency and brings some perturbations to the input impedance at very high frequencies.

2. Middle-ear attic ligaments

For the attic ligaments we include the anterior malleal ligament, anchoring the manubrium and TM at the tympanic ring to the attic, and the posterior incudal ligament, anchoring the incus to the attic. These two ligaments are represented by a single compliance, in series with the malleus mass. The attic ligaments account for the residual stiffness for the short-circuit boundary condition corresponding to the disarticulated stapes experiment (Puria and Allen, 1998).

This two port is special because it does not have any explicit shunt admittance: consequently, the usual "ABCD matrix" method cannot be directly applied, because the method requires the estimate of the characteristic impedance for each section (i.e., Eq. (6)). The resolution of this has been the addition of a shunt (rotational) mass, making the first two port a high-pass filter (Campbell, 1922). Our reasoning as to why this element must be a mass is that every horn results from a spatially varying impedance having a cutoff, leading to a high-pass behavior, presumably with a low cutoff frequency (e.g., well below 1 kHz). This mass element could then be associated with the TM rather than the OC. This is an interesting problem, related to that described in Rosowski *et al.* (1988). Its value is equally difficult to estimate, thus it has been adjusted empirically: the authors have decided to simply decrease the cutoff frequency so that the mass does not disrupt the system response. The mass is then set to very high values, resulting in a cutoff frequency around 100 Hz. This ensures a proper reflection of low frequencies, without a noticeable attenuation above 1 kHz. This problem clearly requires an experimental investigation, well beyond the scope of the present modeling effort. The present implementation of this two port is to be seen as a stepping stone toward the resolution of the TM cutoff, rather than an actual solution.

3. Stapes and cochlear load

The cochlear impedance model is taken from Lynch *et al.* (1982), modeled by a cochlear resistor R_c, shunted by a series mass and resistor M_0 and R_0, representing the helicotrema. The round window is represented by a series compliance C_{rw}. The model published by Lynch *et al.* (1982) also includes another mass, in series with the shunt association which had little utility, and thus was dropped, as by Puria and Allen (1991). The final load is treated as a simple one port, since the wave transmitted into the cochlea is not developed in the present analysis. A single frequency-dependent reflection coefficient $\mathcal{R}_{\text{load}}(s)$ is derived from this load impedance:

$$Z_{\text{load}}(s) = M_s s + \frac{1}{C_{al}s} + R_{al} + \frac{R_c(M_0 s + R_0)}{R_c + M_0 s + R_0} + \frac{1}{C_{rw}s}, \tag{A1}$$

giving a cochlear reflectance of

$$\mathcal{R}_{\text{load}}(s) = \frac{Z_{\text{load}}(s) - z_{0_i}}{Z_{\text{load}}(s) + z_{0_i}}, \tag{A2}$$

where z_{0_i} is the incus characteristic impedance, $\sqrt{M_i/C_{is}}$; in fact, since the stapes is included in the final load impedance, the reflection coefficient must be computed using the incus impedance, not the stapes. The reflected sample is obtained by filtering the forward output of the incus two port. Note that the round window compliance is not expected to have much influence here since, according to both Puria and Allen (1998) and Lynch *et al.* (1982), its value is around 50 times lower than the annular ligament compliance. It is included in the model, however, because it is associated with an actual physical element of the ear and could have a more serious influence in further work on pathologic ears.

4. Sampling rate

With the lumped-parameter approach, delay is no longer simulated by shifting samples along the transmission line, but directly by filtering. It is then critical to set the filters parameters properly so that the phase shift of each two port is physically relevant. A critical factor is the sampling rate.

TABLE I. Model parameter values (1/2).

Parameter	Value
Sampling rate [Hz]	$f_s=1.96\times10^6$
Ear canal:	
Canal length [cm]	$L_{ec}=0.15$
Spatial period [cm]	$\Delta x_{ec}=0.018$
Speed of sound [cm/s]	$\mathcal{C}_{ec}=34{,}720$
Canal diameter [cm]	$D_{ec}=0.46$
Canal cross-section area [cm^2]	$A_{ec}=0.17$
Characteristic impedance [g/(cm^4 s)]	$Z^a_{ec}=245.0$
Tympanic membrane:	
TM diameter [cm]	$D_{tm}=0.72$
TM area [cm^2]	$A_{tm}=0.41$
Number of annuli	$N=71$
Ratio of speeds	$q=3.4$
Wave speed [cm/s]	$\mathcal{C}_{tm}=10{,}212$
Spatial period [cm]	$\Delta r_{tm}=0.005$
Angle with respect to canal axis [°]	$\alpha=40$
Impedance ratio	$\text{Ratio}_{\text{TM}}=7.5$
Impedance at TM output [g/(cm^4 s)]	$Z^a_{tm_{out}}=1837.3$
Ossicular chain:	
Middle ear cavity cross-section area [cm^2]	$A_{me}=A_{tm}=0.41$ (0.41)
Ossicular chain lever ratio	$N_{\ell r}=2$ (2)
Malleus ligaments resistance [dyne s/cm]	$R^m_m=60$ (30, 2)
Malleus ligaments compliance [cm/dyn]	$C^m_m=1.90\times10^{-6}$ $(8.89\times10^{-7}, 2.14)$
Malleus ligaments inertial mass [g]	$M^m_{mh}=1.39$
Malleus mass [g]	$M^m_m=0.0028$ (0.0028)
IM-joint resistance [dyne s/cm]	$R^m_{im}=7.50$ (7.50)
IM-joint compliance [cm/dyn]	$C^m_{im}=2.91\times10^{-8}$ $(2.87\times10^{-8}, 1.01)$
Malleus impedance [g/(cm^4 s)]	$Z^m_m=308.86$

TABLE II. Model parameter values (2/2).

Parameter	Value
Incus mass [g]	$M^m_i=8.25\times10^{-4}$ (8.25×10^{-4})
IS-joint resistance [dyne s/cm]	$R^m_{is}=37.5$ (75, 2)
IS-joint compliance [cm/dyn]	$C^m_{is}=5.41\times10^{-10}$ $(5.39\times10^{-10}, 1.00)$
Stapes mass [g/cm^4]	$M^a_s=24.75$ (24.75)
Annular ligament resistance [dyne s/cm^5]	$R^a_{al}=7.5\times10^5$(7.5$\times$ 10^5)
Annular ligament compliance [cm^5/dyn]	$C^a_{al}=3.81\times10^{-11}$ $(2.52\times10^{-11}, 1.51)$
Stapes footplate area[cm^2]	$A_{fp}=0.0126$ (0.0126)
Helicotrema mass [g/cm^4]	$M^a_0=3750$ (16875, 0.22)
Helicotrema resistance [dyne s/cm^5]	$R^a_0=7.5\times10^5$ (2.10$\times$ 10^6, 0.36)
Cochlea input impedance [dyne s/cm^5]	$R^a_c=9.0\times10^6$ (9.0$\times$ 10^6)
Round window compliance [cm^5/dyn]	$C^a_{rw}=1.33\times10^{-9}$ (1.33×10^{-9})

Typical delays in the ossicles are in the order of the microsecond (Puria and Allen, 1998): proper handling of such small delays requires high sampling rates, namely, in the order of 2 MHz (more precisely, $f_s=1.96$ MHz). Once derived, f_s is used to compute the spatial sampling on the TM and in the canal. This may be a detail where some future simplification might occur. For now we remain detailed, so that we can accurately simulate the fine structure of various pathologies, rather than be fast (a possible goal of such future models).

5. Choice of parameters

Model parameters are summarized in Tables I and II. Acoustical units are specified with superscript a, mechanical units with superscript m, following the representation used by Puria and Allen (1998). The various middle ear model parameters are inspired by the values from Puria and Allen (1998), and Lynch *et al.* (1982) for the cochlea model (i.e., M_0, R_0, R_c, and C_{rw}). Since these two studies do not take into account the TM impedance transformation ratio, Ratio_{TM}, their values have been scaled by this factor before being used in our model. Note that for the two joint compliances, C_{im} and C_{is}, the formula used in our work (Eq. (6)) differs slightly from the ones mentioned by Puria and Allen (1998, Eqs. 16 and 17), because we do not allow an IM joint slippage factor, and our IM joint is located on the left side of the OC transformer. Our approach, as in (Puria and Allen, 1998), is to start from the earlier values and then to adjust them to best match experimental impedance data of Allen (1986). The parameters were adjusted manually, and the influence of each one is intuitively understood. In Tables I and II, previously published (scaled) values are indicated in brackets, along with the relative variation we have applied to get the values that are actually used in the model (not indicated when values are equal). In most cases, this factor lies between 0.5 and 2, which shows that our model is consistent with the previous analysis. Only helicotrema parameters had to be decreased significantly to obtain a proper match, as discussed further.

It is interesting to question the difference between the malleus and incus mass. The values reported here are very close to the ones from Puria and Allen (1998). In their paper, they compare their malleus mass to experimental measurements by Lynch (1981); Lynch *et al.* (1994): their mass is smaller by a factor of around 4. According to them, "this factor might be accounted for by the smaller size of animals used in [their] study in comparison to those of Lynch *et al.* (1994). Another explanation might be the differences in the radius of gyration between Lynch's measurements and those in [their] study." However, the ratio of malleus mass to incus mass is consistent with the values from Lynch. No reference is made to the stiffness values. It is probably relevant to say that the incus is smaller than the malleus, hence the difference of mass.

Allen, J. B. (**1986**). "Measurement of eardrum acoustic impedance," in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 44–51.

Allen, J. B. (**2001**). "Nonlinear cochlear signal processing," in *Physiology of the Ear*, 2nd ed. (Singular Thomson Learning, San Diego), Chap. 19, pp. 393–442.
Allen, J. B. (**2003**). "Amplitude compression in hearing aids," in *MIT Encyclopedia of Communication Disorders* (MIT Press, Boston), Chap. Part IV, pp. 413–423.
Allen, J. B., Jeng, P. S., and Levitt, H. (**2005**). "Evaluating human middle ear function via an acoustic power assessment," J. Rehabil. Res. Dev. **42**(4), 63–78.
Bekesy, G. v., and Rosenblith, W. A. (**1951**). "The mechanical properties of the ear," in *Handbook of Experimental Psychology*, edited by S. S. Stevens (Wiley, New York).
Beranek, L. L. (**1954**). *Acoustics* (McGraw-Hill, New York).
Bilbao, S. D. (2001). "Wave and scattering methods for the numerical integration of partial differential equations," Ph.D. thesis, Stanford University.
Brillouin, L. (**1953**). *Wave Propagation in Periodic Structure*, 2nd ed. (Dover, New York).
Campbell, G. A. (**1922**). "Physical theory of the electric wave-filter," Bell Syst. Tech. J. **1**(2), 1–32.
Fay, J. P. (2001). "Cat eardrum mechanics," Ph.D. thesis, Stanford University.
Fay, J. P., Puria, S., and Steele, C. R. (**2002**). "Cat eardrum response mechanics," *Presented at the Calladine Festschrift*, edited by S. Pellegrino (Kluwer, The Netherlands).
Fay, J. P., Puria, S., and Steele, C. R. (**2006**). "The discordant eardrum," Proc. Natl. Acad. Sci. U.S.A. **103**(52), 19743–19748.
Funnel, W. R. J., and Decraemer, W. F. (**1996**). "On the incorporation of moire shape measurements in finite element models of the cat eardrum," J. Acoust. Soc. Am. **100**(2), 925–932. Part 1.
Funnel, W. R. J., Decraemer, W. F., and Khanna, S. M. (**1987**). "On the damped frequency response of a finite-element model of the cat eardrum," J. Acoust. Soc. Am. **81**(6), 1851–1859.
Funnel, W. R. J., and Laszlo, C. A. (**1978**). "Modeling of the cat eardrum as a thin shell using the finite-element method," J. Acoust. Soc. Am. **63**, 1461–1466.
Goode, R. L., and Killion, M. C. (**1987**). "The middle ear from the standpoint of the surgeon and the acoustician," *Paper presented at the 113th meeting of the Acoustical Society of America*, Indianapolis, Indiana.
Guinan, Jr., J. J., and Peake, W. T. (**1967**). "Middle-ear characteristics of anesthetized cats," J. Acoust. Soc. Am. **41**(5), 1237–1261.
Kelly, J. L., and Lochbaum, C. C. (**1963**). "Speech synthesis," in *Proceedings of the Fourth International Congress on Acoustics* (ICA, Copenhagen), Chap. 42, pp. 1–4.
Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**2000**). *Fundamentals of Acoustics*, 4th ed. (Wiley, New York).
Leach, W. M. (**1996**). "A two-port analogous circuit and spice model for salmon's family of acoustic horns," J. Acoust. Soc. Am. **99**(3), 1459–1464.
Lim, D. J. (**1968a**). "Tympanic membrane electron microscopic observation part i: Pars tensa," Acta Oto-Laryngol. **66**, 181–198.
Lim, D. J. (**1968b**). "Tympanic membrane part ii: Pars flaccida," Acta Oto-Laryngol. **66**, 515–532.
Lynch, T. J., III. (1981). "Signal processing by the cat middle-ear: Admittance and transmission, measurements and model," Ph.D. thesis, Massachusetts Institute of Technology.
Lynch, T. J., III, Nedzelnitsky, V., and Peake, W. T. (**1982**). "Input impedance of the cochlea in cat," J. Acoust. Soc. Am. **72**(1), 108–130.
Lynch, T. J., III, Peake, W. T., and Rosowski, J. J. (**1994**). "Measurements of the acoustic input impedance of cat ears: 10 Hz to 20 kHz," J. Acoust. Soc. Am. **96**, 2184–2209.
Olson, E. S. (**1998**). "Observing middle ear and inner ear mechanics with novel intracochlear pressure sensors," J. Acoust. Soc. Am. **103**, 3445–3463.
Oppenheim, A. V., Schafer, R. W., and Buck, J. R. (**1999**). *Discrete-Time Signal Processing*, 2nd ed. (Prentice–Hall, Upper Saddle River, NJ).
Parent, P. (2005). "Wave model of the tympanic membrane," Master's thesis, University of Illinois at Urbana-Champaign.
Parent, P., and Allen, J. B. (**2006**). "Tympanic membrane model," Abstract presented at the *29th annual midwinter meeting of the Association for Research in Oto-laryngology*, Baltimore.
Puria, S. (**1991**). "A physical model for the middle ear cavity," J. Acoust. Soc. Am. **89** Suppl., 1864.
Puria, S. (**2003**). "Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustics emissions," J. Acoust. Soc. Am. **113**(5), 2773–2789.
Puria, S., and Allen, J. B. (**1991**). "A parametric study of cochlear input impedance," J. Acoust. Soc. Am. **89**, 287–309.
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**(6), 3463–3481.
Puria, S., and Fay, J. (**2001**). "Human eardrum and ossicles: Two-port matrix measurements," *Abstract presented at the Association for Research in Otolaryngology*.
Puria, S., and Rosowski, J. J. (**1996**). "Measurement of reverse transmission in the human middle ear: Preliminary results," in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. L. Hecht-Poinar (World Scientific, Singapore).
Rabbitt, R. D., and Holmes, M. H. (**1986**). "A fibrous dynamic continuum model of the tympanic membrane," J. Acoust. Soc. Am. **80**(6), 1716–1728.
Rabbitt, R. D., and Holmes, M. H. (**1988**). "Three-dimensional acoustic waves in the ear canal and their interaction with the tympanic membrane," J. Acoust. Soc. Am. **83**, 1064–1080.
Rosowski, J. J., Carney, L. H., and Peake, W. T. (**1988**). "The radiation impedance of the external ear of cat: Measurements and applications," J. Acoust. Soc. Am. **84**(5), 1695–1708.
Rosowski, J. J., Cosme, F., Ravicz, M. E., and Rodgers, M. T. (**2006**). "Real-time opto-electronic holographic measurements of the sound-induced displacements of tympanic membranes," Paper presented at the *29th annual midwinter meeting of the Association for Research in Otolaryngology*, Baltimore.
Rosowski, J. J., Davis, P. J., Merchant, S. N., Donahue, K. M., and Coltrara, M. D. (**1990**). "Cadaver middle ears as models for living ears: Comparisons of middle-ear input immittance," Ann. Otol. Rhinol. Laryngol. , 403–412.
Salmon, V. (**1946a**). "Generalized plane wave horn theory," J. Acoust. Soc. Am. **17**(3), 199–211.
Salmon, V. (**1946b**). "A new family of horns," J. Acoust. Soc. Am. **17**(3), 212–218.
Sen, D., and Allen, J. B. (**2006**). "Functionality of cochlear micromechanics – as elucidated by the upward spread of masking and two tone suppression," Acoust. Aust. **34**(1), 43–51.
Shaw, E. A. G. (**1977**). "Eardrum representation in middle-ear acoustical networks," J. Acoust. Soc. Am. **62** Suppl. 1, S102.
Shaw, E. A. G., and Stinson, M. R. (**1981**). "Network concepts and energy flow in the human middle ear," J. Acoust. Soc. Am. **69**(S44), S43.
Stinson, M. R. (**1990**). "Revision of estimates of acoustic energy reflectance at the human eardrum," J. Acoust. Soc. Am. **88**(4), 1773–1778.
Stinson, M. R., and Daigle, G. A. (**2005**). "Comparison of an analytic horn equation approach and a boundary element method for the calculation of sound fields in the human ear canal," J. Acoust. Soc. Am. **118**(4), 2405–2411.
Stinson, M. R., and Khanna, S. M. (**1989**). "Sound propagation in the ear canal and coupling to the eardrum, with measurements on model systems," J. Acoust. Soc. Am. **85**(6), 2481–2491.
Stinson, M. R., and Khanna, S. M. (**1994**). "Spatial distribution of sound pressure and energy flow in the ear canals of cats," J. Acoust. Soc. Am. **96**(1), 170–180.
Stinson, M. R., Shaw, E., and Lawton, B. (**1982**). "Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal," J. Acoust. Soc. Am. **72**, 766–773.
Wegel, R. L., and Lane, C. E. (**1924**). "The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear," Phys. Rev. **23**, 266–285.
Zwislocki, J. (**1957**). "Some impedance measurements on normal and pathological ears," J. Acoust. Soc. Am. **29**(12), 1312–1317.
Zwislocki, J. (**1962**). "Analysis of the middle-ear function. Part I: Input impedance," J. Acoust. Soc. Am. **34**(8, Part 2), 1514–1523.

Transmission matrix analysis of the chinchilla middle ear

Jocelyn E. Songer[a)]
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles St., Boston, Massachusetts 02114 and Speech and Hearing Bioscience and Technology, Health Sciences and Technology, Harvard-MIT, Cambridge, Massachusetts 02138

John J. Rosowski
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles St., Boston, Massachusetts 02114, Speech and Hearing Bioscience and Technology, Health Sciences and Technology, Harvard-MIT, Cambridge, Massachusetts 02138 and Department of Otology and Laryngology, Harvard Medical School, Boston, MA 02114

(Received 4 December 2006; revised 8 May 2007; accepted 8 May 2007)

Despite the common use of the chinchilla as an animal model in auditory research, a complete characterization of the chinchilla middle ear using transmission matrix analysis has not been performed. In this paper we describe measurements of middle-ear input admittance and stapes velocity in ears with the middle-ear cavity opened under three conditions: intact tympano-ossicular system and cochlea, after the cochlea has been drained, and after the stapes has been fixed. These measurements, made with stimulus frequencies of 100–8000 Hz, are used to define the transmission matrix parameters of the middle ear and to calculate the cochlear input impedance as well as the middle-ear output impedance. This transmission characterization of the chinchilla middle ear will be useful for modeling auditory sensitivity in the normal and pathological chinchilla ear. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2747157]

PACS number(s): 43.64.Ha, 43.64.Tk, 43.64.Bt, 43.80.Lb [BLM] Pages: 932–942

I. INTRODUCTION

In this paper we describe a two-port model of the normal chinchilla middle ear that is derived from measurements in both intact and altered conditions. A two port is a network with two input terminals and two output terminals. For the middle ear, the input terminals are located at the tympanic membrane and the output terminals are located at the oval window. The assumption we make regarding the middle-ear two port is that all elements within the network are passive and linear; this assumption is supported by previous assessments of middle-ear linearity (Guinan and Peake, 1967; Nedzelnitsky, 1980). Further support for the use of transmission matrices to characterize the mechanics of the middle ear comes from prior studies conducted in other species (Shera and Zweig, 1992; Peake *et al.*, 1992; Puria, 2004, 2003; Voss and Shera, 2004). The middle ear is composed of multiple structures with complex motions, and a detailed analysis of the behavior of the individual components of the middle ear is similarly complex (Zwislocki, 1962; Kringlebotn, 1988; Goode *et al.*, 1994; Koike *et al.*, 2002; Ladak and Funnell, 1996; Gan *et al.*, 2002). An advantage of a two-port analysis of the middle ear compared to other middle-ear models is that the two port provides a complete description of the middle ear that can be characterized using a small set of measurements. A two port is an appropriate solution for this work because we are interested in relating the chinchilla middle ear inputs to its outputs, and we are not interested in the individual elements within the middle ear.

We use a transmission matrix to characterize the middle-ear two port. To determine the transmission matrix parameters we rely on a fundamental property of transmission matrices, that they are independent of the load (Desoer and Kuh, 1969; Shera and Zweig, 1992). In this case, that means that the matrix is independent of both the ear-canal load and the cochlear load. We exploit this property to estimate the transmission matrix parameters by manipulating the cochlear load, i.e., draining the cochlea and fixing the stapes footplate. These two processes are considered manipulations of the load on the middle ear and as such do not alter the transmission matrix characterization of the middle ear. We then estimate the transmission matrix parameters with measurements of the middle-ear input admittance ($Y_{ME}=\frac{U_{TM}}{P_{TM}}$, where U_{TM} refers to the volume velocity of the tympanic membrane and P_{TM} refers to the pressure at the tympanic membrane) and the middle-ear transfer ratio ($H_p=\frac{V_s}{P_{TM}}$, where V_s refers to stapes velocity) in three different conditions: with the inner ear intact (similar to previously reported measurements (Rosowski *et al.*, 2006; Songer and Rosowski, 2006)), with the cochlea drained, and with the stapes fixed. These measurements and manipulations allow us to solve for the four parameters necessary to characterize the two port.

The two port is then used, with the data from the intact ear, to characterize the normal cochlear input impedance as well as the middle-ear output impedance. As we will see, due to limitations of the manipulations used to create the stapes fixation, the model is most accurate for frequencies below 1500 Hz.

The goal of this work is to improve the characterization of the chinchilla middle ear using a two-port representation. An example of the usefulness of this type of middle-ear

[a)]Author to whom correspondence should be addressed. Electronic mail: jocelyns@paradoxical.net

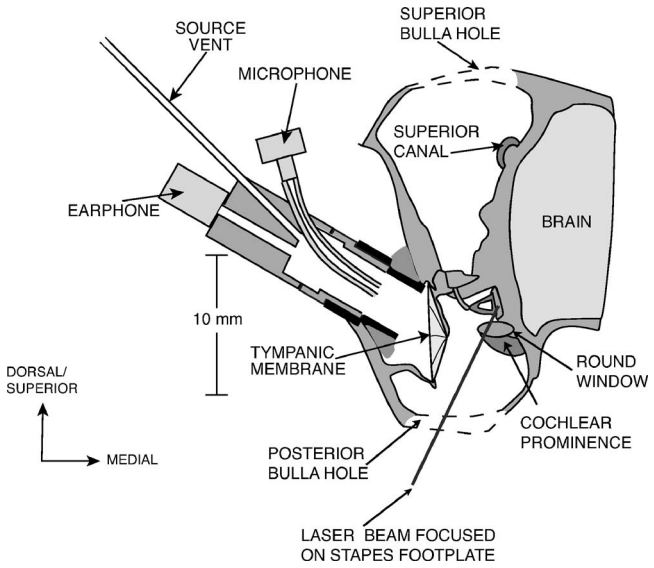


FIG. 1. A schematic of a coronal section through the chinchilla ear illustrating: accesses to the middle ear through the two surgically produced bulla holes, the placement of the earphone and the laser focused on the stapes footplate.

analysis is illustrated in a companion paper where the two port and the estimate of cochlear impedance determined here are used in a mechano-acoustic model of the effect of superior canal dehiscence (an opening in the bony labyrinth surrounding the superior semicircular canal) on hearing in chinchilla (Songer and Rosowski, 2007).

II. METHODS

A. Animal preparation

We collected data from each experimental condition from a total of five chinchillas (partial data sets were acquired in three additional animals). Chinchillas were chosen as the animal model for this study because they are commonly used in auditory research, their range of hearing is similar to that of humans (100 Hz–20 kHz at 30 dB sound pressure level (SPL) (Miller, 1970)), and we are interested in developing a chinchilla model of auditory sensitivity both in the normal ear and in response to superior canal dehiscence (Songer and Rosowski, 2007).

Experiments were carried out in accordance with the National Institute of Health Animal Care and Use Guidelines and were approved by the Massachusetts Eye and Ear Infirmary animal care and use committee. The surgical preparation was previously described (Songer and Rosowski, 2005, 2006) and is summarized here. The chinchillas were anesthetized with pentobarbital (50 mg/kg) and ketamine (40 mg/kg) with boosters every 2 h, or as necessary. After the animal was anesthetized and tracheotomized the superior bulla was exposed and a hole was made in it to visualize the medial surface of the tympanic membrane (Fig. 1). The tendon of the tensor tympani was then severed with a small knife and the stapedius muscle was paralyzed by sectioning the facial nerve between its genu and the stapedial nerve branch (Songer and Rosowski, 2005).

A hole was then introduced into the posterior bulla to visualize the round window and the lenticular process of the incus. Both the superior and posterior bullar holes were left wide open (open cavity) throughout the measurements. The bony wall to which the stapedius tendon attaches was then resected between the tendon attachment, the horizontal canal, and the round window in order to visualize the stapes footplate and crura. Three to six reflective beads (each 50 μm in diameter) were then placed on the stapes footplate. Any fluid accumulated in or around the footplate was removed using fine absorbent paper points prior to measurement of the stapes velocity using laser-Doppler vibrometry (LDV).

The pinna and cartilaginous ear canal were resected and the bony ear canal shortened. A probe-tube microphone and calibrated sound source were coupled to the ear canal via a short brass tube that was cemented into the remnant of the bony ear canal. A broadband stimulus with a low-frequency emphasis (log chirp) with frequency components at 11 Hz intervals between 11 Hz and 24 Hz was used as the stimulus. The limited high-frequency output of our sound source restricted the usable frequency range to below 8000 Hz. At frequencies below 100 Hz noise became a problem, especially in the laser measurements. Both the sound pressure at the tympanic membrane (P_{TM}) and the stapes velocity (V_s) were measured in response to log-chirp stimuli at three levels (covering a 20 dB range from 74 to 94 dB SPL) to check for repeatability and linearity. Minor deviations in linearity were observed near 160 Hz. These deviations are consistent with previous work (Ruggero *et al.*, 1996; Rosowski *et al.*, 2006; Songer and Rosowski, 2006) and are usually attributed to an inner ear nonlinearity.

After a series of base line measurements, the cochlea was drained by creating a large hole (0.5 mm $\times$ 1.0 mm) in the bony cochlear prominence near the round window (Fig. 1) and then using paper points to remove fluid from the cochlea. After draining the cochlea, the round window was perforated in order to ensure continued drainage. In some animals a hole in the superior semicircular canal was introduced prior to creating the hole in the cochlear prominence. After the cochlea was drained, the middle-ear input admittance (Y_{ME}) and middle-ear transfer function (H_p) were measured repeatedly to test for preparation stability.

The stapes footplate was then fixed by applying either dental cement or Superglue gel ®, forming a connection between the stapes and the adjacent petrous bone. The cement or glue was allowed to dry for at least 10 min. After the glue had hardened, a series of measurements of Y_{ME} were made. H_p measurements were not possible after fixation because the glue covered the stapes footplate, which was the former location of the reflectors used for measurements of V_s. In two cases new reflectors were placed either on the glue fixing the footplate or on the crus of the stapes to estimate V_s after fixation. To evaluate the effectiveness of the fixation we also measured the cochlear potential at the round window using a silver wire electrode both before and after fixation in three ears using procedures described previously (Songer and Rosowski, 2005). We were also able to compare our measurements of the magnitude of the admittance with the stapes fixed ($|Y^F_{\mathrm{ME}}|$) in these ears to that measured in ears in which

the inner ear was drained prior to fixation and demonstrate that after fixation the state of the fluid in the inner ear does not impact our measurements of Y_{ME}^{F}.

B. Instrumentation

1. Source and admittance measurements

The sound stimulus and instrumentation used have been described previously (Songer and Rosowski, 2005, 2006) but will be summarized here. The stimulus we used was a train of 100 log chirps with frequency components between 11 Hz and 24 kHz. The stimulus was generated with a series of LabView scripts and presented to the ear with a hearing-aid earphone (Knowles ED-1913) that was part of a sound-source/measurement assembly with a high acoustic output impedance (Ravicz *et al.*, 1992; Rosowski *et al.*, 2006). A hearing-aid microphone (Knowles EK-3027) was built into the sound source/measurement assembly and used to measure the resultant sound pressure in the ear canal.

The microphone data recorded at the tympanic membrane in conjunction with the Norton equivalent of the source were utilized to calculate the middle-ear input admittance (Y_{ME}). The Norton equivalent of the acoustic source was determined using previously described procedures (Songer and Rosowski, 2006; Lynch *et al.*, 1994) where the response of the sound source is characterized using a series of acoustic loads of known impedance. To determine Y_{ME}, the volume velocity of the Norton equivalent representation of the sound source (U_{src}) was divided by the sound pressure at the tympanic membrane (P_{TM}) and then the equivalent admittance of the source (Y_{src}) was subtracted from the total admittance

$$Y_{\mathrm{ME}} = \frac{U_{\mathrm{src}}}{P_{\mathrm{TM}}} - Y_{\mathrm{src}}. \tag{1}$$

The admittance was then corrected for the residual ear canal and earphone-coupler space using a transmission-line correction (Lynch *et al.*, 1994; Voss and Shera, 2004) where the length of the residual canal and sound coupler was estimated to be 6.0 mm and the radius of the tube was 2.4 mm (Songer and Rosowski, 2006).

2. Stapes velocity measurements

Laser-Doppler vibrometry (LDV) was used to detect the sound-induced stapes velocity (V_s) based on the Doppler shift of the light reflected from glass beads placed on the stapes footplate according to a previously described procedure (Songer and Rosowski, 2006). We used a single point LDV from Polytech PI (OFV 5000) to measure V_s. The voltage output from the vibrometer is proportional to the velocity of the moving object and the sensitivity of the LDV system was determined through comparison with a reference accelerometer. A micromanipulator was used to focus and direct the laser light on the stapes footplate. The approximate measurement angle was between 50 and 60° relative to horizontal. No corrections for the angle were performed. Where necessary, V_s was converted to a volume velocity (U_s) by multiplying it by the nominal area of the stapes footplate ($A_s = 1.98$ mm^2 (Vrettakos *et al.*, 1988)).[1]

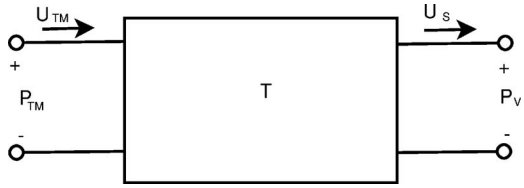


FIG. 2. The mechano-acoustic two-port network used to describe the middle ear of the chinchilla. P_{TM} is the sound pressure at the tympanic membrane, U_{TM} is the volume velocity of the tympanic membrane, P_V is the sound pressure in the vestibule (and equal to P_O in the normal condition), and U_S is the stapes volume velocity. T represents the transmission matrix with matrix elements of A, B, C, and D.

The middle-ear transfer function was defined as $H_p = \frac{V_s}{P_{\mathrm{TM}}}$ where V_s was measured as described above and P_{TM} is the pressure measured in the ear canal and was calculated from the microphone voltage recorded in the ear canal, and converted to Pascals using a previous calibration performed against a reference microphone.

All of the data presented in this paper were normalized for measurement system gains and attenuations (Songer and Rosowski, 2005, 2006). Measurements of microphone voltage and stapes velocity were made for stimulus levels of 74, 84, and 94 dB SPL in each ear and in each condition, except for the stapes fixed condition in which velocity measurements were not possible. Over the majority of the frequency range described here (100–8000 Hz) both the microphone signal and the laser signal were generally linear with a signal that exceeded the noise floor by at least 10 dB.

III. DEFINITION OF TRANSMISSION MATRIX ELEMENTS

A. Matrix in the normal ear

The transmission matrix (T) we use to describe the middle ear is illustrated in Fig. 2, and defined by Eq. (2) where P_{TM} is the pressure at the tympanic membrane, U_{TM} is the volume velocity of the tympanic membrane, P_O is the sound pressure produced by the middle ear at its output, and U_S is the stapes volume velocity. In the normal condition P_O is equivalent to the sound pressure within the cochlear vestibule, P_V. We have defined both the volume velocity going into the middle ear (U_{TM}) and the stapes volume velocity (U_S) leaving the two port as positive (Fig. 2). The transmission matrix elements A, B, C, and D are four complex matrix elements that can be used to characterize the relationship between the variables at the ports, i.e.

$$\begin{bmatrix} P_{\mathrm{TM}} \\ U_{\mathrm{TM}} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} P_O \\ U_S \end{bmatrix}. \tag{2}$$

In the matrix equation above, we can directly measure three of the four input/output variables: P_{TM}, U_{TM} and U_S.

B. Matrix with the cochlea drained

We assume that by draining the cochlea, we reduce the pressure on the medial side of the stapes to zero, thereby

setting P_O to zero and simplifying the matrix to Eq. (3), where P_{TM}^D, U_{TM}^D, and U_S^D are response variables measured with the cochlea drained:

$$\begin{bmatrix} P_{\mathrm{TM}}^D \\ U_{\mathrm{TM}}^D \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} 0 \\ U_S^D \end{bmatrix}. \tag{3}$$

Determining the resultant equations for P_{TM} and U_{TM} allows us to solve for both B and D in terms of the measured quantities: Y_{ME}^D and H_p^D, where the superscript D marks values measured with the cochlea drained, and the area of the stapes footplate (A_s)

$$B = \frac{P_{\mathrm{TM}}^D}{U_S^D} = \frac{1}{H_p^D A_s}, \tag{4}$$

$$D = \frac{U_{\mathrm{TM}}^D}{U_S^D} = \frac{Y_{\mathrm{ME}}^D}{H_p^D A_s}. \tag{5}$$

C. Matrix with fixed stapes

By fixing the stapes with cement we are able to greatly diminish $|U_S|$. If we assume that the resulting stapes velocity is negligible compared to the normal velocity, Eq. (2) reduces to Eq. (6) where the superscript F denotes measured response variables after fixation

$$\begin{bmatrix} P_{\mathrm{TM}}^F \\ U_{\mathrm{TM}}^F \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} P_O^F \\ 0 \end{bmatrix}, \tag{6}$$

where P_O^F is the effective middle-ear output pressure acting on the fixed stapes, and is equivalent to the pressure acting on a rigid inner-ear load. Equation (6) leads us to equations for both A and C

$$A = \frac{P_{\mathrm{TM}}^F}{P_O^F}, \tag{7}$$

$$B = \frac{U_{\mathrm{TM}}^F}{P_O^F}. \tag{8}$$

Since we do not measure P_O^F directly, we can calculate the relationship between A and C as

$$A = C\frac{P_{\mathrm{TM}}^F}{U_{\mathrm{TM}}^F} = \frac{C}{Y_{\mathrm{ME}}^F}. \tag{9}$$

Using Eqs. (4), (5), and (7) as well as the reciprocity theorem, we can calculate the fourth matrix parameter (Shera and Zweig, 1992). The reciprocity theorem states that in any passive linear network the forward transfer admittance is equal to the reverse transfer admittance or, in the case of our network representation,

$$AD - BC = 1. \tag{10}$$

Assuming reciprocity, we can solve for C independently of A and P_O as demonstrated in Eq. (11)

$$\frac{1}{C} = \left(D\frac{P_{\mathrm{TM}}^F}{U_{\mathrm{TM}}^F} - B \right) = \frac{D}{Y_{\mathrm{ME}}^F} - B. \tag{11}$$

D. Definition of transfer matrix and load impedances

Using our measurements of Y_{ME} and H_p in conjunction with calculations of the transmission matrix parameters and the impedance of our source, we can calculate the cochlear input impedance (Z_c), the middle-ear input impedance $\left(Z_{\mathrm{ME}} = \frac{1}{Y_{\mathrm{ME}}}\right)$ and the middle-ear output impedance (Z_{out}).

Z_c is the load on the middle ear in the intact state. Using the transmission matrix as well as our measurements of Y_{ME} and H_p we can solve for Z_c.[2]

$$Z_c = \frac{P_V}{U_S} = \frac{DP_{\mathrm{TM}} - BU_{\mathrm{TM}}}{-CP_{\mathrm{TM}} + AU_{\mathrm{TM}}} = \frac{B - \frac{D}{Y_{\mathrm{ME}}}}{\frac{C}{Y_{\mathrm{ME}} - A}}. \tag{12}$$

In addition to determining Z_{ME} from our measurements of Y_{ME}, we can also calculate Z_{ME} using the matrix in Eq. (2) as demonstrated in Eq. (13)

$$Z_{\mathrm{ME}} = \frac{P_{\mathrm{TM}}}{U_{\mathrm{TM}}} = \frac{AP_O + BU_S}{CP_O + DU_S} = \frac{AZ_c + B}{CS + D}. \tag{13}$$

Z_{out} is equal to P_O/U_S in the reverse direction and describes the load the middle ear would place on a sound source in the inner ear. Z_{out} is calculated using the two-port model with the input port loaded by the ear canal and the earphone impedances (the impedance of the sound source within the ear). This load is denoted as Z_{src} (Songer and Rosowski, 2006) and the equation for Z_{out} follows:

$$Z_{\mathrm{out}} = \frac{DZ_{\mathrm{src}} + B}{CZ_{\mathrm{src}} + A}. \tag{14}$$

IV. RESULTS OF THE PHYSIOLOGICAL MEASUREMENTS

Data were collected from eight ears. In three of the ears, draining the cochlea led to significant fluid accumulation in the oval-window niche, occluding our reflectors and preventing us from obtaining stapes-velocity data. The data used to calculate the transmission matrix parameters are from the five ears in which we were able to successfully make measurements in all three experimental conditions: 1) TM, ossicles and cochlea intact, 2) cochlea drained, and 3) stapes fixed.

To compare our measurements to previous results, the data are presented in terms of transfer functions: the middle-ear input admittance ($Y_{\mathrm{ME}} = U_{\mathrm{TM}}/P_{\mathrm{TM}}$) and the middle-ear transfer function ($H_p = V_S/P_{\mathrm{TM}}$).

A. Normal ear Y_{TM} and H_p

The data from five ears were acquired in the intact state. The mean magnitude and phase of Y_{ME} for the normal ear is illustrated in Fig. 3. This mean result is similar to other measurements of Y_{ME} with open middle-ear spaces (Songer and Rosowski, 2006; Rosowski *et al.*, 2006) and is presented with units of acoustic Siemens ($1S = 1m^3 - s^{-1} - \mathrm{Pa}^{-1}$).

There are three prominent features in the mean $|Y_{\mathrm{ME}}|$ data from the normal or intact ear: a minimum near 160 Hz,

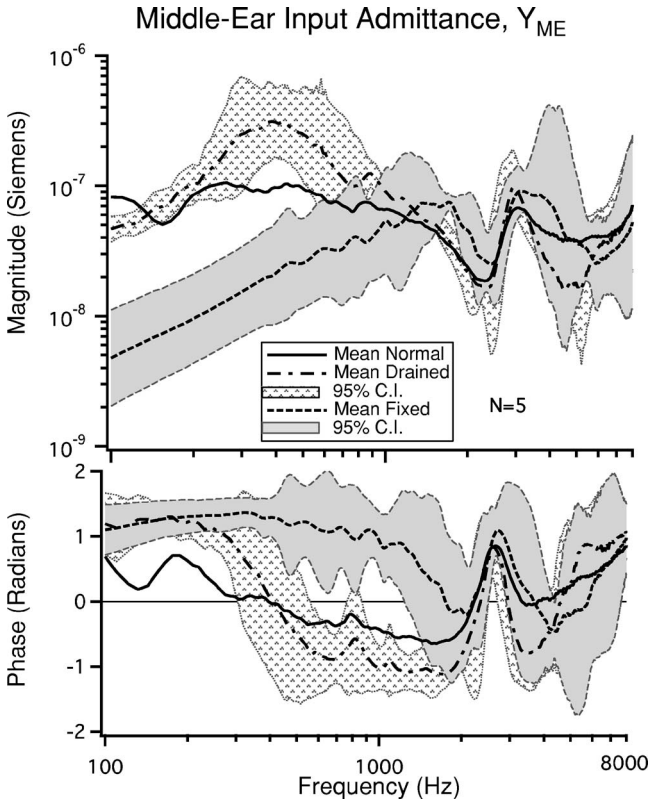


FIG. 3. The mean magnitude and phase of Y_{ME} for each condition: normal (intact), cochlea drained, and stapes fixed (N=5). The 95% confidence intervals (C.I.s) are shown for both the drained and stapes fixed conditions. Figure 8 shows the 95% C.I. for Y_{ME} the intact condition.

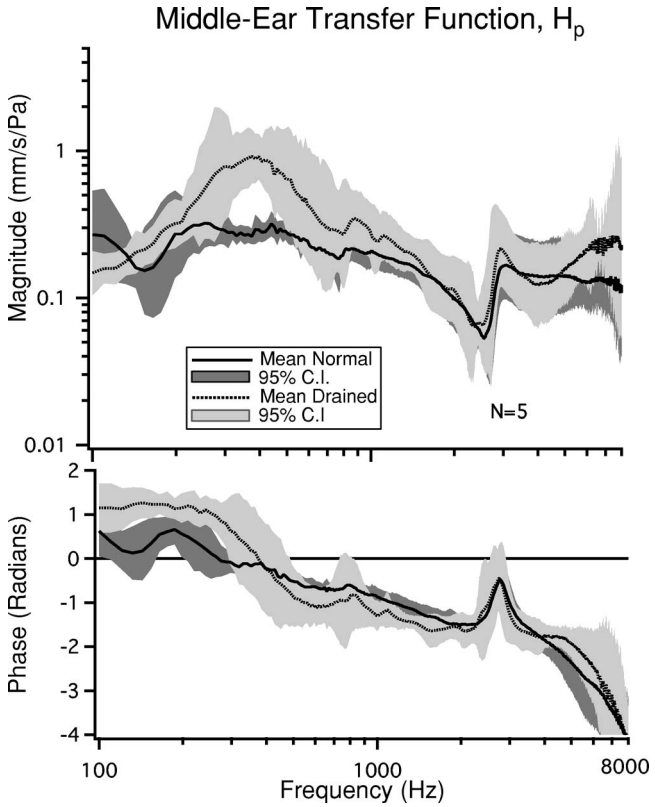


FIG. 4. The mean magnitude and phase for the H_p measured in each condition. The 95% confidence intervals (C.I.) for both the normal and drained states are also illustrated.

a second minimum near 2400 Hz, and a maximum near 3 kHz. These extrema in the Y_{ME} magnitude are associated with changes in the phase of Y_{ME}. The prominent features in the mean are present in all five ears, with slight differences in the sharpness of the extrema. Minima in middle-ear transmission near 160 Hz have been observed previously in measurements of Y_{ME} (Songer and Rosowski, 2006) as well as in measures of cochlear microphonic (Dallos, 1970) and cochlear potential (Songer and Rosowski, 2005). It is hypothesized that these minima are due to the influence of the helicotrema on low-frequency cochlear input impedance (Dallos, 1970; Songer and Rosowski, 2006; Rosowski *et al.*, 2006). The minimum near 2400 Hz is attributed to an antiresonance produced by the open holes in the bulla. Similar minima have been observed previously in chinchilla (Songer and Rosowski, 2006; Rosowski *et al.*, 2006) and other species (Moller, 1965; Guinan and Peake, 1967;Ravicz *et al.*, 1992).

The mean magnitude and phase of the H_p in five intact ears is illustrated in Fig. 4. The frequency response of H_p is similar to the frequency response observed for Y_{ME} and to previous measurements (Songer and Rosowski, 2006; Ruggero *et al.*, 1990): minima in $|H_p|$ occur at 160 and 2400 Hz and a maximum is apparent near 3000 Hz. The phase is primarily decreasing throughout the frequency range except near 160 Hz, and between 2400 and 3000 Hz where there are phase changes associated with the extrema in the magnitude. Above 5 kHz the phase of H_p decreases more rapidly.

B. Drained cochlea Y_{ME} and H_p

The mean magnitude and phase as well as the 95% confidence intervals[3] for Y_{ME} after the cochlea has been drained (Y^D_{ME}) are illustrated in Fig. 3. The mean $|Y^D_{ME}|$ has a maximum at 400 Hz and a minimum near 2400 Hz. The phase at 100 Hz is near 1 rad and remains nearly constant until approximately 300 Hz, above which there is a rapid phase transition to near -1 rads where it remains until peaking near 2400 Hz. Between 100 and 300 Hz, the upward slope in magnitude and the positive angle between 1 and $\pi/2$ radians suggest a stiffness controlled admittance. Between 600 and 2000 Hz, the downward slope in magnitude and the angles between -1 and $-\pi/2$ suggest a mass-controlled admittance. The minimum near 2400 Hz is consistent with the open-cavity resonance observed in the normal ear.

The H_p after draining the inner ear (H^D_p) has many similarities to Y^D_{ME}. Draining the inner ear results in a mean increase in $|H^D_p|$ for frequencies below 1400 Hz (Fig. 4). This increase has a peak near 400 Hz, below which the slope is positive and the phase is between 1 and $\pi/2$ rad and above which the slope is negative and the phase is between -1 and $-\pi/2$ rad. For frequencies above 1500 Hz there is little difference between the H_p measured in the intact ear and that observed after draining. The minimum near 2400 Hz associated with the open-cavity resonance is still observed. For frequencies above 5 kHz the phase decreases rapidly.

Both Y^D_{ME} and H^D_p show a low-frequency resonance (peak near 400 Hz with a π change in phase) that is consistent with measurements of middle-ear mechanics associated

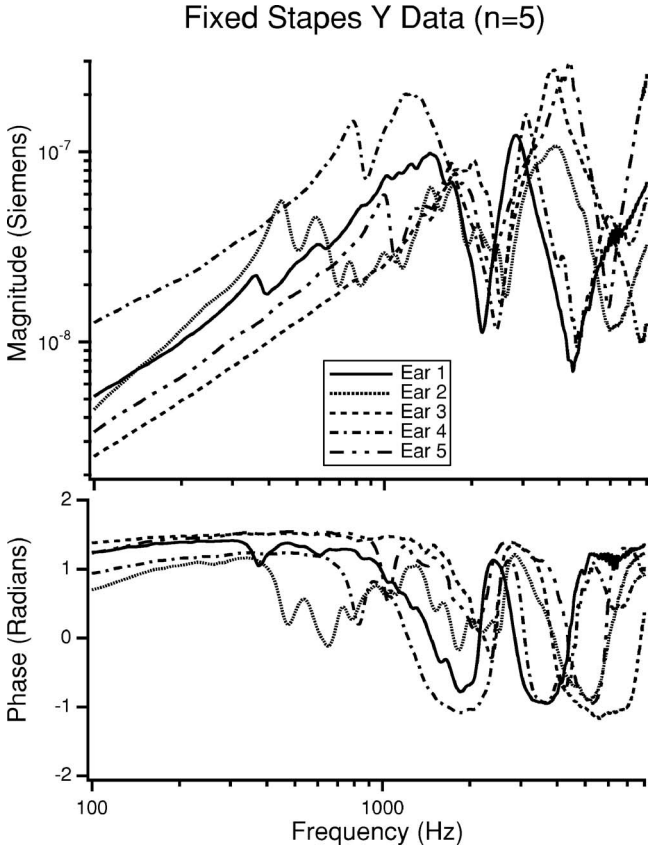


FIG. 5. The magnitude and phase of Y_{ME} after the stapes has been fixed ($n=5$). Ear 3 has the lowest admittance and is considered to have the best fixation. Note the sharp anti resonance near 2400 Hz which is associated with the open bulla holes in our preparation.

with cochlear draining (Moller, 1965; Lynch, 1981; Allen, 1986; Puria and Allen, 1998). The appearance of this resonance after cochlear draining is consistent with the removal of cochlear damping (Moller, 1965; Rosowski *et al.*, 2006).

C. Fixed stapes Y_{ME} and H_p

The mean magnitude and phase of Y_{ME} after fixing the stapes (Y_{ME}^F) is illustrated in Fig. 3. For low frequencies, fixing the stapes results in a decrease in $|Y_{ME}|$ and leads to a phase close to $\pi/2$ rad. This change in Y_{ME} is consistent with previous work (Moller, 1965; Lynch, 1981) that suggests the middle-ear input impedance becomes stiffer as a result of fixation for frequencies below 1000 Hz. The mean $|Y_{ME}^F|$ increases proportionately with frequency above 100 Hz, reaching a peak near 1500 Hz. The peak is followed by a minimum at 2400 Hz. The phase at 100 Hz is between 1 and 1.5 rad and remains nearly constant until near 1500 Hz above which it rapidly transitions to a value less than 0 rad.

The $|Y_{ME}^F|$ varies between the five ears (Fig. 5). This variation is likely due to differences in the effectiveness of fixation. The more complete the fixation, the smaller the low-frequency admittance is expected to be. According to this criterion, ear 3 has the most complete fixation of the ears we tested.

It was not possible to measure H_p after fixation of the stapes because the glue overlapped the location of stapes reflectors. In two cases, we placed additional reflectors on the crus to record the stapes velocity after a partial fixation and observed a 20–25 dB decrease in H_p for frequencies below 500 Hz. Between 500 and 1500 Hz the effect of fixation diminishes and above 1500 Hz, the change in H_p as a result of fixation was less evident.

Since measurements of stapes velocity were not routinely possible after fixation, we also measured the cochlear potential (CP) normalized by $P_{TM}\left(G=\mathrm{CP}/P_{TM}\right)$ before and after fixation in three ears. The fixation in these ears resulted in a 10–30 dB reduction in G for frequencies below 1500 Hz. Above 2 kHz there was little difference between G before and after fixation. This indicates that our fixation[4] is most effective for frequencies below 1500 Hz.

D. Summary of physiological findings

Figure 3 illustrates Y_{ME} in response to each of the experimental conditions: intact ear, cochlea drained, and stapes fixed. Above 1500 Hz, the difference between the means of the conditions is not significant. This suggests that for frequencies above 1500 Hz the state of the inner ear is not affecting our measurements. For frequencies below 1500 Hz we observe differences in Y_{ME} both in magnitude and phase in response to our manipulations as described above. In both the intact ear and after the cochlea is drained, H_p (Fig. 4) has a similar frequency dependence to Y_{ME}.

V. COMPUTATION OF TRANSMISSION MATRIX PARAMETERS

The transmission matrix parameters (A, B, C, and D) are calculated for each of the five ears based on the individual Y_{ME} and H_p measurements. The mean and 95% confidence intervals of the A, B, C, and D calculations from the five ears are our first estimate of these four parameter values and their variance. Our measurements of both G and H_p after fixation suggest that the assumed large reduction in stapes motion after our fixation procedure is only strictly valid for frequencies below 1500 Hz. Since our estimates of A, C, Z_{ME}, Z_{out}, and Z_c are dependent on the fixation measurements and this assumption, they are most reliable for frequencies below 1500 Hz.

A. Two-port description as transmission matrix

B and D are calculated from measurements taken after the cochlea is drained. B is the ratio of the drained P_{TM} to drained U_S and has units of $Pa\ s\ m^{-3}$. The mean value of our five estimates of B along with the 95% confidence intervals are illustrated in Fig. 6(B). The most prominent feature of $|B|$ is a broad minimum near 350 Hz which is associated with a transition in phase from near $-\pi/2$ to near $\pi/2$. As discussed earlier, this minimum is related to a middle-ear resonance that is observed when the cochlear damping is removed.

D is the ratio of the drained U_{TM} to the drained U_S and is dimensionless. The mean value of the five estimates of D is illustrated in Fig. 6(D) along with the 95% confidence intervals. For frequencies below 1500 Hz the average $|D|$ is about 165, with a phase angle near zero. There is a peak in

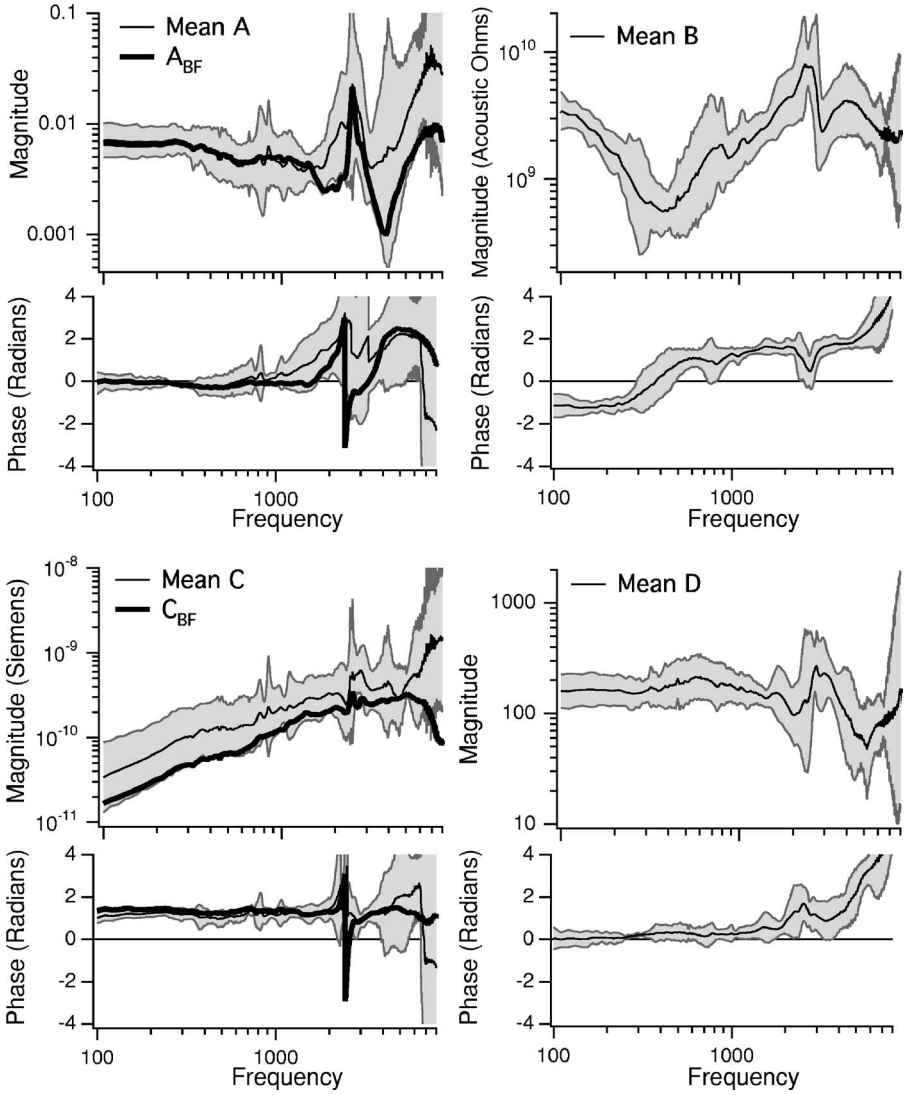


FIG. 6. The mean values of the transmission parameters (A, B, C, and D) along with the 95% confidence intervals estimated from five ears. The mean A and C are plotted as well as the parameters calculated from the best fixation measurements, A_{bf} and C_{bf}. The calculations of B and D do not depend on fixation, therefore only the mean and C.I.s are illustrated.

the average $|D|$ near 3 kHz with a value of 250, with a phase angle at the peak of about $\pi/2$. At higher frequencies $|D|$ decreases and the phase angle grows.

The transmission parameters A and C are calculated using B and D as well as measurements taken after fixing the stapes. C is defined in Eq. (11) and has units of admittance ($m^3-s^{-1}-\mathrm{Pa}^{-1}$). The mean and 95% confidence intervals of the five individual C computations are illustrated in Fig. 6(C). $|C|$ is roughly proportional to frequency and the phase angle of C is near $\pi/2$ over the majority of the frequency range. C was also calculated using the measurements from the fixation that produced the largest reduction in $|Y_{ME}|$ (ear 3) and the mean B and D. We refer to this estimate as C_{bf}, where the bf stands for the "best fixation." The C_{bf} calculation is intended to reduce the effect of potential procedural problems (ineffective fixation) on model predictions and results. $|C_{bf}|$ is about 60% of the mean $|C|$ across the entire frequency range. At 2500 Hz there is a phase discontinuity in C_{bf} due to residual effects of the sharp bulla-hole-cavity anti resonance observed in the stapes-fixed data from ear 3 (Fig. 5).

A is unitless and calculated using Eq. (7) and Eq. (9). The mean and 95% confidence intervals of the five individual estimates of A are illustrated in Fig. 6(A). Below 1500 Hz, A is fairly constant with a magnitude near 0.007, a phase near zero and is approximately $1/D$. Above 1500 Hz, there is an increase in both the magnitude and phase of A. Due to the variable quality of our fixations for frequencies above 1500 Hz, we also calculated A_{bf} using the best fixation measurements of ear 3 (also plotted in Fig. 6(A)). For frequencies below 1500 Hz both A and A_{bf} are nearly identical. Above 1500 Hz, however, $|A_{bf}|$ is smaller than $|A|$, has a more prominent peak near 2500 Hz and a phase discontinuity at the same frequency. The sharp phase transition in A_{bf} near 2500 Hz can again be attributed to a strong cavity anti resonance observed in the fixed-stapes data from ear 3 (Fig. 5).

B. Cochlear input impedance

The mean and 95% confidence interval of the five Z_c estimates calculated using Eq. (12) are illustrated in Fig. 7. The magnitude and phase are relatively constant with frequency. The Z_c calculated from the best fixation, Z_c^{bf}, is also illustrated. One feature of our calculation of Z_c is a peak near 160 Hz and an associated change in phase. This peak occurs

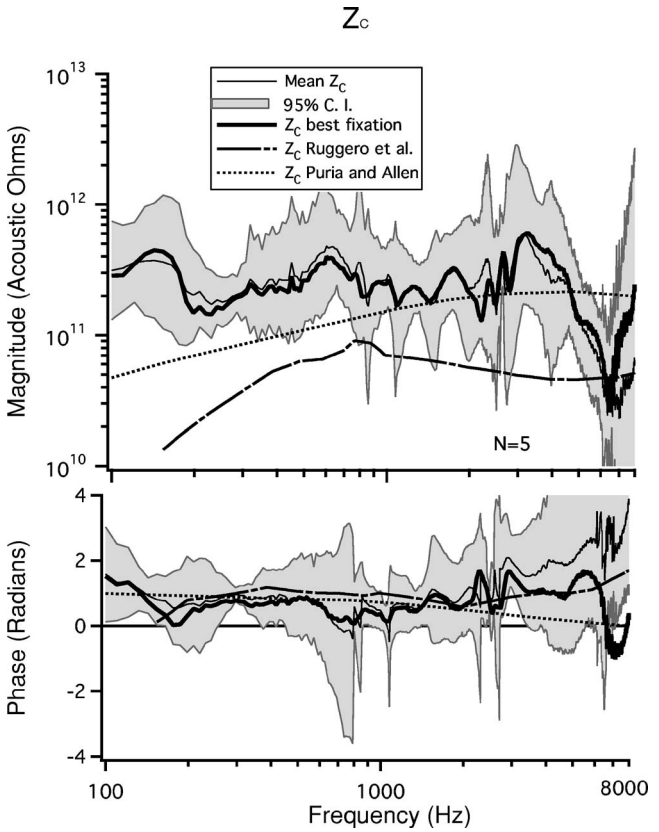


FIG. 7. The mean calculated $Z_c(n=5)$ with 95% confidence intervals. Z_c^{bf}, the Z_c calculated from the best fixation, as well as Z_c estimated by Ruggero *et al.* (1990) and Puria and Allen (1991) are also illustrated.

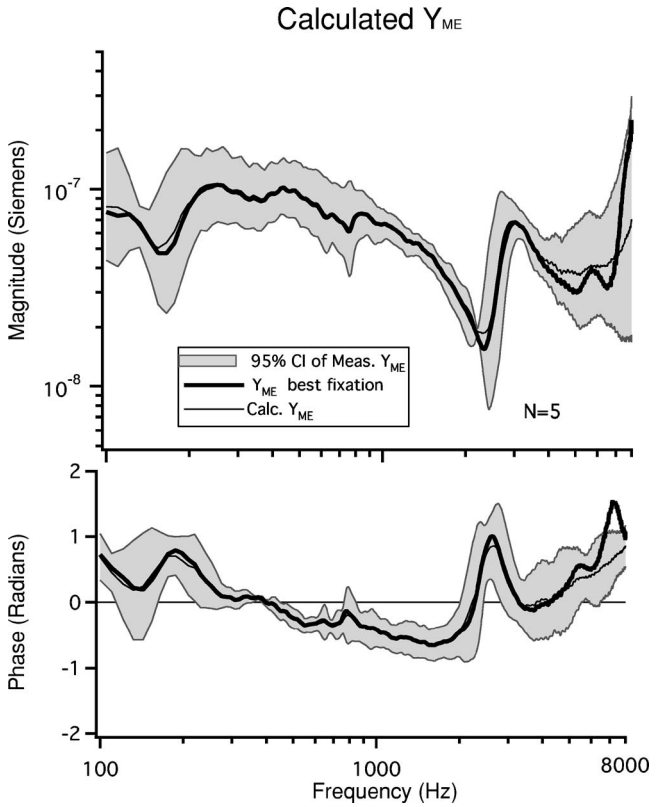


FIG. 8. The 95% confidence intervals of the measured Y_{ME} data as well as the mean calculated Y_{ME} and the Y_{ME}^{bf} are illustrated. Both calculations of Y_{ME} are within the 95% C.I. of the measurement data for frequencies below 6 kHz.

at the same frequency as the observed minimum in $|Y_{ME}|$. The minimum in $|Y_{ME}|$ has been attributed to the effect of the helicotrema in previous studies (Songer and Rosowski, 2006; Dallos, 1970).

Calculations of Z_c presented by Ruggero *et al.* (Ruggero *et al.*, 1990) and Puria and Allen (Puria and Allen, 1991) are also illustrated in Fig. 7. Our calculations of Z_c differ from those presented by Ruggero in three major aspects: 1) our estimate has a magnitude that is 3–4 times larger than that of Ruggero *et al.*, 2) our measurements do not exhibit the low-frequency roll off observed in the Ruggero *et al.* estimates and 3) our measurements have much more fine structure. The Puria and Allen estimates of Z_c have a magnitude that is generally similar to our estimate of Z_c, especially at frequencies between 1000 and 5000 Hz. The Puria and Allen estimates have a low-frequency roll off that is slower than that seen in the Ruggero *et al.* estimates but still suggest lower impedance magnitude than our estimates at frequencies less than 800 Hz. The three estimates of the phase of Z_c are similar, with the largest differences observed at the highest and lowest frequencies.

C. Calculations of Y_{ME}, H_p and middle-ear output impedance

The mean of the Y_{ME} calculated from the five individual sets of impedance parameter estimates ($Y_{ME}=\frac{1}{Z_{ME}}$ Eq. (13)) and the Y_{ME} calculated from A_{bf}, C_{bf}, Z_{bf} and the mean A and B are compared to the 95% confidence intervals from our measurements of Y_{ME} in Fig. 8. The magnitude and phase of the two computations and our measurement of Y_{ME} show excellent agreement for frequencies between 100 Hz and 6 kHz, thereby demonstrating self-consistency within the calculations. The Y_{ME} calculated with the best fixation measurement, Y_{ME}^{bf} shows some differences from the measurements and the other estimate at frequencies above 6 kHz.

A second consistency test is to compare our calculated middle-ear transfer function (H_p) to the measured values of H_p as illustrated in Fig. 9. The calculations of H_p are a good fit to the data, however, for frequencies above 2 kHz, the calculation of H_p using the best fixation measurement is closer in magnitude and phase to the measured data. Thus, our calculations of both Y_{ME} and H_p suggest that our transmission matrix parameter values are internally consistent. A slight improvement in agreement between the measured and predicted values is achieved at high frequencies using parameters calculated based on the best fixation measurements, however the Y_{ME} estimated with the best-fixation data differs slightly from the mean estimates above 6 kHz.

The mean calculated Z_{out} along with the 95% confidence intervals are illustrated in Fig. 10. Z_{out} has a magnitude that is inversely proportional to frequency and a phase near $\pi/2$ for frequencies below 1 kHz. This is consistent with Z_{out} being compliance dominated within this frequency range. Z_{out} is dependent on the impedance of the sound source (Z_{src}) as well the ear canal since they make up the load on the middle ear in the reverse direction (from the oval window). The relationship between Z_{out} and Z_{src} is explicitly stated in

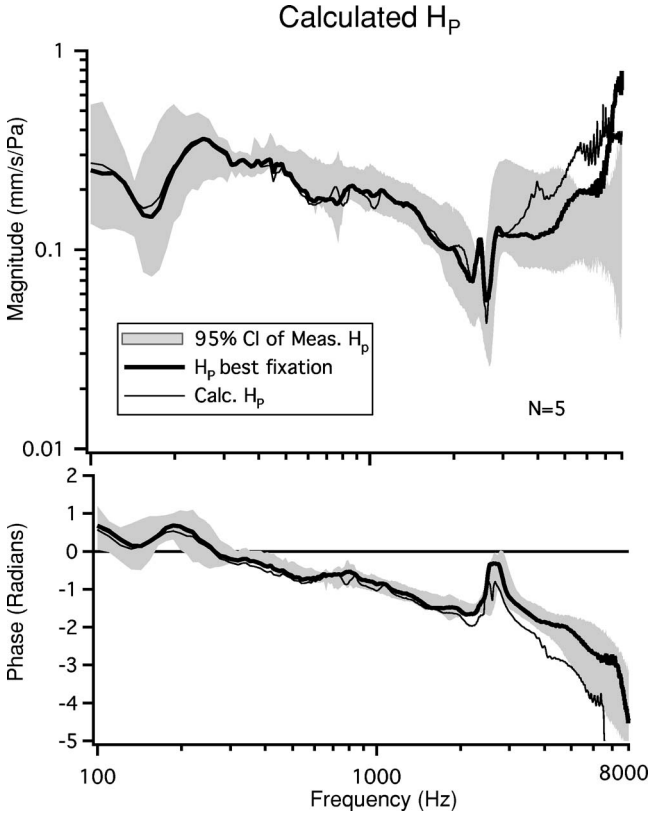


FIG. 9. The 95% confidence intervals of the measured H_p, the mean calculated H_p, and the H_p^{bf}. The H_p^{bf} data have a better fit to the measured H_p data at high frequencies.

Eq. (14) and Z_{src} is illustrated in Fig. 10. For frequencies below 2400 Hz, Z_{out} generally appears to be a scaled version of Z_{src} where the mean scaling factor between 100 and 8000 Hz is 82 dB.

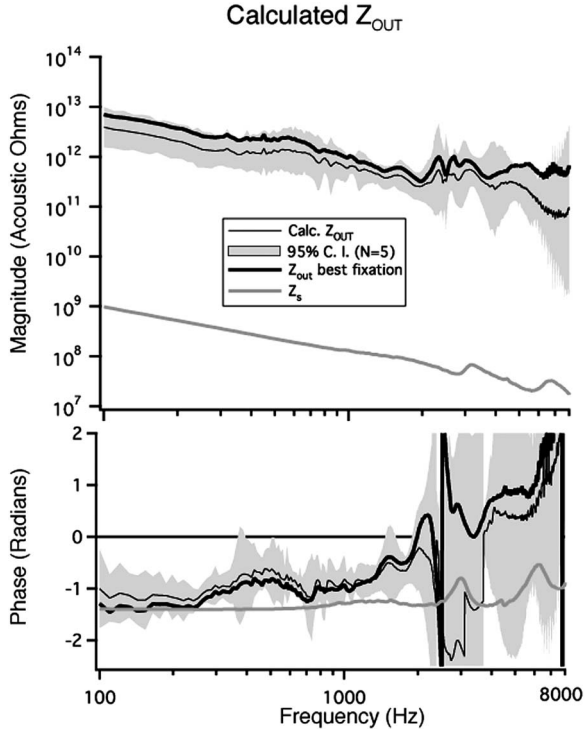


FIG. 10. The mean calculated Z_{out} with 95% confidence intervals as well as Z_{out}^{bf}. Z_{src} (Z_s) is also illustrated.

The Z_{out} calculated from the best fixation measurement (Z_{out}^{bf} is also plotted. $|Z_{out}^{bf}|$ is larger than $|Z_{out}|$ over the entire frequency range. The phase of $|Z_{out}^{bf}|$ has some rapid phase cycling near 2400 Hz that is associated with a strong antiresonance in the original measurements. Since Z_{out} is a driving-point impedance and the middle-ear is a passive system, we expect the angle of Z_{out} to vary between $\pm\pi/2$. However, artifact in the handling of the phase data associated with small changes in the cavity resonance near 2400 Hz causes the angle to exceed these bounds at some frequencies above 2400 Hz.

VI. DISCUSSION OF MODEL ESTIMATES OF NORMAL EAR FUNCTION

A. Comparison to other Z_c estimates

Both Ruggero *et al.* (Ruggero *et al.*, 1990) and Puria and Allen (Puria and Allen, 1991) report values of Z_c in chinchilla. The Z_c estimate published by Ruggero *et al.* (Ruggero *et al.*, 1990) was calculated based on measurements of V_s and previously published measurements of P_V. Between 200 and 2000 Hz our estimate of Z_c is, on average, 11.6 dB (3.8 times) greater than that published by Ruggero *et al.* Puria and Allen use a spatially varying transmission line model together with measurements of the dimensions of scala vestibuli and scala tympani to estimate the value of Z_c (Puria and Allen, 1991) and arrive at a value of Z_c that is similar to our estimate at frequencies between 1000 and 5000 Hz, and as much as 12.9 dB (4.4 times) greater than that reported by Ruggero *et al.* For frequencies between 400 Hz and 2 kHz the three estimates of Z_c, though they differ in magnitude, exhibit similar frequency responses both in magnitude and phase. One factor that would reduce the magnitude of our estimate of Z_c by 4–6 dB (1.6–2 times), would be if we used a correction for the angle of our observed stapes motion. The stapes velocities in our study were measured with an approximate angle of 50–60° between our laser and the axis of "piston-like" motion of the stapes. The cosine of these angles describes the error in estimating the piston-like motion. However, given that the motion of the stapes in humans, cats, and gerbils (Decraemer and Khanna, 2004) and probably chinchilla becomes complicated at frequencies above a few kHz, such a correction is problematic (Chien *et al.*, 2006).

While the Puria and Allen Z_c model does suggest a decrease in Z_c magnitude above 2 kHz, the minimum they compute is at 15 kHz instead of the minima between 6 and 8 kHz that we observe. The phase above 3 kHz that we predict from the transmission matrix analysis is similar to that reported by Ruggero *et al.* Below 500 Hz the data from Ruggero *et al.* exhibit a positive slope of 7 dB/octave, whereas our estimates suggest a relatively flat frequency response in this range.

B. Utility of Z_{out} estimates

The transmission matrix analysis of the middle ear allows us to estimate Z_{out} for the normal chinchilla ear in addition to Z_{ME} and Z_c. Z_{out} has been calculated using similar methodologies in human cadaveric temporal bones (Puria, 2003). The Z_{out} in human temporal bones (Puria, 2003) ex-

hibits a similar frequency dependence to that presented here but with a magnitude of 10^{10} mks ohms near 2 kHz (Puria, 2003), which is two orders of magnitude lower than our estimated chinchilla Z_{out}. A potential difference between the measurements made in this study and the observations in temporal bones could be from the acoustic source. Z_{out} can be strongly affected by the impedance of the sound source in the ear canal (Voss and Shera, 2004).

Measurements of Z_c in conjunction with measurements of Z_{out} can be used to estimate the stapes reflection coefficient (Puria, 2003) and the round-trip middle-ear gain (Voss and Shera, 2004) to provide further insight into otoacoustic emission (OAE) generation mechanisms. In addition to the utility of calculations of Z_{out} in improving our understanding of OAEs, our calculated measurements of Z_{out} in chinchilla can be used to develop a better understanding of the middle-ear load on the inner ear in response to bone-conducted sound stimuli (Songer, 2006).

C. Benefits and limitations of transmission matrix characterization of the middle ear

Transmission matrix analysis has been used to describe the middle ears of cats (Voss and Shera, 2004) as well as human cadaveric temporal bones (Puria, 2003). The transmission matrix characterization of the chinchilla middle ear reported here has many similarities to those reported previously, but differs in some aspects. In this report we did not rely on driving the middle ear "in reverse" with OAEs. Instead, we focused on the forward transmission of the middle ear. This led us to use fixation as one of our experimental conditions. As discussed previously, the degree of fixation varied between ears (Fig. 5) and was most effective at frequencies less than 1.5 kHz. Despite these drawbacks, the predicted values of Z_c and Z_{out} appear reasonable in terms of their frequency dependence. Additional tests of the high-frequency ($>$1500 Hz) characteristics of the chinchilla middle-ear two port will be useful in applications where the high-frequency responses of the middle ear are critically important and can be implemented by directly measuring the pressure in the vestibule instead of relying on measurements of stapes fixation. However, cases where the low-frequency responses are of greatest interest, such as superior canal dehiscence syndrome (Songer and Rosowski, 2007), will benefit from the existing transmission matrix analysis.

The transmission matrix characterization of the middle ear makes no assumptions about the load on either side of the middle ear and can be used in a wide variety of applications ranging from developing models of disease (e.g., superior semicircular canal dehiscence) to improving our understanding of OAEs.

ACKNOWLEDGMENTS

This work has been supported by a NSF graduate student fellowship, NIH training grant, and additional NIH grants. S.N. Merchant, C.A. Shera, W.T. Peake, and M.E. Ravicz provided insights and suggestions. M.L. Wood assisted with data collection and animal surgery.

[1]As discussed later, this simple conversion to volume velocity assumes that the piston-like component of stapes motion dominates the single-point measurement we make. When this assumption was tested in cats and human temporal bones (Heiland *et al.*, 1999; Hato *et al.*, 2003; Decraemer and Khanna, 2004; Chien *et al.*, 2006; Voss *et al.*, 2000) it was generally found to be valid for frequencies less than 2 kHz and somewhat questionable at higher frequencies.

[2]Z_c can be solved for in two additional ways as illustrated below; however, our calculations of Z_c are made using Eq. (12) because these values were empirically determined to be the most stable and repeatable:

$$Z_c = \frac{P_V}{U_S} = \frac{\frac{P_{TM}}{U_S} - B}{A}$$

$$Z_c = \frac{\frac{U_{TM}}{U_S} - D}{C} = \frac{\left.\frac{Y_{ME}}{H_p}\right|_{normal} - D}{C}.$$

[3]The 95% confidence interval (C.I.) can be defined as 1.96 times the standard error and indicates the probability ($p<0.05$) that the actual mean is within the specified range.

[4]We have demonstrated that, despite our application of Superglue to the stapes footplate and surrounding bone, we were unable to create a complete fixation of the stapes. However, in order to maintain a simple nomenclature we continue to refer to this process as fixation and as data with a "fixed stapes."

Allen, J. B. (**1986**). "Measurements of eardrum acoustic impedance," in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. H. Hall, A. Hubbard, S. T. Nealy, and A. Tubis (Springer-Verlag, New York) pp. 44–51.

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.

Dallos, P. (**1970**). "Low-frequency auditory characteristics: Species dependence," J. Acoust. Soc. Am. **48**(2), 489–499.

Decraemer, W. F., and Khanna, S. M. (**2004**). "Measurement, visualization and quantitative, analysis of complete three-dimensional kinematical data sets of human and cat middle ear," in *The Proceedings of the Third International Symposium on Middle Ear Research and Oto-Surgery*, edited by K. Gyo, H. Wada, H. Hato, and T. Koike (World Scientific, Singapore), pp. 3–10.

Desoer, C. A., and Kuh, E. S. (**1969**). *Basic Circuit Theory* (McGraw-Hill, New York).

Gan, R. Z., Sun, Q., Dyer, R. K., Jr., Chang, K. H., and Dormer, K. J. (**2002**). "Three-dimensional modeling of middle-ear biomechanics and its applications," Otol. Neurotol. **23**, 271–280.

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.

Guinan, J. J., Jr., and Peake, W. T. (**1967**). "Middle-ear characteristics of anesthetized cats," J. Acoust. Soc. Am. **41**(5), 1237–1261.

Hato, N., Stenfelt, S., and Goode, R. L. (**2003**). "Three-dimensional stapes footplate motion in human temporal bones," Audiol. Neuro-Otol. **8**(3), 140–152.

Heiland, K. E., Asai, R. L., and Huber, A. M. (**1999**). "A human temporal bone study of stapes footplate movement," Am. J. Otol. **20**, 81–86.

Koike, T., Wada, H., and Kobayashi, T. (**2002**). "Modeling of the human middle ear using the finite-element method," J. Acoust. Soc. Am. **111**(3), 1306–l317.

Kringlebotn, M. (**1988**). "Network model for the human middle ear," Scand. Audiol. **17**, 75–85.

Ladak, J., and Funnell, W. R. J. (**1996**). "Finite-element modeling of the normal and surgically repaired cat middle ear," J. Acoust. Soc. Am. **100**(2), 933–944.

Lynch, T. J., III. (**1981**). Signal Processing by the Cat Middle Ear: Admittance, and Transmission, Measurements and Models. Doctoral thesis, Massachusetts Institute of Technology.

Lynch, T. J., III, Peake, W. T., and Rosowski, J. J. (**1994**). "Measurements of the acoustic input impedance of cat ears: 10 Hz–20 Khz," J. Acoust. Soc. Am. **96**(4), 2184–2209.

Miller, J. D. (**1970**). "Audibility curve of the chinchilla," J. Acoust. Soc. Am. **48**(2), 513–523.

Moller, A. R. (**1965**). "An experimental study of the acoustic impedance of the middle ear and its transmission properties," Acta Oto-Laryngol. **60**, 129–149.
Nedzelnitsky, V. (**1980**). "Sound pressures in the basal turn of the cat cochlea," J. Acoust. Soc. Am. **68**, 1676–1689.
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**(5), 2773–2789.
Puria, S. (**2004**). "Middle-ear two-port measurements in human cadaveric temporal bones: Comparison with models," in *The Proceedings of the Third International Symposium on Middle Ear Research and Oto-Surgery*, edited by K. Gyo, H. Wada, H. Hato, and T. Koike (World Scientific, Singapore), pp. 43–50.
Puria, S., and Allen, J. B. (**1991**). "A parametric study of cochlear input impedance," J. Acoust. Soc. Am. **89**(1), 287–309.
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.
Ravicz, M. E., Rosowski, J. J., and Voigt, H. (**1992**). "Sound-power collection by the auditory periphery of the mongolian gerbil meriones unguiculatus: I. middle-ear input impedance," J. Acoust. Soc. Am. **92**(1), 157–177.
Rosowski, J. J., Ravicz, M. E., and Songer, J. E. (**2006**). "Structures that contribute to middle-ear admittance in chinchilla," J. Comp. Physiol. **192**(12), 1287–1311.
Ruggero, M. A., Rich, N. C., Robles, L., and Shivapuja, B. G. (**1990**). "Middle-ear response in the chinchilla and its relationship to mechanics at the base of the cochlea," J. Acoust. Soc. Am. **87**(4), 1612–1629.
Ruggero, M. A., Rich, N. C., Shivapuja, B. G., and Temchin, A. (**1996**). "Auditory nerve responses to low-frequency tones: Intensity dependence," Aud. Neurosci. **2**, 159–185.
Shera, C. A., and Zweig, G. (**1992**). "Middle-ear phenomeonology: The view from the three windows," J. Acoust. Soc. Am. **1992**, 1356–1370.
Songer, J. E. (**2006**). Superior Semicircular Canal Dehiscence: Auditory Mechanisms. Doctoral thesis, Massachusetts Institute of Technology.
Songer, J. E., and Rosowski, J. J. (**2005**). "The effect of superior canal dehiscence on cochlear potential in response to air-conducted stimuli in chinchilla," Hear. Res. **210**, 53–62.
Songer, J. E., and Rosowski, J. J. (**2006**). "The effect of superior-canal opening on middle-ear input admittance and air-conducted stapes velocity in chinchilla," J. Acoust. Soc. Am. **120**(1), 258–269.
Songer, J. E., and Rosowski, J. J. (**2007**). *A Mechano-Acoustic Model of Superior Canal Dehiscence in Chinchilla*. J. Acoust. Soc. Am. **122**(2)
Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2000**). "Acoustic responses of the human middle ear," Hear. Res. **150**(1–2), 43–69.
Voss, S. E., and Shera, C. A. (**2004**). "Simultaneous measurement of middle-ear input impedance and forward/reverse transmission in cat," J. Acoust. Soc. Am. **16**(4), 2187–2198.
Vrettakos, P. A., Dear, S. P., and Saunders, J. C. (**1988**). "Middle ear structure in the chinchilla: A quantitative study," Am. J. Otolaryngol. **9**, 58–67.
Zwislocki, J. (**1962**). "Analysis of the middle-ear function, Part I. Input impedance," J. Acoust. Soc. Am. **34**(8), 1514–1523.

A mechano-acoustic model of the effect of superior canal dehiscence on hearing in chinchilla

Jocelyn E. Songer[a)]
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles St., Boston, Massachusetts 02114, and Speech and Hearing Bioscience and Technology, Health Sciences and Technology, Harvard-MIT, Cambridge, Massachusetts 02138

John J. Rosowski
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles St., Boston, Massachusetts 02114, Speech and Hearing Bioscience and Technology, Health Sciences and Technology, Harvard-MIT, Cambridge, Massachusetts 02138, and Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02114

(Received 4 December 2006; revised 8 May 2007; accepted 8 May 2007)

Superior canal dehiscence (SCD) is a pathological condition of the ear that can cause a conductive hearing loss. The effect of SCD (a hole in the bony wall of the superior semicircular canal) on chinchilla middle- and inner-ear mechanics is analyzed with a circuit model of the dehiscence. The model is used to predict the effect of dehiscence on auditory sensitivity and mechanics. These predictions are compared to previously published measurements of dehiscence related changes in chinchilla cochlear potential, middle-ear input admittance and stapes velocity. The comparisons show that the model predictions are both qualitatively and quantitatively similar to the physiological results for frequencies where physiologic data are available. The similarity supports the third-window hypothesis of the effect of superior canal dehiscence on auditory sensitivity and mechanics and provides the groundwork for the development of a model that predicts the effect of superior canal dehiscence syndrome on auditory sensitivity and mechanics in humans. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2747158]

PACS number(s): 43.64.Ha, 43.64.Tk, 43.64.Bt, 43.80.Lb [BLM] Pages: 943–951

I. INTRODUCTION

Superior canal dehiscence (SCD) syndrome is an inner-ear disorder in which patients develop a hole (a dehiscence) in the bony wall between the superior semicircular canal and the brain case. Patients afflicted with this disorder present to the clinic with vestibular and/or auditory symptoms. The auditory symptoms include a conductive hearing loss (Minor *et al.*, 2003; Mikulec *et al.*, 2004) and abnormal middle-ear mechanics (Rosowski *et al.*, 2004; Songer and Rosowski, 2006; Chien *et al.*, 2007). It has been proposed that the SCD acts as a pathological "third window" into the inner ear (the round window and oval window are the two normal windows) giving rise to both the auditory and vestibular symptoms associated with SCD syndrome (Rosowski *et al.*, 2004; Minor *et al.*, 1998; Songer and Rosowski, 2005). In general, a third window is an additional compressibility located somewhere within the inner ear that leads to changes in the volume velocities at the two normal cochlear windows (the round and oval windows). Depending on its location, the third window can decrease the stimulus to the cochlea. For example, if the third window is opened on the vestibular side of the inner ear (as in SCD, posterior canal dehiscence (Krombach *et al.*, 2003; Brantberg *et al.*, 2006), and enlarged vestibular aqueduct syndrome (Mimura *et al.*, 2005)) we predict that sound energy will be shunted through the new window away from the cochlea. On the other hand, if the new window were opened in the scala tympani near the round window, it should have little effect on hearing.

The purpose of this paper is to assess mechanisms by which an opening in the normally continuous bony wall of the superior canal (i.e., a dehiscence) can affect hearing sensitivity and middle-ear sound conduction. To address these issues we codify the third-window hypothesis of SCD using a simple circuit model (Fig. 1). Specifically, we create a lumped-element model of the superior semicircular canal (SC) including the SC hole or dehiscence (SCD) and combine it with a two-port model of the middle ear (Songer and Rosowski, 2007) to predict the effect of a SCD on auditory sensitivity and mechanics in response to air-conducted sound stimuli in chinchilla. The effect of a SCD on the model volume velocities and pressures is related to changes in auditory sensitivity and middle-ear mechanics and then compared to previously reported data in chinchilla (Songer and Rosowski, 2005, 2006). The specific goal of this work is to test the third-window hypothesis of SCD-induced hearing loss by comparing model predictions with the physiologic effects of the SCD (Songer and Rosowski, 2005, 2006).

II. ANATOMICALLY BASED MODEL OF Z_{SCD}

The model structure for the chinchilla ear that we used is illustrated in Fig. 1. A model of a dehiscent ear is coupled to a two-port representation of the chinchilla middle ear

[a)]Author to whom correspondence should be addressed. Electronic mail: jocelyns@paradoxical.net

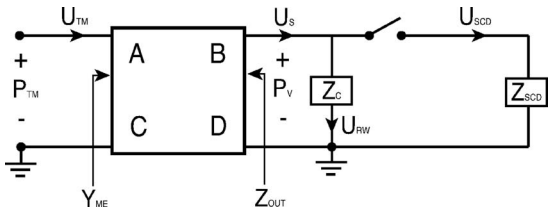


FIG. 1. Schematic of the mechano-acoustic model of the effect of SCD on auditory sensitivity. The middle ear is represented by a two port with four transmission matrix parameters (A, B, C, and D), the cochlea by Z_c, and the SCD by Z_{SCD}. Y_{ME} stands for the middle-ear input admittance and Z_{out} stands for the middle-ear output impedance. P_{TM} and P_v represent the sound pressures at the tympanic membrane and in the vestibule, respectively. U_{TM}, U_S, U_{RW}, and U_{SCD} represent the volume velocities of the tympanic membrane, stapes, round window and within the dehiscent canal.

(Songer and Rosowski, 2007). The structure of the dehiscence model is based upon a codification of the third-window hypothesis, which suggests that the SCD adds a shunt pathway to the inner ear that reduces the stimulus to the cochlea. In order to characterize the impedance of the SCD (Z_{SCD}) we assume that the impedance of this additional pathway is due to fluid flow through the dehiscent SC. Since the wavelengths (in fluid) of the sound stimuli are much much greater than the dimensions of the canal, a lumped-element model characterization of Z_{SCD} is appropriate. To define our lumped-element model, a more complete description of chinchilla SC anatomy was necessary.

A. Anatomical reconstruction

We created a histological reconstruction of one chinchilla inner ear in which the effect of a surgically induced SCD on hearing had been evaluated (Songer and Rosowski, 2006). The reconstruction allowed us to determine the precise size and location of the dehiscence. To create the reconstruction the chinchilla head was fixed, decalcified, embedded in celloidin, and cut into 20 μm sagittal sections. Registered digital photographs of the sections were then created, and every fourth section was imported into *Amira*®, a program designed for three-dimensional (3D) reconstructions of histological images. Within *Amira*, we segmented the bone and fluid filled spaces of the inner ear, and produced a scaled 3D reconstruction of the cochlear and vestibular structures. The 3D reconstruction allowed us to measure precisely the size and location of the dehiscence as well as other relevant canal dimensions illustrated in Fig. 2.

Table I lists the estimates of the anatomical dimensions gathered from the reconstruction of one chinchilla ear. A schematic illustrating how each component was defined is illustrated in Fig. 2. The length of the ampulla (ℓ_{amp}) is the distance between the utricle and the narrowing of the canal. The length of the lateral branch of the SC (ℓ_{lsc}) is the length of the SC between the end of the ampulla and the dehiscence. The length of the dehiscence (ℓ_{dehis}) is the distance from one end of the dehiscence to the other. The length of the wide segment of the medial SC (ℓ_{mscw}) is the length from the beginning of the SC at the utricle to the narrowing of the SC (immediately after the common crus). The length of the medial branch of the SC (ℓ_{msc}) is the length of the canal from the medial end of the dehiscence to the beginning of the common crus. The sum of the lengths ℓ_{msc}, ℓ_{dehis}, and ℓ_{lsc} is 7.7 mm.

The length, diameter and location of the dehiscence can vary. The dimensions affected by these variations are: ℓ_{lsc}, ℓ_{msc}, and ℓ_{dehis}, which are illustrated in Fig. 2 with double-sided arrows. In our surgical preparation ℓ_{lsc} could extend about 5.0 mm from the ampulla to where the SC enters the

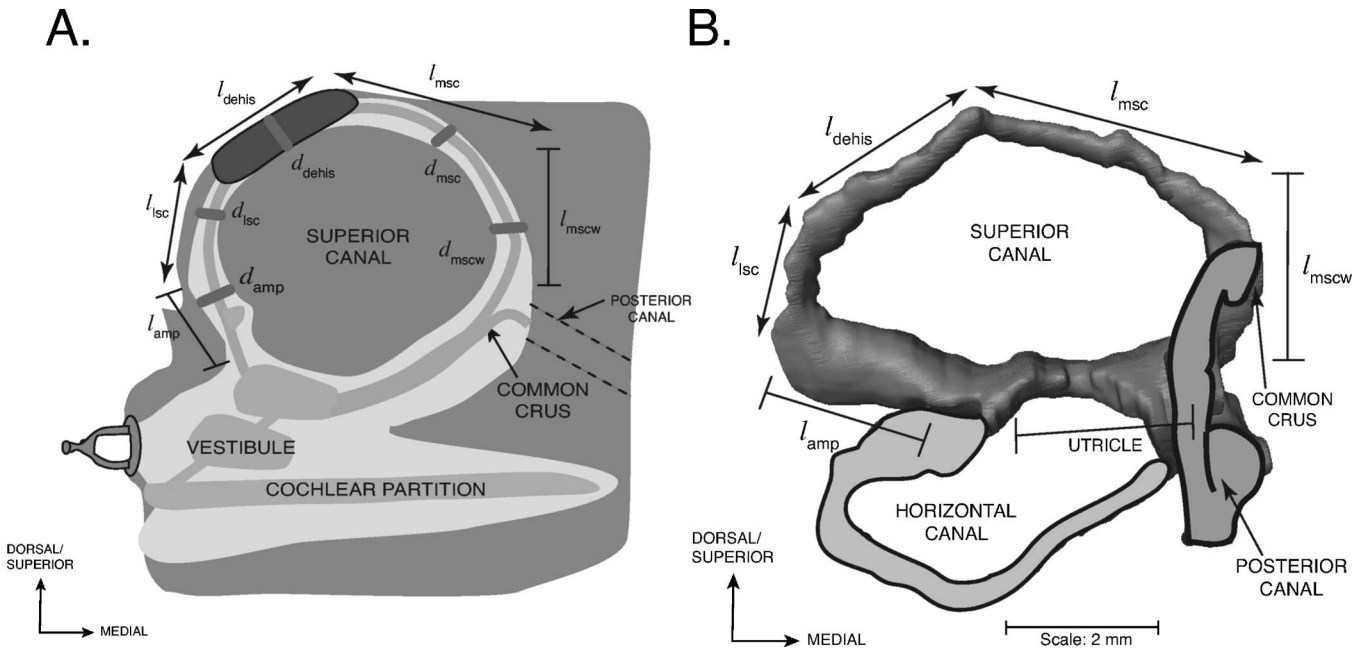


FIG. 2. A) A schematic of the reconstructed superior canal (SC) of the left ear of a chinchilla in the canal's plane. Labels with the first letter l are lengths; the ds are canal diameters. The dimensions are provided in Table I. Lengths with double-sided arrows represent variable dimensions that depend on the size and location of the dehiscence and lengths with bars represent fixed dimensions. The utricle and saccule sit within the vestibule of the inner ear. B) A reconstruction of the chinchilla bony superior semicircular canal illustrating the defined parameter lengths. The outlines of the membranous posterior and horizontal canals as well as the superior portion of the vestibule (near the utricle) are included for reference.

TABLE I. Anatomical parameters determined from the reconstruction of the SC from one ear along with the values used in the model. The lengths are listed showing a range of values where applicable. The radii are listed showing the mean and standard deviations.

Parameter	Definition	Measurements (mm)	Model (mm)
ℓ_{amp}	Length of ampulla	2.5	2.5
a_{amp}	Radius of ampulla	0.42±0.17	0.42
ℓ_{lsc}	Length of lateral SC branch	$0.1<\ell_{lsc}<5.0$	2.5
a_{lsc}	Radius of lateral SC branch	0.18±0.02	0.18
ℓ_{dehis}	Length of dehiscence	$0.1<\ell_{dehis}<2.0$	1.0
a_{dehis}	Radius of dehiscence	$0<a_{dehis}<a_{lsc}$	0.18
ℓ_{msc}	Length of medical SC branch	$7.7-\ell_{lsc}-\ell_{dehis}$	4.2
a_{msc}	Radius of medical SC branch	0.14±0.02	0.14
ℓ_{mscw}	Length of wide medial SC branch	3.0	3.0
a_{mscw}	Radius of wide medial SC branch	0.55±0.42	0.55
ℓ_{bone}	Width of the SC wall	0.18±0.02	0.18

cranial cavity. This 5.0 mm range is the region where it was surgically possible to place a dehiscence. However, our dehiscences were typically located 2.5 mm from the ampulla resulting in an ℓ_{lsc} of 2.5 mm and were typically 1.0 mm in length resulting in an ℓ_{dehis} of 1.0 mm. This resulted in an ℓ_{msc} value of 4.2 mm (7.7−2.5−1).

The diameter of each bony canal segment was the average diameter of multiple locations (approximately 100 μm intervals). Since the bony canal is not perfectly circular in cross section, the mean of the multiple diameter measurements and the measurements of the diameter in the perpendicular plane were averaged for each tube segment to define the diameter of the segment. These measurements were made in one chinchilla ear and variations between animals may exist. Figure 2 illustrates the measured diameters from which the radii listed in Table I are determined. The radius of the canal near the dehiscence was considered to be equal to a_{lsc} with a value of 0.18 mm. The radius of the dehiscence itself could range between 0 and the radius of the canal, 0.18 mm. The surgically induced dehiscences in this study typically had radii, a_{dehis}, of 0.18 mm (which was sometimes rounded to 0.2 mm) and lengths, ℓ_{dehis}, of 1.0 mm.

B. Model of the SC with dehiscence

Figure 3 shows the model structure of Z_{SCD}. The model consists of three blocks. The first block consists of two two-port tube models (Egolf, 1977) that account for the lateral branch of the SC including the ampulla of the SC (Z_{amp}) and the SC segment lateral to the dehiscence (Z_{lsc}). In parallel with these two ports is the second block, another set of two ports which model sound flow through the medial branch of the SC, first the wide segment (Z_{mscw}) and then the narrower segment (Z_{msc}). In series with the upper output nodes of the canal segments is the third block representing the impedance of the dehiscence, Z_{dehis}. Z_{dehis} has two parts: the acoustic impedance associated with sound flow through the fluid-filled hole in the bone (Z_{hole}) and the radiation of sound away from the hole (Z_{rad}) into the air-filled middle ear.

For the purpose of this model we assume that the SC extends from the boundary between the utricle and the wide segment of the medial branch of the SC to the boundary between the ampulla and the utricle. These anatomical boundaries were defined to simplify the tube calculations.

1. Defining Z_{dehis}

The dehiscence itself (Z_{dehis}) is modeled as the series combination of the impedance of the hole in the bone and the radiation of sound from the hole into air. The radiation of sound from the hole depends on the medium outside the hole and the size of the hole. In our chinchilla measurements, this medium is the air within the opened middle-ear cavity. We use a radiation impedance computed for sound in air to account for sound radiation from the hole into the open cavity. The radiation impedance is approximated using the equations for a piston (where the fluid interface inside the hole is considered to act like a piston) in an infinite baffle (Beranek, 1998, p. 121).[1] The radiation impedance represents the im-

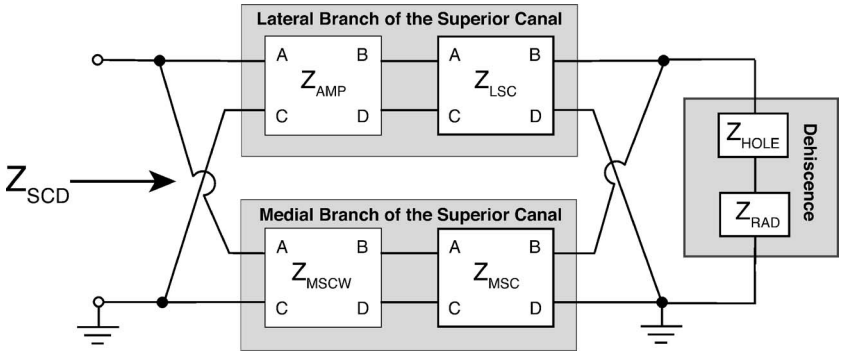


FIG. 3. Model of the SC impedance. Z_{amp}, Z_{lsc}, Z_{mscw}, and Z_{msc} are transmission matrices used to determine the impedance of both the medial and lateral branches of the SC, each of which extends to the dehiscence. Z_{dehis}, a combination of Z_{hole}, which represents the impedance of the fluid in the hole in the bony canal wall, and Z_{rad}, the radiation impedance from the hole to the air-filled middle-ear space of the chinchilla.

pedance at the boundary between the inner ear fluid within the dehiscence and the air-filled middle-ear cavities.

The hole in the bone is then modeled as a small-sized fluid-filled tube with a length equal to the thickness of the bony wall surrounding the canal (ℓ_{bone}). The radius of the tube (a_{hole}) is the effective radius of the dehiscence, usually an average of a_{dehis} (half the width of the dehiscence) and half the length of the dehiscence $\left(a_{\text{hole}}=\frac{a_{\text{dehis}}+\ell_{\text{dehis}}/2}{2}\right)$. Both the width of the dehiscence and the length of the dehiscence have been estimated for each individual ear and the length and width used in the model are typical of those seen in the individual ears. In ears in which we wanted to evaluate the effect of dehiscence size, we also looked at dehiscences where a_{dehis} was smaller than the radius of the canal, and in these cases $a_{\text{hole}}=a_{\text{dehis}}$ and $\ell_{\text{dehis}}=a_{\text{dehis}}$.

Once a_{hole} has been defined, the impedance of the hole is calculated using equations from Beranek, 1998,[2] using his small-size tube approximation (p.135) and values for the density and viscosity of water at 37°C. This specification of the hole impedance yields an intuitive result in that smaller holes correspond to larger impedances. Specifically, as the size of the dehiscence approaches zero the sound flow through the dehiscence reduces to zero.

Z_{dehis} is then evaluated as the simple series combination of the fluid-filled hole, Z_{hole}, and the radiation of sound into the air-filled middle ear, Z_{rad}.[3]

2. Branches of the SC and Z_{SCD}

Two ports are used to model the two branches of the SC resulting from dehiscence. The modeled tube segments are the lateral tube, including the ampulla and the lateral segment of the SC, and the medial tube, including both the wide and narrow portions of the medial segment of the SC (Fig. 3). The two ports are described in terms of lossy transmission matrices that account for the viscosity of the fluids.[4,5] The transmission matrix approach allows us to represent the entire length of either the medial or lateral tube as the product of transmission matrices representing smaller tube segments, assuming that neither the cross-sectional area of the tube nor its contents change significantly from one segment to the next. Since the canal diameter does not exhibit any abrupt widenings or narrowings in the regions we are looking at, the first condition is met and since all of the tubes are filled with the same fluid contents, the second condition is also held to be true. We then assume that the cross-sectional dimensions of the vestibule and common crus are large relative to the dimensions of the SC and that these large cross sections contribute little to the impedance associated with fluid motion through the dehiscent canal. Based on these assumptions we model the acoustic effects of the fluid in the vestibule and common crus as an end correction to the length of the SC (Beranek, 1998, p. 132). Specifically, an end correction is added to both ℓ_{mscw} and ℓ_{amp}.

To calculate Z_{SCD}, Z_{hole} and Z_{rad} are placed in series, and are the load on the parallel combination of impedance of the medial and lateral branches of the superior semicircular canal. We simplified this calculation by replacing the transmission matrix representations of the medial and lateral branches with lumped element impedances equivalent to the input impedance of each tube section. A two port can be replaced with an equivalent lumped element parameter if the current at the input and output terminals are equal ($U_{\text{in}}=U_{\text{out}}$).[6] Using our parameter values, U_{out} is within 1% of U_{in} in each branch tube in the 100–8000 Hz frequency range. The equivalent lumped elements of the two canal branches are well approximated by simple acoustic masses over this frequency range.

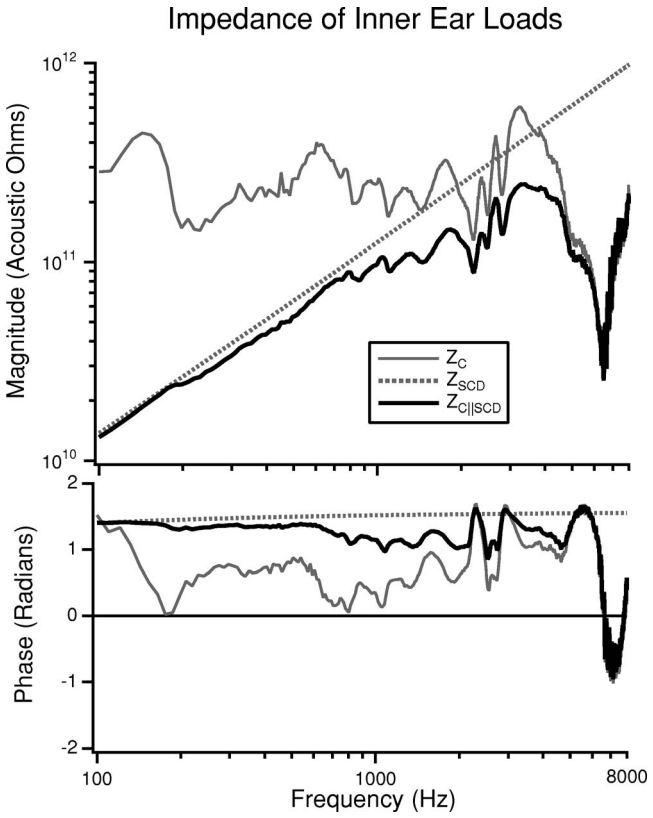


FIG. 4. The load on the middle ear due to the inner ear is illustrated both before and after the introduction of a SCD. In the intact ear, the load is Z_c (Songer and Rosowski, 2007). After the introduction of the dehiscence the load is the parallel combination of Z_c and Z_{SCD} (Fig. 3) and is illustrated as $Z_{c\|\text{scd}}$.

III. MODEL PREDICTIONS OF THE EFFECT OF SCD ON THE INNER EAR LOAD

When a dehiscence is placed in the SC, Z_{SCD}, which is represented by the model in Fig. 3, is placed in parallel with the normal cochlear input impedance (Fig. 1). This parallel combination defines the load of the dehiscent inner ear on the middle ear. Figure 4 illustrates the Z_c for the intact ear, as described by our earlier analysis (Songer and Rosowski, 2007), the Z_{SCD} associated with a 1-mm-long dehiscence located 2.5 mm medial to the ampulla (using the dehiscence parameters described in Table I) and the parallel combination of Z_c and Z_{SCD}, $Z_{c\|\text{scd}}$. In the case of a 1 mm dehiscence at this location, Z_{SCD} is dominated by the acoustic mass of the fluid within the medial and lateral tube segments; there is little contribution from the hole and the radiation impedance (Z_{dehis} is at least 30 dB less than Z_{SCD} for frequencies between 100 and 8000 Hz). The mass-like Z_{SCD} controls the

impedance of the dehiscent inner ear at frequencies below 800 Hz. Above 800 Hz, the impedance of the parallel combination of $Z_{c\|\mathrm{scd}}$ depends on both the mass-dominated Z_{SCD} and the more resistive Z_c. The dehiscence has nearly no effect at frequencies above 4 kHz, where the lower impedance of the cochlea dominates the load of the dehiscent inner ear on the middle ear.

IV. MODEL PREDICTIONS OF THE EFFECT OF DEHISCENCE ON MIDDLE-EAR SOUND TRANSMISSION

Our model of the effect of SCD on auditory sensitivity and mechanics uses the lumped-element representation of Z_{SCD} described in the previous sections as well as an experimentally derived representation of the middle ear and Z_c (Songer and Rosowski, 2007). In order to establish the validity of the model, predictions are made of the effects of SCD on three model variables and compared to previous measurements of the same: middle-ear input admittance, Y_{ME} (Songer and Rosowski, 2005), middle-ear transfer function, H_p (Songer and Rosowski, 2006), and cochlear potential normalized by P_{TM}, G (Songer and Rosowski, 2005).[7] The model without the dehiscence has previously been demonstrated to fit normal measurements of admittance and middle-ear transfer function (Songer and Rosowski, 2007). In our comparisons here, we concentrate on dehiscence induced changes.

We evaluate the SCD-induced change in Y_{ME} as the dB difference between Y_{ME} after the introduction of SCD and Y_{ME} in the intact ear and refer to this value as ΔY_{ME} where $|\Delta Y_{\mathrm{ME}}| = 20 \log_{10} \frac{|Y_{\mathrm{ME}}|_{\mathrm{SCD}}|}{|Y_{\mathrm{ME}}|}$ and $\angle \Delta Y_{\mathrm{ME}} = \angle Y_{\mathrm{ME}}|_{\mathrm{SCD}} - \angle Y_{\mathrm{ME}}$. The model predictions of ΔY_{ME} and the ΔY_{ME} measured in eight chinchilla ears (Songer and Rosowski, 2006) are illustrated in Fig. 5. The model prediction of ΔY_{ME} fit within the 95% confidence intervals for the data over almost the entire frequency range. This suggests that the model accurately predicts dehiscence related changes in middle-ear input admittance.

The model prediction of SCD-induced change in H_p is represented as ΔH_p (the difference in H_p before and after the introduction of an SCD) and is compared to the 95% confidence interval of measurements of ΔH_p ($n=6$) from a previous study (Songer and Rosowski, 2006) (Fig. 6). The model prediction of ΔH_P is within the 95% confidence intervals of the measurements over almost the entire frequency range.

The close fit between the model predictions of both ΔY_{ME} and ΔH_p indicate that our model (Fig. 1) effectively predicts changes in auditory mechanics that result from the introduction of a dehiscence.

In addition to evaluating changes in auditory mechanics, we also evaluate the change in hearing sensitivity predicted by the model and compare it to the measured change in cochlear potential (CP) published previously (Songer and Rosowski, 2006). The effect of SCD on hearing is modeled as the change in volume velocity through Z_c normalized by sound pressure at the tympanic membrane $\left(\frac{U_c}{P_{\mathrm{TM}}}\right)$. We assume that the changes in volume velocity through Z_c reflect changes in the effective stimulus to the cochlea and are proportional to changes in cochlear potential normalized by sound pressure (i.e., $G = \frac{\mathrm{CP}}{P_{\mathrm{TM}}} \propto \frac{U_c}{P_{\mathrm{TM}}}$). The model predicted SCD-induced change in G is presented as a dB difference, ΔG, and compared to measurements of ΔG from a previous study ($n=6$) (Songer and Rosowski, 2005) (Fig. 7).

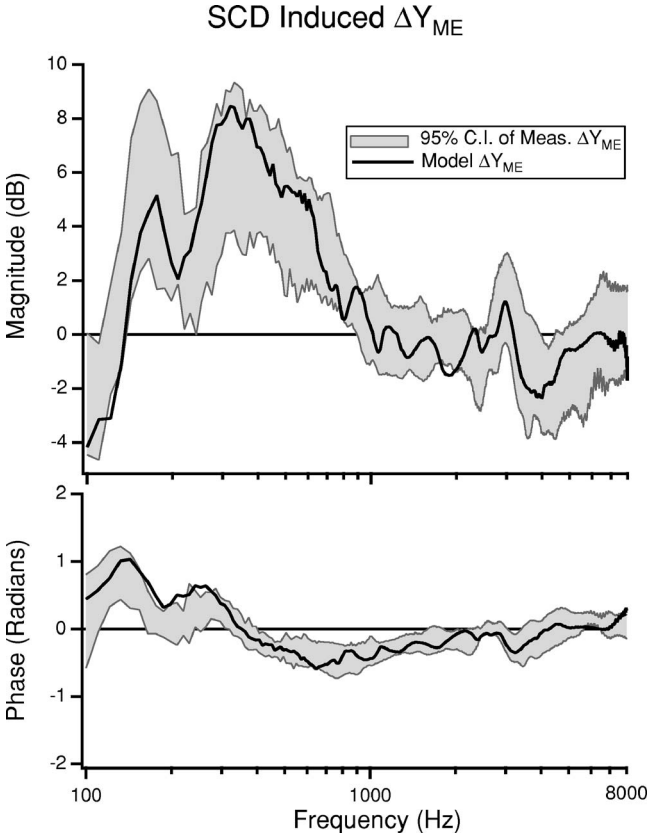


FIG. 5. Model prediction of ΔY_{ME} resulting from dehiscence presented in dB. The model prediction is plotted with the 95% confidence intervals of the ΔY_{ME} measured in eight chinchilla ears (Songer and Rosowski, 2006) for comparison.

Below 200 Hz the measurements of G are close to the noise floor (within 10 dB) and are not plotted. For frequencies above 300 Hz the model prediction fits within the 95% confidence intervals of the data. Below 300 Hz our prediction overestimates the decrease in ΔG resulting from dehiscence. One reason why the model fit might be off for low frequencies is the difficulty in acquiring low-frequency CP data (Songer and Rosowski, 2005) that are not contaminated by noise. Another possible source of discrepancy may arise from inner ear nonlinearities that have been observed between 80 and 300 Hz in chinchilla (Dallos, 1970; Songer and Rosowski, 2006; Rosowski, Ravicz, and Songer, 2006) that may affect the experimental results, but may not be effectively captured by the model. Overall, however, the model succeeds in predicting the observed decrease in auditory sensitivity to air-conducted sound.

V. DISCUSSION

We report a model of the auditory effects of dehiscence in chinchilla. The model predicts a decrease in hearing sensitivity as well as an increase in both middle-ear input admittance and the middle-ear transfer function magnitude as-

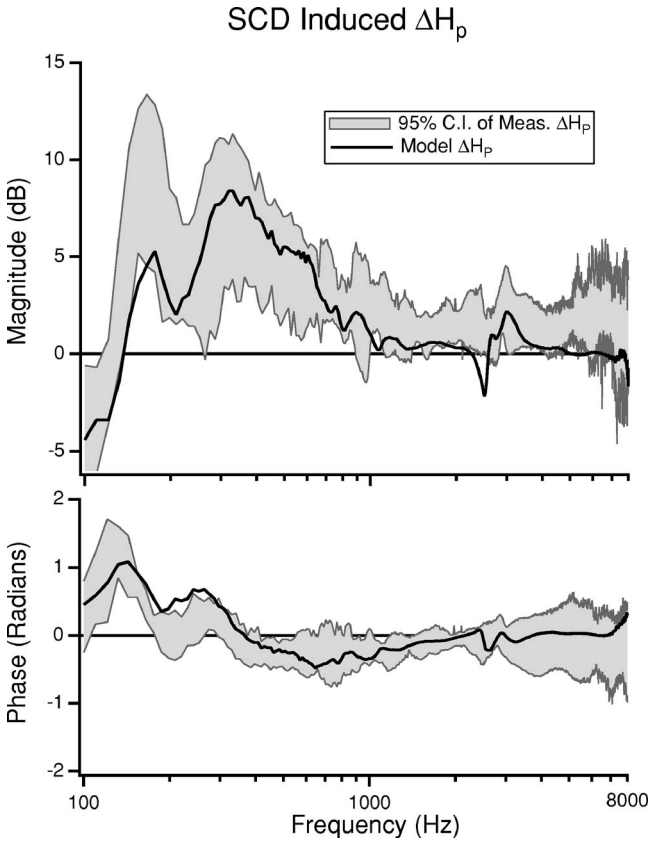


FIG. 6. Model prediction of ΔH_p resulting from dehiscence. The 95% confidence intervals from previous measurements of ΔH_p in six chinchilla ears (Songer and Rosowski, 2006) are presented for comparison.

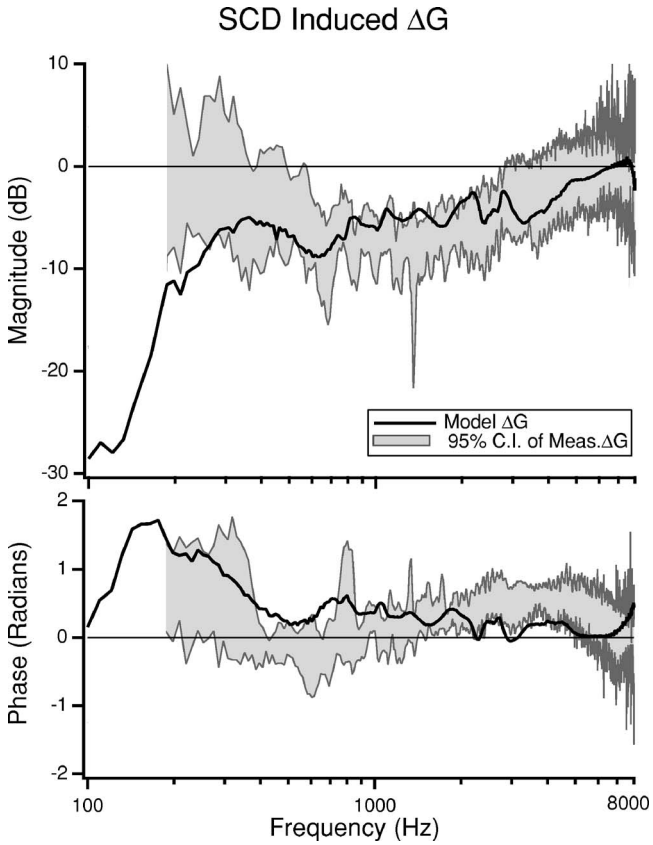


FIG. 7. Model prediction of the SCD induced change in auditory sensitivity (ΔG) presented in dB. The 95% confidence intervals of measurements of ΔG from six chinchilla ears made in a previous study are also illustrated (Songer and Rosowski, 2005). Measurement data for frequencies below 200 Hz are near the noise floor and are not illustrated.

sociated with SCD. These predictions are a good match to the experimentally observed data (Songer and Rosowski, 2005, 2006) as presented above.

A. Model limitations and adaptations

The model described in this paper consists of three major elements: the middle ear represented as a two port, the cochlea represented by Z_c, and the superior semicircular canal, including the dehiscence, represented as Z_{SCD}. Each segment of the model is subject to a number of assumptions that may affect the reliability of the model predictions. One limitation of the model is the frequency range of validity for the transmission matrix representation of the middle ear as well as the calculated value for Z_c. As described in detail in Songer and Rosowski, 2007, the physiologic measurements from which they were determined were most reliable for frequencies between 100 and 1500 Hz; it is therefore possible that errors in these model elements could be introduced into our predictions at frequencies above 1500 Hz. Due to the high degree of similarity between the measured and predicted responses for Y_{ME}, H_p, and auditory sensitivity across the entire frequency range, we do not believe that this potential restriction is negatively influencing our results.

The model structure and layout presented here for chinchilla may be adapted to predict the effect of SCD on auditory parameters in humans. In order to implement a human model of SCD based on this work, an existing middle-ear model could be used in conjunction with the lumped-element model of the SCD presented here, but adapted to reflect human anatomical and physiological parameters. Important anatomical and physiological differences that need to be addressed include: differences in sound transmission through the human middle ear, the addition of closed middle-ear air spaces in human patients, the state of the middle-ear muscles (they have been inactivated in our animal measurements), the dimensions of the semicircular canal, differences in the size and location of the dehiscence, and the fact that the dehiscence in humans opens into the cranial cavity and is bounded by the brain and its surrounding membranes and fluid spaces. Implementing this model structure for humans and accounting for the differences described above will allow us to test the hypothesized mechanism of SCD-induced changes in auditory sensitivity and may provide insight into the variability of patient presentation.

B. Clinical relevance

Previous work has compared the effect of SCD in chinchilla (Songer and Rosowski, 2005, 2006) with clinical data obtained from human patients (Minor *et al.*, 2003; Mikulec *et al.*, 2004) and from temporal bones (Chien *et al.*, 2007). These comparisons demonstrate decreases in our measures of auditory sensitivity in response to SCD that are qualitatively similar: a decrease in auditory sensitivity measured as a decrease in cochlear potential in chinchilla (Songer and

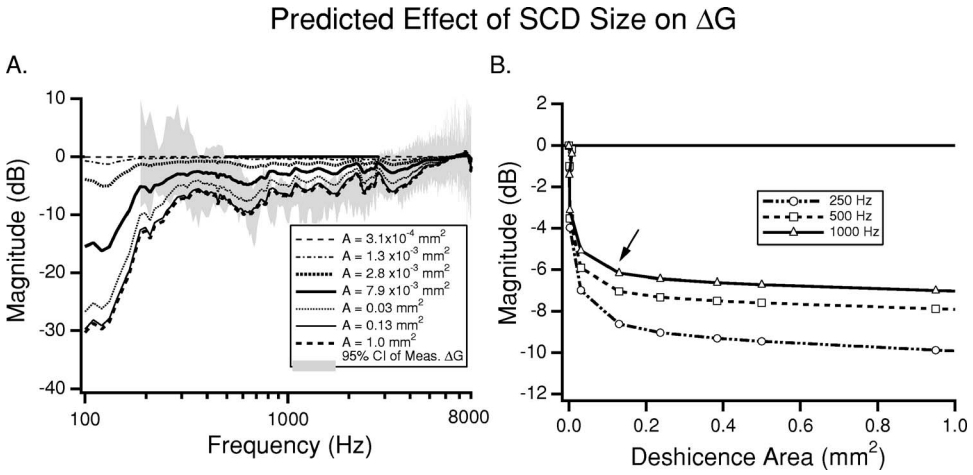


FIG. 8. The predicted effect of SCD size on ΔG. For reference, the estimated cross sectional area of the SC is 0.13 mm^2. A) The model predictions as a function of frequency along with the 95% confidence interval (CI) from measurements (areas between 0.24 and 1.0 mm^2) (Songer and Rosowski, 2005) are illustrated for comparison. Note that the curves for areas of 0.13 and 1.0 mm^2 are difficult to distinguish because there is little difference between them. B) The model predictions at select frequencies (250, 500, and 1000 Hz) plotted as a function of SCD area. The arrow indicates the cross-sectional area of the canal.

Rosowski, 2005) and as an air-bone gap in human patients (Minor *et al.*, 2003; Mikulec *et al.*, 2004). Qualitative comparisons of the effect of SCD on auditory mechanics show increases in both H_p and Y_{ME} in chinchilla (Songer and Rosowski, 2006), increases in umbo velocity in humans (Rosowski *et al.*, 2004), and increases in stapes velocity in human temporal bones (Chien *et al.*, 2007).

Despite the qualitative similarities between the chinchilla and human physiological data, it is difficult to make quantitative comparisons. Part of this difficulty is due to the wide variation in patient symptoms. Some patients with SCD syndrome have large air-bone gaps and some patients have no auditory symptoms (no air-bone gap). We hypothesize that these differences may be due to differences in the structure and state of the middle ear, the size of the dehiscence, the location of the dehiscence and other anatomical parameters. In the next section we look at how some of these differences (dehiscence size and location) affect auditory sensitivity in chinchilla.

C. Predicted effect of variations in dehiscence size and location

A question relevant to the diagnosis of SCD is "How do patient symptoms vary with dehiscence size and location?" Using our model we can predict the effect of variations in dehiscence size on auditory sensitivity in chinchilla. These effects are modeled by varying the radius and length of the dehiscence, resulting in changes in SCD area and the lengths of the lateral and medial limbs of the canal remnants. In Fig. 8(A) the model predictions of the effect of dehiscence size on the frequency dependence of the hearing loss are plotted along with the 95% confidence intervals from measurements with areas between 0.24 and 1.0 mm^2 Songer and Rosowski, 2005. Figure 8(A) demonstrates that small dehiscences have little effect on ΔG, and large dehiscences result in predicted changes similar to those observed experimentally Songer and Rosowski, 2005. Figure 8(A) also suggests that the SCD induced hearing loss is largest at low frequencies. Figure 8(B) plots the predicted change in auditory sensitivity at three frequencies as a function of dehiscence area. Figure 8(B) demonstrates that very small dehiscences have little effect on auditory sensitivity, however, as the dehiscence size approaches the cross-sectional area of the canal auditory sensitivity decreases. Once the dehiscence size exceeds the cross-sectional area of the canal ($a_{dehis} > 0.2$ mm, *area* $= 0.13$ mm^2) there is little additional change in ΔG.

The effect of dehiscence location can also be predicted using the model. Figure 9 shows the predicted effect of dehiscence location on ΔG where the size of the SCD is fixed at $a_{hole} = 0.2$ mm, $\ell_{dehis} = 1$ mm. It predicts that as the SCD gets closer to the ampulla, the decrease in ΔG resulting from the dehiscence gets slightly larger. These changes are small and most of the curves fall within the 95% confidence interval of the measured data indicating that they fall within predicted inter-animal differences.

All of the dehiscences reported in human patients with SCD syndrome are large, typically with short dimensions equal to the radius of the canal and lengths on the order of millimeters. Our model predictions suggest that slight variations in these large dehiscence sizes are not expected to account for the wide variations in hearing sensitivity observed in human patients with SCD syndrome. The model does suggest small changes in auditory sensitivity resulting from

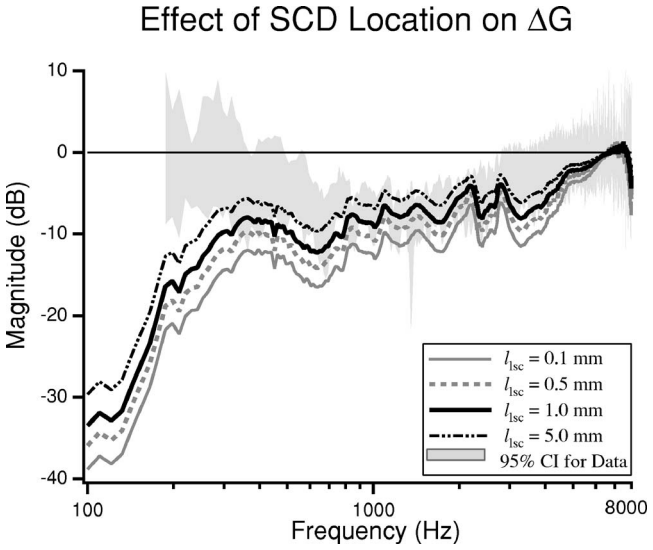


FIG. 9. Predicted effect of dehiscence location on ΔG. All the predictions are for dehiscences that are 0.4 mm in diameter and 1.0 mm in length; ℓ_{lsc} refers to the distance between the ampulla and the dehiscence and is varied as illustrated in the figure. The 95% CI from the previously reported measured ΔG (Songer and Rosowski, 2005) is illustrated for comparison.

variations in dehiscence location, but such small changes cannot account for large variations observed in the presentation of auditory symptoms in patients (Minor *et al.*, 2003; Mikulec *et al.*, 2004). One caveat is that we only tested the effect of dehiscence location within the range where we had surgical access to the SC through the chinchilla middle-ear air space. These results suggest that developing a model of SCD in humans where dehiscence location can be manipulated may lead to a better understanding of the potential effect of dehiscence location on auditory sensitivity. The model does not predict whether changes in size and location of dehiscence would affect vestibular symptoms.

VI. CONCLUSIONS

We created a histological reconstruction of the chinchilla inner ear and used it to define anatomical parameters for a lumped-element model of the chinchilla superior semicircular canal, including dehiscence. The model was then used to predict the effects of a SC hole or dehiscence on both auditory sensitivity and mechanics. These predictions are consistent with previous physiological data showing decreases in the magnitude of cochlear potential in response to dehiscence as well as increases in both middle-ear input admittance and middle-ear transfer function magnitude in response to dehiscence. Additionally, the model predicts the effect of dehiscence location and size on auditory sensitivity and mechanics and provides a framework for generating a model of the effect of SCD on human patients.

ACKNOWLEDGMENTS

This work has been supported by a NSF graduate student fellowship, NIH training grant, and additional NIH grants. S. N. Merchant, C. A. Shera, W. T. Peake, M. E. Ravicz, and M. L. Wood provided insights and suggestions.

[1]The radiation impedance we use is described by Beranek, 1998, p. 121, and is defined below:

$$R_{a1} = \frac{0.1404\rho_{\text{air}}c_{\text{air}}}{\pi a_{\text{hole}}^2},$$

$$R_{a2} = \frac{\rho_{\text{air}}c_{\text{air}}}{\pi a_{\text{hole}}^2},$$

$$C_{a1} = \frac{5.94a_{\text{hole}}^3}{\rho_{\text{air}}c_{\text{air}}^2},$$

$$M_{a1} = \frac{0.27\rho_{\text{air}}}{a_{\text{hole}}},$$

$$Zc_{a1} = \frac{1}{j\omega C_{a1}},$$

$$Zm_{a1} = j\omega M_{a1},$$

$$Z_{\text{branch 1}} = \frac{R_{a1}Zc_{a1}}{R_{a1} + Zc_{a1}} + R_{a2},$$

$$Z_{\text{rad}} = \frac{Z_{\text{branch1}}Zm_{a1}}{Z_{\text{branch1}} + Zm_{a1}}.$$

[2]The impedance of the hole as defined by Beranek, 1998, p. 135, is calculated as follow:

$$R_{\text{hole}} = \frac{8\mu_{H_2O}\ell_{\text{bone}}}{\pi a_{\text{hole}}^4},$$

$$M_{\text{hole}} = \frac{\rho_{H_2O} * \ell_{\text{bone}}}{\pi a_{\text{hole}}^2},$$

$$Z_{\text{hole}} = R_{\text{hole}} + j\omega\frac{4}{3}M_{\text{hole}}.$$

Note that as dehiscence size approaches zero, Z_{dehis} approaches infinity, effectively closing the third window. The appropriateness of the small-size tube approximation was evaluated by comparing it to the exact solution. Differences between the exact solution and the above approximation were less than 2.5 dB in magnitude.

[3]In these ears the membranous labyrinth and endosteum are typically still intact, which could lead to a membranous covering over the hole. This is not explicitly accounted for in the model. The decision not to include these membranes was based on observations that presence or absence of a membrane at the interface between the fluid and air did not impact our measured results and we wanted to keep the model as simple as possible.

[4]The lossy transmission-line models of tube segments that we use were first described by Egolf, 1977, and adapted for fluid filled tubes by Dickens, 1986, as follows:

$$Y = j\omega\frac{\pi a^2}{\rho c^2}\left[1 + \frac{2(\gamma-1)}{k_p a}\frac{J_1(k_p a)}{J_0(k_p a)}\right],$$

$$Z = j\omega\frac{\rho}{\pi a^2}\left[1 - \frac{2}{k_s a}\frac{J_1(k_s a)}{J_0(K_s a)}\right]^{-1},$$

$$G = \ell\sqrt{Y \times Z},$$

$$Z0 = \sqrt{\frac{Z}{Y}},$$

$$A = \cosh(G),$$

$$D = A,$$

$$B = Z0 \times \sinh(G),$$

$$C = \frac{\sinh(G)}{Z0},$$

$$z_{in} = \frac{Az_o + B}{Cz_o + D}.$$

$J_1()$ and $J_0()$ are Bessel functions of order 1 and 0 and ω refers to 2π times the frequency. The arguments to the Bessel functions are dimensionless: $k_p = \sqrt{-\frac{j\omega}{h^2}}$ and $k_s = \sqrt{\frac{-j\omega\rho}{\mu}}$. $h^2 = \frac{\kappa}{\rho c_p}$ which is the thermal diffusivity of the medium; z_o is the terminating impedance of the tube.

[5]The medium of interest for this work is inner ear lymph at body temperature. We assume that lymph is approximated by water and has the following physical properties:

$$\gamma = 1,$$

$$c = 1560 \text{ m/s},$$

$$\rho = 993 \text{ kg/m}^3,$$

$$\kappa = 0.6340 \text{ W/}(m - K),$$

$$\mu = 2 \times 10^{-3} \text{ kg/}(\text{m} - \text{s}),$$

$c_p = 4.16 \times 10^{-3}$ J/(kg – K).

γ is ratio of specific heats of the fluid medium, c is the speed of sound in water, ρ is the density of water, κ is the thermal conductivity of water, μ is the absolute viscosity and c_p is the specific heat.

[6]The equality of U_{in} and U_{out} depends on the transmission matrix parameters $A(\omega)$ and $D(\omega)$ of the two port equaling 1 throughout the frequency range of interest.

[7]All of the previously reported data referred to here had dehiscence sizes that were approximately 0.4 mm, in diameter and 1.0 mm in length.

Beranek, L. L. (**1998**). *Acoustics* (American Institute of Physics, New York).

Brantberg, K., Bagger-Sjoback, D., Mathiesen, T., Witt, H., and Pansell, T. (**2006**). "Posterior canal dehiscence syndrome caused by an apex cholesteatoma," Otol. Neurotol. **27**(4), 531–534.

Chien, W., Ravicz, M. E., Rosowski, J. J., and Merchant, S. N. (**2007**). "Measurements of human middle- and inner-ear mechanics with dehiscence of the superior semicircular canal," Otol. Neurotol. **28**, 250–257.

Dallos, P. (**1970**). "Low-frequency auditory characteristics: Species dependence," J. Acoust. Soc. Am. **48**, 489–499.

Egolf, D. (**1977**). "Mathematical modeling of a probe-tube microphone," J. Acoust. Soc. Am. **61**, 200–205.

Krombach, G. A., DiMartino, E., Schmitz-Rode, T., Prescher, A., Haage, P., and Kinzel, S. *et al.* (**2003**). "Posterior semicircular canal dehiscence: A morphological cause of vertigo similar to superior canal dehiscence," Eur. Radiol. **13**, 1444–1450.

Mikulec, A., McKenna, M., Ramsey, M., Rosowski, J. J., Herrmann, B., and Rauch, S. *et al.* (**2004**). "Superior semicircular canal dehiscence presenting as conductive hearing loss without vertigo," Otol. Neurotol. **25**, 121–129.

Mimura, T., Sato, E., Sugiura, M., Yoshino, T., Naganawa, S., and Nakashima, N. (**2005**). "Hearing loss in patients with enlarged vestibular aqueduct: Air-bone gap and audiological bing test," Int. J. Audiol. **44**(8), 466–469.

Minor, L., Carey, J., Cremer, P., Lustig, L., and Streubel, S. (**2003**). "Dehiscence of bone overlying the superior canal as a cause of apparent conductive hearing loss," Otol. Neurotol. **24**(2), 270–278.

Minor, L., Solomon, D., Zinreich, J., and Zee, D. (**1998**). "Sound- and/or pressure-induced vertigo due to bone dehiscence of the superior semicircular canal," Arch. Otolaryngol. Head Neck Surg. **124**, 249–258.

Rosowski, J. J., Ravicz, M. E., and Songer, J. E. (**2006**). "Structures that contribute to middle-ear admittance in chinchilla," J. Comp. Physiol. **192**(12), 1287–1311.

Rosowski, J. J., Songer, J. E., Nakajima, H. H., Brinsko, K. M., and Merchant, S. N. (**2004**). "Investigations of the effect of superior semicircular canal dehiscence on hearing mechanisms," Otol. Neurotol. **25**, 323–332.

Songer, J. E., and Rosowski, J. J. (**2005**). "The effect of superior canal dehiscence on cochlear potential in response to air-conducted stimuli in chinchilla," Hear. Res. **210**, 53–62.

Songer, J. E., and Rosowski, J. J. (**2006**). "The effect of superior-canal opening on middle-ear input admittance and air-conducted stapes velocity in chinchilla," J. Acoust. Soc. Am. **120**(1), 258–269.

Songer, J. E., and Rosowski, J. J. (**2007**). "Transmission matrix analysis of the chinchilla middle ear." J. Acoust. Soc. Am. **122**(2).

Intracochlear pressure and derived quantities from a three-dimensional model

Yong-Jin Yoon,[a] Sunil Puria, and Charles R. Steele
Department of Mechanical Engineering and Department of Otolaryngology-Head and Neck Surgery, Stanford University, Stanford, California 94305-4035

(Received 10 October 2006; revised 9 May 2007; accepted 10 May 2007)

Intracochlear pressure is calculated from a physiologically based, three-dimensional gerbil cochlea model. Olson [J. Acoust. Soc. Am. **103**, 3445–3463 (1998); **110**, 349–367 (2001)] measured gerbil intracochlear pressure and provided approximations for the following *derived quantities*: (1) basilar membrane velocity, (2) pressure across the organ of Corti, and (3) partition impedance. The objective of this work is to compare the calculations and measurements for the pressure at points and the derived quantities. The model includes the three-dimensional viscous fluid and the pectinate zone of the elastic orthotropic basilar membrane with dimensional and material property variation along its length. The arrangement of outer hair cell forces within the organ of Corti cytoarchitecture is incorporated by adding the feed-forward approximation to the passive model as done previously. The intracochlear pressure consists of both the compressive fast wave and the slow traveling wave. A Wentzel–Kramers–Brillowin asymptotic and numerical method combined with Fourier series expansions is used to provide an efficient procedure that requires about 1 s to compute the response for a given frequency. Results show reasonably good agreement for the direct pressure and the derived quantities. This confirms the importance of the three-dimensional motion of the fluid for an accurate cochlear model. [DOI: 10.1121/1.2747162]

PACS number(s): 43.64.Kc, 43.64.Bt [BLM] Pages: 952–966

I. INTRODUCTION

The cochlea is a snail-shaped, fluid-filled duct which is divided along its longitudinal direction by the compliant basilar membrane (BM), on which is located the organ of Corti (OC). The fluid and compliant structures within the cochlea are set in motion in response to sound input at the stapes, and the detection of this motion by inner hair cells initiates hearing through afferent auditory nerve firing transmitted to the auditory brainstem. The pressure difference across the OC is one of the driving forces of the motion of OC, and this motion has been the subject of intracochlear experimentation and cochlear models for analysis.

This study is motivated by the measurements of intracochlear pressure (Olson, 1998, 2001). Pressure near the stapes at scala vestibule (SV), which is the "input" pressure to the inner ear, and pressure at the scala tympani (ST) through the round window (RW) opening was measured from the gerbil cochlea *in vivo*. Intracochlear pressure at a number of positions spaced by tens of micrometers was measured to obtain the localized pressure and the pressure gradients which indicate fluid motion in the base of the gerbil cochlea.

In this study, the mechanical behavior of the cochlea is simulated with a physiologically-based, three-dimensional (3D) cochlear model. Results are compared with the experimental data for the best frequency [best frequency (BF)]-to-place map, BM velocity, intracochlear pressure, and quantities derived from the pressure, using the formulas of: (1) BM velocity, (2) pressure difference across OC, and (3) OC impedance.

Numerous cochlear models have been used to explain the biomechanical behavior of the cochlea. Models extend the passive cochlear model with the inclusion of the motion of the OC, particularly the active behavior of the outer hair cells (OHCs). The simplified one-dimensional model with negative damping by de Boer (1983) was extended to include nonlinearity in the activity using a quasilinear method (Kanis and de Boer, 1996, 1997). Higher dimensional active models have also been developed. Two-dimensional finite difference models were constructed by using a feedback law (Neely, 1985, 1993). Numerically intense 3D finite-element models had been developed with the inclusion of varying details and complexities of the OC (Kolston and Ashmore, 1996; Böhnke and Arnold, 1998). However, the fluid was modeled as inviscid, which does not require as fine a mesh. Finally, models including the activity in the OC as a feed-forward mechanism which took into account the longitudinal tilt of the OHCs had been developed. Two-dimensional (Geisler and Sang, 1995) and 3D models with the active feed-forward mechanism has been developed (Steele *et al.*, 1993; Steele and Lim, 1999; Lim and Steele, 2002).

The present study uses the physiologically -based, linear 3D feed-forward model. The model uses a combination of the asymptotic phase integral method that is commonly known as the Wentzel–Kramers–Brillouin (WKB) method and the fourth-order Runge-Kutta (RK4) numerical forward integration. This hybrid approach provides significantly

[a]Electronic mail: yongjiny@stanford.edu

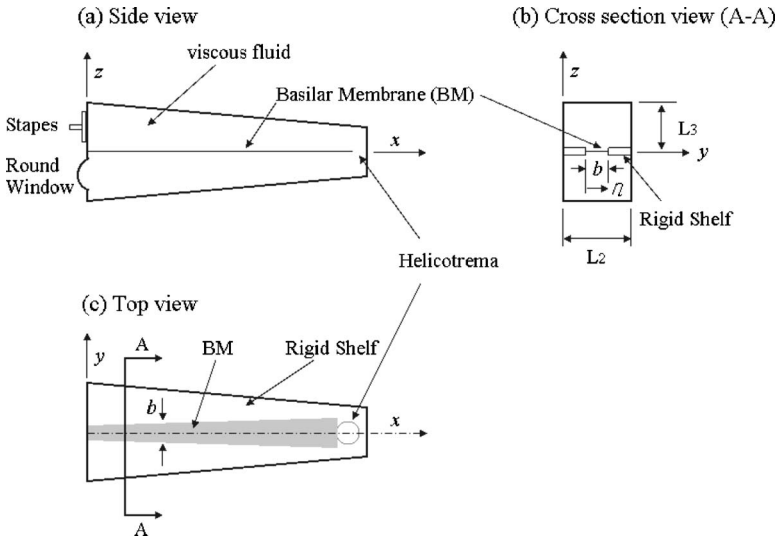


FIG. 1. Schematic drawing of the passive cochlear model geometric layout. Distances are parametrized by the Cartesian coordinates $\{x,y,z\}$, which represent the distance from the stapes, the distance across the scala width, and the height above the partition, respectively. (a) Side, (b) cross section (*A-A*), and (c) top views of the model.

faster computations than the finite difference or finite element methods and more accuracy than the WKB alone (Lim and Steele, 2002).

The present model is as simple as possible to capture the essential features in the cochlea. Included in the model are the variation of the dimensions and material properties along the cochlear duct, and 3D viscous fluid effects. On the organ of Corti, only one degree of freedom, the flexing of the pectinate zone of the orthotropic BM, is considered. The spiral coiling of the cochlea is also neglected, as it has been shown, in general, to have no significant effects on the model response (Loh, 1983; Steele and Zais, 1985). The results from this active model successfully demonstrate various aspects of a live cochlea, as observed by *in vivo* measurements.

II. MATHEMATICAL METHODS

The passive and active cochlear models are presented. First, the passive model with macroscopic features of the cochlea without OHCs motility is described. Next, the feedforward active mechanism of the OHCs from the motion of OC is formulated. Quantities of interest, including the BF-to-place map, BM velocity, and intracochlear pressure, were computed with the hybrid method which combines the WKB and RK4 methods. Intracochlear pressure combines the pressure from the compressive fast wave and the slow traveling wave. Finally, the derived quantities from the intracochlear pressure, (1) *BM velocity*, (2) *pressure difference across OC*, and (3) *OC impedance*, are compared with Olson's measurements.

A. Passive model

The physical cochlea consists of a rigid bony housing containing two coiled, fluid-filled ducts, separated by a partition that is composed of rigid and compliant regions. The geometric properties of the ducts and the mechanical properties of the partition vary along the length of the cochlea. The entire system is stimulated when the stapes displaces the cochlear fluid adjacent to the oval window (OW), which lies at the base of the top fluid duct called scala vestibuli (SV). The model is based on these physiological features of the cochlea. A schematic drawing with the side, cross section, and top view is shown in Fig. 1. For simplicity, the duct has been uncoiled and all boundaries made vertical or horizontal.

This model consists of a tapered chamber with rigid walls filled with viscous fluid. The chamber is divided by a cochlear partition into two equal ducts representing scala vestibule and scala tympani (ST). The two fluid ducts are joined at the apical end of the fluid chamber via a hole representing the helicotrema. The cochlear partition represents a collapsed scala media (SM) with its structural properties dominated by the pectinate zone of the BM. The pectinate zone of the BM is considered to be an orthotropic plate, in which the Young's modulus in the transverse direction is much greater than that in the longitudinal direction. By including the variations of BM width, thickness, and fiber density, the stiffness of the partition varies in the longitudinal direction.

Two types of waves are set up in such a model, the symmetric and the anti-symmetric pressure waves (Peterson and Bogert, 1950). The symmetric pressure wave is a fast compression wave with equal pressure on both sides of the partition. The anti-symmetric pressure wave is a slow traveling wave that has pressure of opposite sign acting on the top and bottom of the partition. Consequently, the antisymmetric pressure wave causes a significant displacement of the partition while the symmetric pressure wave does not result in motion of the partition. In the present model, the antisymmetric pressure slow wave is taken into account for the BM motion and the symmetric pressure fast wave is added to the slow traveling wave pressure for the intracochlear pressure calculation.

Due to symmetry present in the model, only one fluid duct needs to be considered in the simulation. Also, taking advantage of its slender nature, the cochlear duct is divided along its length into discrete cross-sectional slices. For each cross section the 3D fluid displacement and pressure fields are computed using a Fourier series expansion. For each cross section, the explicit expressions for the fluid displacements and fluid pressure at the partition are obtained, and these are matched with the plate's displacement and pressure to give an eikonal equation. Solving the eikonal equation

yields the complex wave numbers for each cross section in the cochlear duct. Using the continuity condition for the fluid across the cross-sectional slices, a transport equation is obtained from which the amplitudes of the waves are obtained. The detailed derivations are given in the following.

The fluid displacement vector field $\mathbf{u}$ in the ducts is decomposed into divergence of a scalar field ϕ (irrotational component) and the curl of a vector field $\boldsymbol{\psi}$ (rotational component)

$$\mathbf{u} = \nabla\phi + \nabla \times \boldsymbol{\psi}. \tag{1}$$

The displacement field from ϕ satisfies the rigid wall boundary conditions at $y=0$, $y=L_2$, and $z=L_3$ where the normal fluid displacements are zero. A functional form of ϕ that satisfies the above-presented boundary conditions is

$$\phi(x,y,z,t) = e^{-i\omega t}\sum_j \Phi(x)T_j(x)\cos\left(\frac{j\pi y}{L_2}\right) \times\cosh(\tau_j(x)(L_3 - z)), \tag{2}$$

where ω is the frequency, with a Fourier cosine series expansion used in the y direction. $\Phi(x)$ is the amplitude function along the x direction, while the $T_j(x)$ coefficients allow the fluid to match the arbitrary displacement on the BM. The coefficients $\tau_j(x)$, related to the wave-number n by the continuity equation for the incompressible fluid ($\Delta\phi=0$), reduces to

$$\tau_j(x) = \sqrt{n^2 + \left(\frac{j\pi}{L_2}\right)^2}, \tag{3}$$

where the wave-number n is defined by

$$n^2 = -\frac{\Phi_{xx}}{\Phi}. \tag{4}$$

Due to the low viscosity of the fluid, the boundary layers are localized such that the boundary layers at the rigid walls have no significant effects on the partition motion. Hence, only the boundary layer at the partition is considered. This is described by the vector field $\boldsymbol{\psi}$ (with x, y, and z components, ψ^1, ψ^2, and ψ^3) that assumes the form

$$\boldsymbol{\psi}(x,y,z,t) = \begin{pmatrix}\psi^1\\ \psi^2\\ \psi^3\end{pmatrix} = e^{-i\omega t}\sum_j e^{\kappa_j(x)z}\begin{pmatrix}\psi_j^1 \sin\left(\frac{j\pi y}{L_2}\right)\\ \psi_j^2 \cos\left(\frac{j\pi y}{L_2}\right)\\ 0\end{pmatrix}. \tag{5}$$

Note that vector field $\boldsymbol{\psi}$ describes the rotational component of the fluid displacement field due to viscosity. Here, a harmonic excitation with frequency ω is applied at the stapes. The coefficients ψ_j^1 and ψ_j^2 are related to the amplitude $\Phi(x)$ and $T_j(x)$ through no-slip boundary conditions on the cochlear partition where the tangent displacements are also zero.

The Navier-Stokes equation for no body force, incompressibility, and small displacement is

$$\rho_f(\nabla\phi'' + \nabla \times \boldsymbol{\psi}'') = -\nabla p + \mu\nabla \times (\Delta\boldsymbol{\psi}'). \tag{6}$$

Matching the terms on each side of the equation gives the following:

$$\rho_f\phi'' = -p, \tag{7a}$$

$$\rho_f\boldsymbol{\psi}' = \mu\Delta\boldsymbol{\psi}''. \tag{7b}$$

Equation (7a) relates the pressure acting on the fluid to the scalar potential ϕ and the vorticity Eq. (7b) gives the condition on the vector field, $\boldsymbol{\psi}$.

The vorticity equation [Eq. (7.2)] can be expressed as

$$\kappa_j(x) = \sqrt{\tau_j^2 - i\omega\frac{\rho_f}{\mu}}, \tag{8}$$

where μ is the dynamic viscosity of the fluid.

For the BM partition, the plate bending equation is

$$\rho_p h w'' + \frac{\partial^2}{\partial x^2}\left(D_{11}\frac{\partial^2 w}{\partial x^2}\right) + 2\frac{\partial^2}{\partial x\partial y}\left(D_{12}\frac{\partial^2 w}{\partial x\partial y}\right) + \frac{\partial^2}{\partial y^2}\left(D_{22}\frac{\partial^2 w}{\partial y^2}\right) = p_p, \tag{9}$$

where p_p is the pressure acting on the pectinate zone, ρ_p is the density of the plate, and D_{11}, D_{12}, and D_{22} are the bending stiffness components which take into account the fiber density (f) and sandwiched construction of the BM,

$$D_{ij} = \frac{fE_{ij}}{1-\nu^2}I, \tag{10}$$

where E_{ij} is the Young's modulus, ν is Poisson's ratio, and the area moment of inertia $\boldsymbol{I}$ for symmetric layers is $I = 2\int_{h/2-g}^{h/2}\zeta^2 d\zeta$, where h is the thickness of the membrane, g is the layer thickness, and ζ is the distance through the thickness. Unlike most mammals, the gerbil BM is not symmetric, but has radial fibers concentrated in a curved tympanic band and a flat band on the OC side in the pectinate zone. Schweitzer *et al.* (1996) find that the thickness of the tympanic band correlates with the postnatal maturity of hearing. Since the fiber density for gerbil is not known, the details of the BM mechanics into an effective volume fraction f, consistent with the values from Cabezudo (1978) for cat, are lumped with $I=h^3/12$.

From the plate bending equation [Eq. (9)], the displacement profile of the partition in harmonic motion is

$$w_p(x,\eta,t) = e^{-i\omega t}W(x)\sin\left(\frac{\pi\eta}{b}\right), \tag{11}$$

where $W(x)$ is the amplitude function. The fluid and partition displacements are matched at their interface (with $W(x) = \Phi(x)$) and the coefficients $T_j(x)$ are determined from this assumed shape function of the displacement.

Integrating the pressure across the width and summing up the Fourier harmonics gives the force per unit length in the time domain for the partition and fluid. For the pectinate zone (PZ) of the partition (Lim and Steele, 2002),

$$F_{\mathrm{PZ}}(x,t) = \sum e^{-i\omega t}F_{\mathrm{PZ}}(x,\omega_j), \tag{12}$$

where $F_{\mathrm{PZ}}(x,\omega_j) = (K_p(n,x,\omega_j) - \omega_j^2 M_p(x))W(x,\omega_j)$ with the plate stiffness given by

$$K_p(n,x,\omega) = \frac{\pi}{2b}\left(-\rho_p\omega^2 h + D_{11}n^4 + 2D_{12}n^2\left(\frac{\pi}{b}\right)^2 + D_{22}\left(\frac{\pi}{b}\right)^4\right) \tag{13}$$

and mass given by

$$M_p = \frac{\pi}{2b}\rho_p h. \tag{14}$$

For the fluid

$$F^f_{\mathrm{BM}}(x,t) = \sum e^{-i\omega t}F_f(x,\omega), \tag{15}$$

where

$$F_f(x,\omega_j) = \rho_f\omega_j^2 bH_f W(x,\omega_j) = \omega_j^2 M_f W(x,\omega_j) \tag{16}$$

with M_f and H_f being the effective mass and thickness of the fluid layer over the width of the plate (Lim and Steele, 2002).

Taking into account the asymmetric fluid pressure from the two fluid ducts (SV and ST), and matching the coefficients of stresses in Eqs. (12) and (15) give

$$F_{\mathrm{PZ}}(x,t) = 2F^f_{\mathrm{BM}}(x,t). \tag{17}$$

Equation (17) represents the eikonal equation for the passive model, and it provides a physically consistent and systematic reduction of the 3D model to a one-dimensional formulation in spatial coordinates. The stiffness and mass quantities are reminiscent of those used in lumped parameter models, but these are derived from the physics, and there are no free parameters except for the exact value of volume fraction f and the effective thickness h that are selected for the frequency-to-place map.

B. Feed-forward active model

The active elements in the cochlea are presumed as the OHCs which act like piezoelectric actuators that push on the BM partition to improve the cochlea's sensitivity and frequency selectivity. In this model, the force applied by the OHCs on the BM partition is assumed to be proportional to the total force acting on the BM. Equation (17) states the total force acting on the (PZ) results from the fluid force difference across the two scalae. To include forces resulting from the OHCs' motility, an effective OHC force on the BM, F^C_{BM}, is added to Eq. (17):

$$F_{\mathrm{PZ}} = 2F^f_{\mathrm{BM}} + F^C_{\mathrm{BM}}. \tag{18}$$

The OHC force acting at $x+\Delta$ is proportional to the BM displacement sensed at x by the effect of the OHC longitudinal tilt shown in Fig. 2:

$$F^C_{\mathrm{BM}}(x+\Delta) = \alpha(x)F_{\mathrm{PZ}}(x), \tag{19}$$

where α is the feed-forward gain factor and Δ is the longitudinal distance between the apex and base of the OHC. This depends on the length of the OHC (l_{OHC}) and angle to the longitudinal direction (θ),

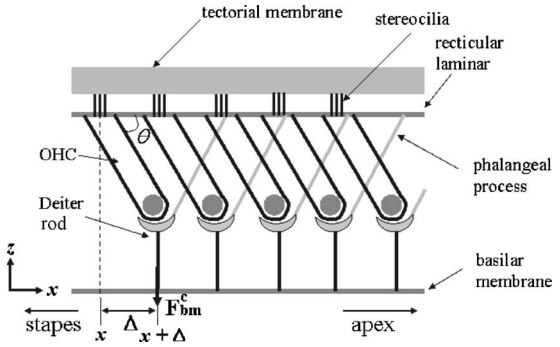


FIG. 2. Schematic of the longitudinal view of organ of Corti, showing the longitudinal tilt of the outer hair cells. The longitudinal distance between the base and apex of the outer hair cells is defined as Δ. The force on the BM to the neighboring OHCs is F^C_{BM}.

$$\Delta = l_{\mathrm{OHC}}\cos\theta. \tag{20}$$

Combining Eqs. (18) and (19) provides a relation that enters the eikonal equation for the feed-forward active model (Lim and Steele, 2002).

C. BM displacement and intracochlear pressure

1. BM displacement

Each individual equation from Eq. (17) or Eq. (18) gives an eikonal equation from which the complex wave numbers can be obtained by Newton-Raphson iterative scheme. In the present formulation, positive and negative real parts correspond to forward and backward propagating waves, respectively. The amplitude of the propagating wave can be obtained from the transport equations which are obtained by considering the volume integral over a thin slice of the duct's cross section with differential volume $\delta V = L_2 L_3 \delta x$, $\int_{\delta V}\Delta\phi dV = 0$ and the transport equation is then reduced to an ODE in x in the form of the well-known reduced wave equation

$$G_{,xx} + n(x,\omega)^2 G = 0, \tag{21}$$

in which $n(x)$ is the local wave numbers, determined by solving the 3D fluid equations for each cross section. The dependent variable $G(x)$ provides the potential $\Phi(x)$ for the fluid,

$$\Phi(x) = \frac{G(x)n}{T_0\sinh(nL_3)}, \tag{22}$$

where $T_0(x)$ is the Fourier coefficient for zeroth component of scalar potential for fluid displacement and L_3 is the height of the fluid chamber. The function $G(x)$ is obtained using a combination of the WKB asymptotic method in the short wavelength region (n is large) and the RK4 forward integration in the long wavelength region (n is small). The boundary conditions of matching the volume displacement at the stapes and zero pressure at the helicotrema are satisfied.

2. Intracochlear pressure

The pressure field in the real cochlea is a summation of two components. The first component is the traveling pressure wave where the fluid displacement is antisymmetric about the partition. The second component is a compressive

wave where the fluid displacement is symmetric. The two pressure wave components are needed to satisfy simultaneous boundary conditions located at the OW and RW.

a. Pressure from the slow traveling wave For a harmonic excitation with a frequency ω, the fluid pressure throughout the cochlear duct associated with the slow traveling wave follows from the functional form of the scalar potential for the fluid displacement that satisfies boundary conditions and the pressure acting on the fluid [Eqs. (2) and (7a) respectively]:

$$p_t(x,y,z) = \rho_f \omega^2 \sum_j \Phi(x) T_j(x) \cosh(\tau_j(x)(L_3 - z)) \times \cos\left(\frac{j\pi y}{L_2}\right). \tag{23}$$

b. Pressure from the compressive fast wave The equilibrium and continuity equations are one dimensional for the fast wave acoustics subject to time-harmonic displacements as follows:

$$\frac{dp_c}{dx} = \rho_f \omega^2 u, \tag{24}$$

$$\frac{dq}{dx} = -\frac{A}{\rho_f c^2} p_c, \tag{25}$$

where p_c is the fluid pressure associated with the compressive fast wave, u is the fluid displacement in the x direction, q is the fluid flux, and A is the cross-sectional area of the ducts. Combining Eqs. (24) and (25) yields the first-order system of differential equations:

$$\begin{bmatrix} p_c \\ q \end{bmatrix}_{,x} = \begin{bmatrix} 0 & \frac{\rho_f \omega^2}{A} \\ -\frac{A}{\rho_f c^2} & 0 \end{bmatrix} \begin{bmatrix} p_c \\ q \end{bmatrix}. \tag{26}$$

c. Intracochlear pressure from the combined waves. The total intracochlear pressure (p_{total}) is obtained by the summation of the pressure field from the slow traveling wave (p_t) and the pressure field from the compressive wave (p_c). The boundary conditions of zero total pressure at the RW and the prescribed total pressure (p_{stapes}) at the stapes yield

$$p_{\text{total}} = a_t p_t + a_c p_c, \tag{27}$$

where two unknown coefficients a_t and a_c are determined by these two boundary conditions,

$$a_t = p_{\text{stapes}} \left(\frac{p_c(x_{\text{rw}})}{p_t(x_{\text{ow}}) p_c(x_{\text{rw}}) - p_t(x_{\text{rw}}) p_c(x_{\text{ow}})} \right) \tag{28}$$

and

$$a_c = p_{\text{stapes}} \left(\frac{p_t(x_{\text{rw}})}{p_t(x_{\text{rw}}) p_c(x_{\text{ow}}) - p_t(x_{\text{ow}}) p_c(x_{\text{rw}})} \right) \tag{29}$$

with x_{rw} and x_{ow} representing the RW and OW coordinates, respectively.

III. RESULTS

The cochlear model is used to calculate the response of a gerbil cochlea. The material property values in Table I were taken from a number of sources (Smith, 1968; Lim, 1980; Miller, 1985; Steele *et al.*, 1995; Karavitaki, 2002) and the dimensions in Table II were from the anatomical measurements for gerbil cochlea (Sokolich *et al.*, 1976; Greenwood, 1990; Dannhof *et al.*, 1991; Cohen *et al.*, 1992; Edge *et al.*, 1998; Thorne *et al.*, 1999).

TABLE I. Material properties for the gerbil cochlear model.

Basilar membrane	$\rho_p = 1.0 \times 10^3$ kg/m^3
	$E_{11} = 1.0 \times 10^{-4}$ GPa
	$E_{22} = 1.0$ GPa
	$E_{12} = 0.0$ GPa
	$\nu = 0.5$
Scala fluid	$\rho_f = 1.0 \times 10^3$ kg/m^3
	$\mu = 0.7 \times 10^{-3}$ Pa s
Outer hair cell	$\theta = 60°, 80°$
	$\alpha = 0.15, 0.28$

The model is meshed into 1200 sections along the 12 mm length of the gerbil cochlea. Forty terms are used in the Fourier expansion across the width of the scala. Calculation with 80 terms for the Fourier expansion shows no difference from 40 terms. Running on an Intel Pentium IX (3.40 GHz) processor, the average time taken for a single harmonic excitation calculation is about 1 s. This method provides a fast and efficient solution compared to a full-scale finite element model. Note that the computation time indicated by Parthasarathi *et al.* (2000) is measured in hours of computing time for the linear solution for a single frequency.

The results include BF-to-place map, BM frequency response, intracochlear pressure, and derived quantities from the intracochlear pressure defined by Olson (1998); (1) BM velocity, (2) pressure difference across OC, and (3) OC impedance. The modeling results are compared with *in vivo* measurements (Olson 1998, 2001).

A. BF-to-place map

The calculation for BF versus location along the gerbil cochlear (BF range: 0.3–50 kHz) is shown in Fig. 3 with the gerbil BF-to-place map (Sokolich *et al.*, 1976; Greenwood,

TABLE II. Anatomical dimensions as a function of longitudinal position (x) for the gerbil cochlear model.

x (mm)	b (mm)[a]	h (mm)[b]	f[c]	L_2, L_3 (mm)[d]	l_{OHC} (μm)[e]
0		0.0210	0.030	1.000	25.0
1.5		0.0175		0.707	
2.9	0.162			0.387	
3.5		0.0131			
5.0				0.316	
5.9		0.0088			
7.2	0.190	0.0073		0.282	
8.4					
9.0		0.0055		0.316	
10.2	0.205	0.0044			
12.0		0.0031	0.007	0.245	65.0

[a] b: Width of plate.
[b] h: Effective thickness of plate.
[c] f: Effective fiber density of plate.
[d] L_2, L_3: Width and height of fluid chamber.
[e] l_{OHC}: Length of outer hair cell length.

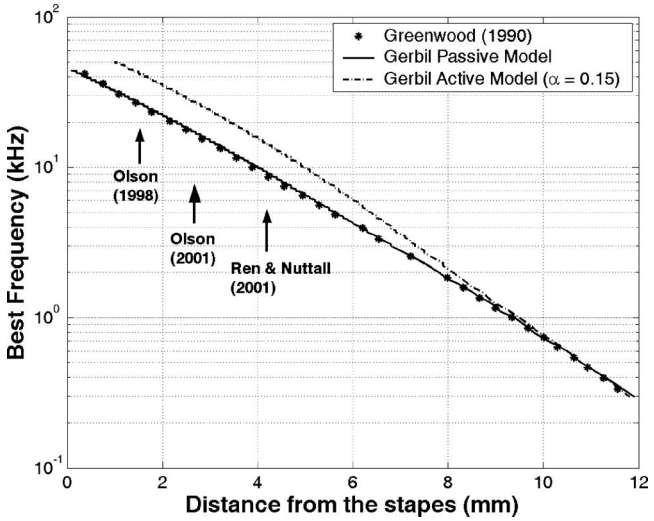


FIG. 3. Best frequency (BF) vs position for the passive cochlear model (solid line) compared to measurements (asterisk), and active cochlear model (dashed-dot line). The present 3D cochlear model represents the cochlear BF-to-place map of gerbil (Sokolich *et al.*, 1976; Greenwood, 1990) over 0.3–50 kHz range spanning a length of 12 mm.

1990) that was constructed from cochlear-microphonic recording. The BF-to-place map from the passive model and measurement are in excellent agreement (Fig. 3). The dashed-dot line represents the BF-to-place map for the feed-forward active model with 0.15 gain factor. Near the stapes (0–4 mm from the stapes), the active model shows approximately 1/2 octave higher BF, whereas there is no BF shift near the helicotrema region. Due to the lower wave number for low frequency, the feed-forward gain from the active model is less in the apical region.

B. Frequency response of BM velocity

The gerbil cochlear BM velocity magnitude and phase for 4.2 mm from the base (BF=9.5 kHz) relative to the stapes displacement are computed over a range of excitation frequencies up to 20 kHz [Figs. 4(a) and 4(b)]. Results from the model are compared with the gerbil measurements (Ren and Nuttall, 2001). The passive model shows quantitatively very good agreement with motion measured at a high stimulus level (100 dB SPL at the ear canal) in magnitude [Fig. 4(a)].

Karavitaki (2002) evaluated the angle of tilt of gerbil OHC (θ) to be approximately 84°, which is close to being perpendicular to the basilar membrane (Fig. 2). The gain from OHCs is calculated for two cases; a nominal mammalian value of $\theta=60°$ and $\theta=80°$. The active model shows fairly good agreement with data at low stimulus level (30 dB SPL at the ear canal) with 30 dB gain for either $\theta=60°$ with feed-forward gain factor $\alpha=0.15$ [dashed line in Fig. 4(a)] or $\theta=80°$ with forward gain factor $\alpha=0.28$ [dotted line in Fig. 4(a)]. Thus only a slightly higher gain, still in the physiologically reasonable range, is needed even when the OHC is nearly perpendicular to the BM.

In the relative BM velocity magnitude plot [Fig. 4(a)], BF place shifts from 9.5 kHz (passive model) to 15 kHz (active model), which is 3/5 octave higher. In the animal measurement BF is also near 9.5 kHz for the high level passive case. For the low level active case, BF place shifts to about 13 kHz, which is only about 2/5 octave higher. So the model appears to overestimate the BF for the active case.

In the model, the phase is normalized to the volume flow rate at $x=0$, as the stapes is assumed to be a piston at the end of the fluid chamber. As shown in Fig. 4(b), the phase of the response obtained from the model shows 2.5 cycles larger roll-off with frequency than the experimental measurements. In the region of the low frequency input, below 4 kHz, the BM velocity phases both from the model and measurement are similar. However, after 4 kHz, the phase of BM velocity from the model shows a larger roll-off than the phase from the data, which corresponds to a higher wave number in the model above 4 kHz. It is well known that the phase excursion to the best place is about 1.5–2 cycles, over a wide

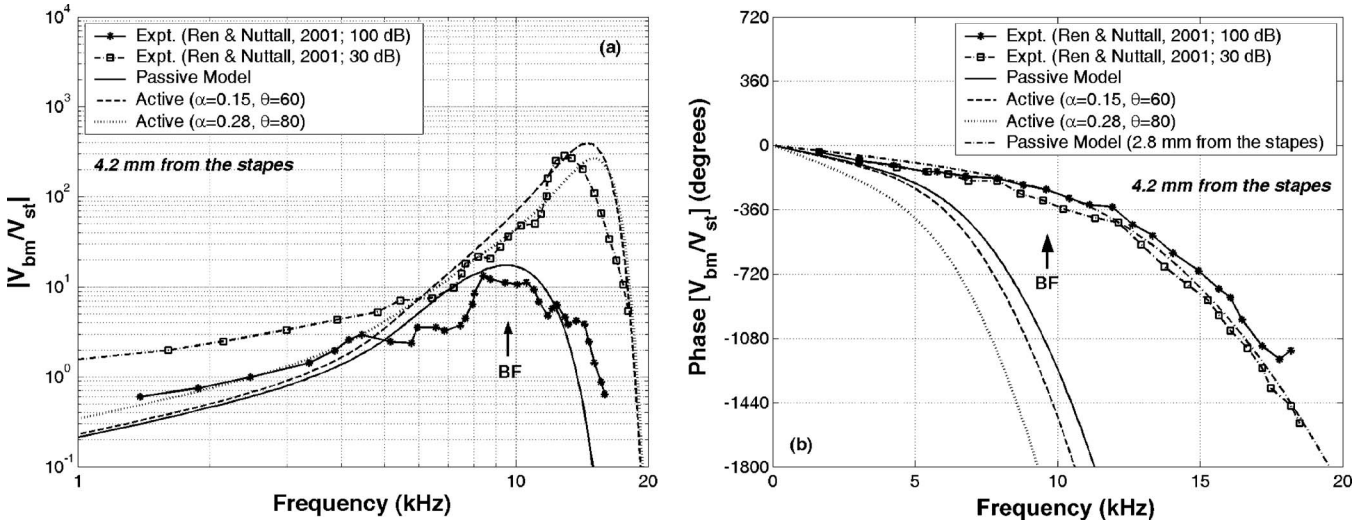


FIG. 4. Basilar membrane (BM) velocity relative to the stapes V_{bm}/V_{st} magnitude (a) and corresponding phase (b) for the gerbil cochlea at 4.2 mm from the base (BF=9.5 kHz). For the active model, $\alpha=0.15$, $\theta=60°$ (dashed line) and $\alpha=0.28$, $\theta=80°$ (dotted line) were used while for the passive case $\alpha=0$. Experimental data (expt.) for 30 and 100 dB SPL corresponding to the active and passive case, respectively, are included for comparison (Ren and Nuttall, 2001). Dashed-dot line in (b): Phase from the model at the 2.8 mm from the stapes.

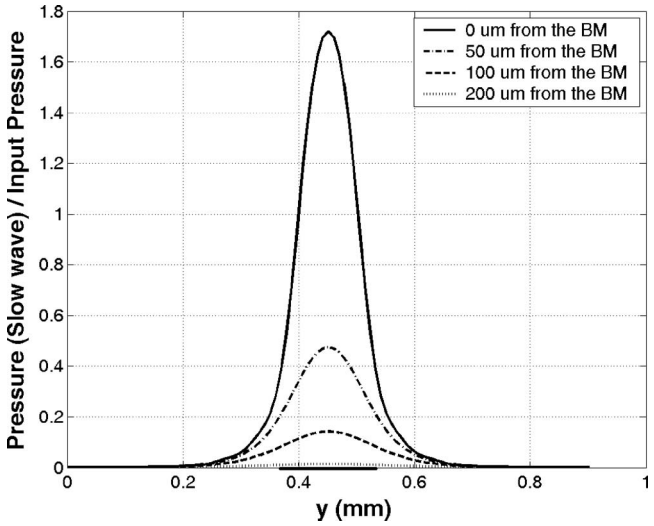


FIG. 5. Radial distribution of intracochlear pressure from the slow wave at different distances from the partition (1.4 mm from the stapes, BF =26 kHz). The location of the BM is indicated by the thickened line (BM width=0.151 mm). The pressure drops exponentially with the distance from the BM in either perpendicular or radial direction.

range of the cochlea (Ren and Nuttall, 2001; Overstreet *et al.*, 2002). Too much phase excursion in the present model may come from the unique shape of the basilar membrane in the gerbil cochlea that is not taken into account. Another possible problem is that the actual position of the stapes in the cochlea extends over a small portion of the basal end of the scala vestibuli which may result in this discrepancy in the phase. The phase calculated at 2.8 mm from the stapes [BF of 15 kHz for the passive case, dashed-dot line in Fig. 4(b)] is very close to the measurements.

C. Intracochlear pressure

The frequency response of the intracochlear pressure from the gerbil model is presented. First, the intracochlear pressure only from the slow wave is calculated for four different locations away from the BM. Next, intracochlear pressure from the combined slow and fast waves is calculated at two different locations which are 1.4 and 2.6 mm from the stapes. These intracochlear pressure simulations for the gerbil model very close to the stapes (1.4 mm from the stapes) show good agreement both in magnitude and phase with the *in vivo* measurements (Olson, 1998), whereas simulation results at 2.6 mm from the stapes shows one cycle more phase excursion at BF location than *in vivo* measurements (Olson, 2001).

1. Slow wave intracochlear pressure

Intracochlear pressure due to the motion of BM from the slow traveling wave is obtained from Eq. (23). Fig. 5 gives the pressure distribution in the SV in the section at 1.4 mm from the stapes (BF=26 kHz). The pressure decreases exponentially with distance from the BM. It is also clear that the pressure depends strongly on the transverse direction to the fluid motion, and is fully three dimensional. This confirms the results first given by Steele and Taber (1979, Fig. 10).

2. Combined slow and fast wave intracochlear pressure

a. Combined intraochlear pressure at 1.4 mm *from the stapes (passive case).* The intracochlear pressure in the ST from the passive model which includes contributions from the traveling wave solution and the compressive wave solution is shown in Figs. 6(a) and 6(b). The intracochlear pressure magnitude and phase at 1.4 mm from the stapes (BF =26 kHz) are calculated for two locations [at 3 and 23 μm away from the BM, Figs. 6(a) and 6(b)] and compared with experimental measurements (Olson, 1998).

Intracochlear pressure from the 3D cochlear model shows good agreement with measurements (Olson, 1998) both in magnitude and phase. Around the BF region, the intracochlear pressure magnitude at 3 μm away from the BM is 5 dB larger than magnitude at 23 μm away from the BM both in the model and measurement. Several distinct peaks and valleys after the peak region (30–40 kHz) are evident in Fig. 6(a). These peaks and valleys are from constructive and destructive interference between the slow and fast wave pressure. However, these are not clearly seen in the data (Olson, 1998). This difference between data and model results may come from (i) the intracochlear pressure calculated at one point compared to the experiment of pressure averaged from the region of the transducer, or (ii) the inadequate frequency sampling in the measurement since more peaks and valleys are found in the most recent intracochlear pressure measurements (Dong and Olson, 2007).

Intracochlear pressure from the model also shows a transition from the slow wave to the compressive fast wave after the BF region which is observed in the measurements. In Fig. 6(a), intracochlear pressure in the model shows that the traveling wave is dominant from the low frequency to the BF region. However, after this BF region, the traveling wave on the BM disappears and the fast acoustic wave becomes the dominant wave as is evident by the approximately constant pressure.

In Fig. 6(b), the pressure phase remains near zero until the frequency exceeds 25 kHz, when the phase changes rapidly. The decrease in phase is characteristic of the traveling wave component of the pressure field, whereas the phase plateau is characteristic of the fast compressive wave. Intracochlear pressure phase from the model at 3 μm away from the BM shows a larger phase accumulation than the measurement because of more oscillation in the model after the peak region [Fig. 6(a), solid line, 30–40 kHz].

b. Combined intraochlear pressure at 2.6 mm *from the stapes (active case).* Intracochlear pressures for the passive and active case at 2.6 mm from the stapes (BF=15 kHz) and 22 μm away from the BM are shown in Figs. 6(c) and 6(d). Model results are compared with Olson's (2001) experimental data of expt. 9-8-98-I-usual. Active model results are calculated at a low stimulus level (50 dB SPL at the ear canal); $\theta=60°$ (OHC angle), $\alpha=0.15$ which was used in the Sec. III B. In the magnitude [Fig. 6(c)], the model results and measurement data show good agreement in (i) nonlinearity with 10 dB gain from the OHCs and (ii) 2/3 octave BF shift in the active case, whereas model shows more peaks and valley than data. In this case, simulation shows 20 dB less gain than BM velocity. This smaller gain in the intracochlear pressure than BM velocity is indication that the organ of Corti is supplying additional force to the BM. Also, this im-

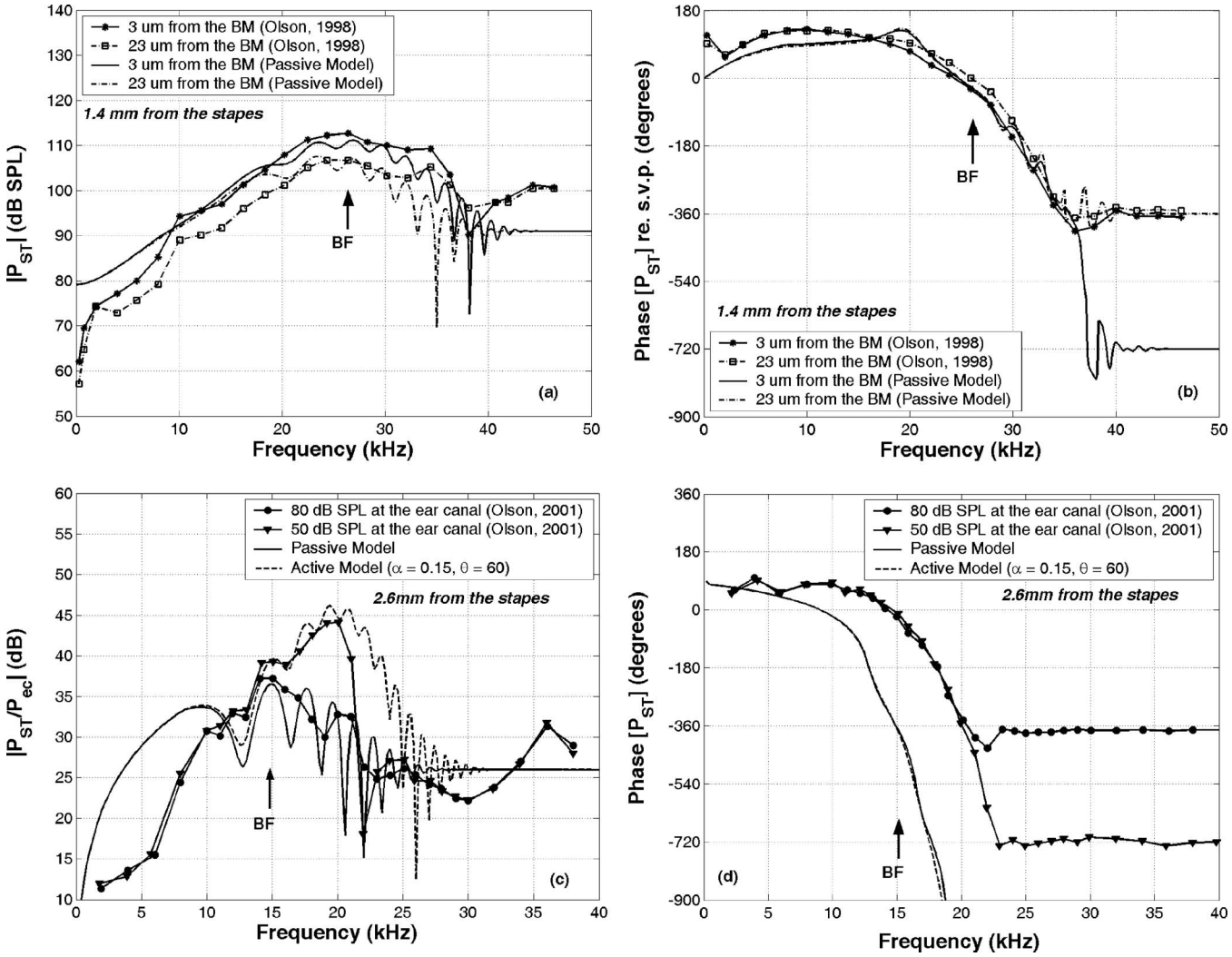


FIG. 6. Combined slow and fast wave intracochlear pressure in the scala tympani (ST) of the gerbil. (a) Intracochlear pressure magnitude and (b) corresponding phase at 1.4 mm from the stapes (BF=26 kHz) and 3 and 23 μm away from the BM (100 dB SPL at the stapes: passive). Data are from Olson, (1998, Fig. 10) expt. 2-26. (c) Intracochlear pressure magnitude and (d) corresponding phase at 2.6 mm from the stapes (BF=15 kHz) and 22 μm away from the BM (80 dB SPL: passive and 50 dB SPL: active case). Data are from Olson (2001, Fig. 7) expt. 9-8-98-I-usual.

plies that the gain from OHCs can be measured more clearly from the BM velocity than intracochlear pressure measurement.

From Fig. 6(d), the model shows approximately one cycle more phase excursion at the BF position than the measurement. The phase difference at the BF for the three different locations are: 0 cycle, 1 cycles, and 2.5 cycles at 1.4 mm (Olson, 1998), 2.6 mm (Olson, 2001), and 4.2 mm (Ren and Nuttall, 2001) from the stapes, respectively. The model shows more phase excursion than measurement with increasing distance from the stapes. These results indicate that the phase difference between model and measurement accumulates with increasing distance from the stapes. These phase excursion issues in the model may be resolved by a more advanced model.

In the following derived quantities analysis, intracochlear pressure both at 1.4 and 2.6 mm from the stapes are used. Especially, intracochlear pressure for the low stimulus level (40 dB at the ear canal) at 1.4 mm from the stapes (Olson, 1998) can be considered as nearly passive since the condition of the cochlea was not optimal. However, there is still a small gain in the low stimulus level in the measurement. Thus, a small gain factor (α=0.05) is used for this case. Since the intracochlear pressure simulation at 1.4 mm from the stapes shows the best agreement both in the phase and magnitude with data, the simulation results at 1.4 mm from the stapes are used for the analysis of exact theoretical OC impedance.

D. Derived quantities from the intracochlear pressure

Olson (1998) developed formulas in terms of the measured intracochlear pressure as approximations for the BM velocity, the pressure difference across the OC, and the OC impedance (Z_{OC}). Since the results from the pressure differences show interesting behavior, the calculation using Olson's formulas and Olson's results are compared directly. Finally, the exact theoretical OC impedance from the model for the passive and active cases is calculated and compared to the result of the difference formula for the estimated theoretical OC impedance.

In this section, quantities derived from the intracochlear pressure as defined by Olson (1998) are calculated and compared with measurement (Olson, 1998, 2001). These are (1) *BM velocity*, (2) *pressure difference across OC*, and (3) OC

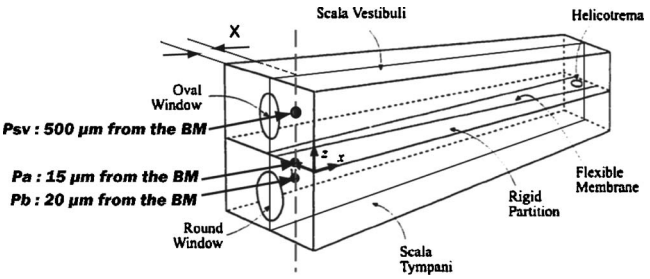


FIG. 7. Schematic 3D drawing of the cochlear model. Intracochlear pressures P_a and P_b are measured and calculated at the indicated positions a and b, 15 and 20 μm away from the BM in the ST respectively, and in SV (P_{SV}). These are used to obtain an approximation (Olson, 1998, 2001) for the BM velocity, pressure difference and impedence, and organ of Corti impedance referred to as derived quantities. Cross indicates distance from the stapes.

impedance. The derivation was presented in a previous study and derived results in this section are based on that study (Olson, 1998).

Derived quantities are estimated from a pair of adjacent ST intracochlear pressures, which are 15 μm (P_a) and 20 μm (P_b) from the BM (Fig. 7). The difference between these two adjacent intracochlear pressures is used to estimate the vertical (z) component of the intracochlear pressure gradient. First, the *BM velocity* (v_{BM}) is found from this intracochlear pressure gradient. Based on the estimation that close to the BM the fluid moves with BM, v_{BM} is considered to be this fluid velocity. From the dimensional analysis, the Navier–Stokes equation in an incompressible fluid is simplified as

$$\nabla P = -\rho_f \frac{\partial v}{\partial t}. \tag{30}$$

For the z component,

$$\frac{\partial P}{\partial z} \approx \frac{(P_a - P_b)}{\Delta z} = -\rho_f \frac{\partial v_z}{\partial t}. \tag{31}$$

By assuming $v_{BM}=v_z$ for a harmonic response,

$$v_{BM} = -\frac{(P_a - P_b)}{\rho_f \omega \Delta z} i \tag{32}$$

was given by Olson for the calculation of v_{BM} from the intracochlear pressure measurements.

Second, the *pressure difference across* OC (ΔP_{OC}) is obtained by assuming zero pressure at the round window (Olson, 1998),

$$\Delta P_{OC} = P_{SV} - 2P_a \tag{33}$$

where P_{sv} is the intracochlear pressure near the stapes at 500 μm from the BM in the scala vestibule. Finally, the *OC impedance* (Z_{OC}) is defined as

$$Z_{OC} = \frac{\Delta P_{OC}}{v_{BM}}. \tag{34}$$

Derived quantities from the gerbil cochlear model are compared with results of expt. 2-26 (Olson, 1998) for the passive case and expt. 9-8-98-I-usual (Olson, 2001) for the more active case.

1. Derived BM velocity

Numerous measurements of basal BM velocity, v_{BM}, have been conducted, and the frequency domain response has been studied under various physiological conditions (Khanna and Loenard, 1986; Cooper and Rhode, 1992a, 1992b; Xue *et al.*, 1995; Nuttall and Dolan, 1996; Ruggero *et al.*, 1996, 1997; Overstreet *et al.*, 2002). Derived v_{BM} from animal intracochlear pressure measurements (2–26) from Olson (1998) represented a similar response to the v_{BM} measured by direct measurement methods by Xue *et al.* (1995).

Figure 8(a) displays the magnitude of the v_{BM} from the model and Olson's 1998 expt. 2-26 measurement. Figure 8(b) shows the v_{BM} phase relative to SV pressure. Stimulus levels were 80 and 40 dB SPL in the ear canal for the measurements, which corresponds to 110 dB and 70 dB SPL at the stapes for the model with the consideration of 30 dB gain from the middle ear ossicles (Olson, 1998). For a moderate OHCs' motility force on the BM, the active model with 0.05 gain factor is used in the calculation of the low level stimulus (40 dB SPL at the ear canal). The derived v_{BM} from the model at 1.4 mm from the stapes shows excellent agreement with the experimental results both in the magnitude [Fig. 8(a)] and phase [Fig. 8(b)]: (i) their peaked shape, (ii) maximum peak region, (iii) the absolute value of maxima, (iv) dropping off rapidly after the peak in the v_{BM} magnitude, and (v) decreasing phase with increasing frequency at a steadily increasing rate above about 8 kHz.

Figures 8(c) and 8(d) show magnitude and phase of the v_{BM} from the model and Olson's 2001 expt. 9-8-98-I-usual measurement. For the passive case, derived BM velocity at 2.6 mm from the stapes shows good agreement with measurement up to 25 kHz. The measurement shows plateau after 25 kHz, which is not observed in the model both in the passive and active cases. However, this is expected because the fast wave should be canceled out in the calculation of derived BM velocity [Eq. (32)]. For the low level stimulus (40 dB SPL at the ear canal), the active model with $\theta=60°$ (OHC angle) and $\alpha=0.15$ is used. Derived BM velocity for the active case shows good agreement with measurement up to BF region. After the BF region, the derived BM velocity magnitude from the active model shows less decrease than the data. The active case simulation also shows no plateau after 25 kHz for the same reason mentioned earlier. Phase of derived BM velocity at 2.6 mm from the stapes shows approximately one cycle more phase excursion than data. The reasons for this are not clear at this point and need to be resolved.

2. Derived pressure across OC

The derived pressure difference across OC (ΔP_{OC}) simulation and experimental results both in the magnitude and phase are shown in Fig. 9. For the calculation of ΔP_{OC}, $P_{sv}-2P_a$ from the current symmetric cochlear model is used [Eq. (33)]. P_a is obtained at the position of 15 μm from the BM in the ST. ΔP_{OC} values from the model are calculated and compared with measurements by Olson (1998, 2001).

ΔP_{OC} magnitude at 1.4 mm from the stapes [Fig. 9(a)] shows (i) a mild peak around the BF region, (ii) notches, and

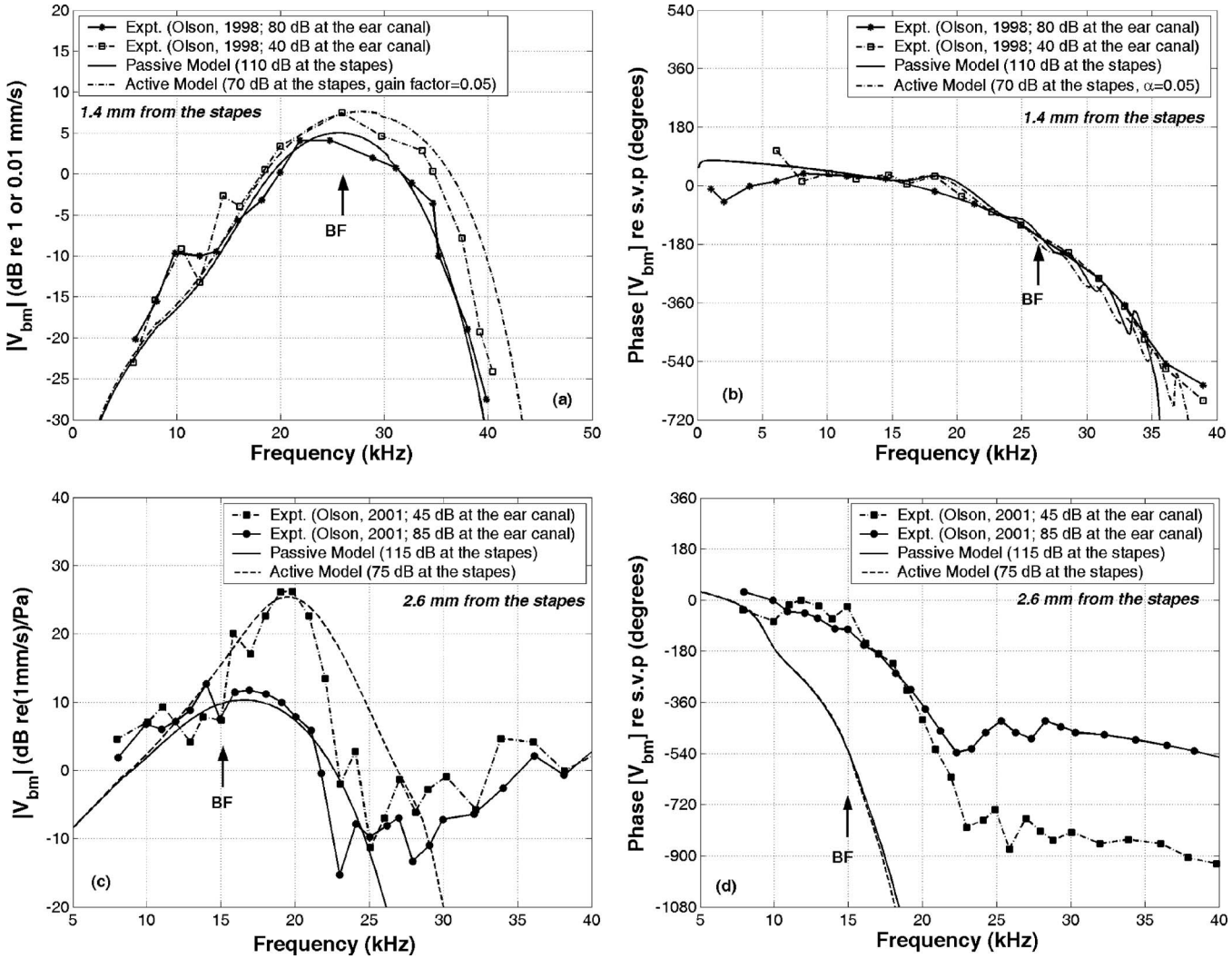


FIG. 8. Derived BM velocity from the gerbil cochlear model and measurements, using the formulas in Olson [1998, Appendix 1, Eq. (A7)]. (a) Magnitude of the measurement results for 40 and 80 dB SPL at the ear canal and model results at 70 and 110 dB at the stapes are plotted re: 0.01 and 1 mm/s, respectively (1.4 mm from the stapes, BF=26 kHz). (b) Phase relative to SV pressure at 1.4 mm from the stapes. Data are from Olson, (1998, Fig. 18) expt. 2-26 at the stimulus levels of 40 and 80 dB SPL at the ear canal. (c) Derived BM velocity magnitude and (d) corresponding phase at 2.6 mm from the stapes (BF=15 kHz) for the passive (85 dB SPL) and active case (45 dB SPL). Data are from Olson (2001, Figs. 15(a) and 15(b)] expt. 9-8-98-I-usual.

(iii) small value in the 0–15 kHz region due to long wave. Long wave has small BM velocity value in the z direction, which induces the small pressure difference across organ of Corti. On the other hand, in the BF region which provides better measurement condition than other frequency region, simulation results show good agreement with measurement.

The estimated ΔP_{OC} magnitude from the model shows decrease from the BF region to 40–45 kHz whereas ΔP_{OC} from the measurements remains fairly flat. Above the 45 kHz region where the fast wave dominates, simulation results of the estimated ΔP_{OC} shows the plateau which comes from the fast wave in the estimation of ΔP_{OC} [Fig. 9(a)]. In Fig. 9(b), the phase is shown relative to the SV pressure. Since the SV pressure is much greater than the ST pressure at low frequencies, the phase is close to zero. Between 20 and 30 kHz, the phase accumulates almost 400°. Similar to ST pressure, the derived ΔP_{OC} is composed of the sum of a fast and slow wave [Eq. (33)]. However, for the exact calculation of ΔP_{OC}, the fast wave component should be canceled out since the magnitudes of the pressure from the fast wave in SV and ST at the same x position are equal.

ΔP_{OC} at 2.6 mm from the stapes is calculated and compared with measurements (Olson, 2001) for the high and low level stimulus which correspond to the passive and active case, respectively. In Fig. 9(c), both simulation and measurement for the passive case (85 dB SPL at the ear canal) show (i) a mild peak near the BF (15 kHz) region, (ii) drop after this region about 15 dB, and (iii) plateau after 23 kHz which is from the fast wave. More peaks and valleys in the simulation of ΔP_{OC} than measurements are still observed. Simulation of ΔP_{OC} magnitude for the low level stimulus (45 dB SPL at the ear canal) shows (i) a peak at 19 kHz with 10 dB gain from the feed-forward mechanism, (ii) less sharp tuning than data near the BF region, (iii) plateau after 30 kHz which is one octave higher than measurement, and (iv) 2/3 octave BF shift in active case. ΔP_{OC} phases relative

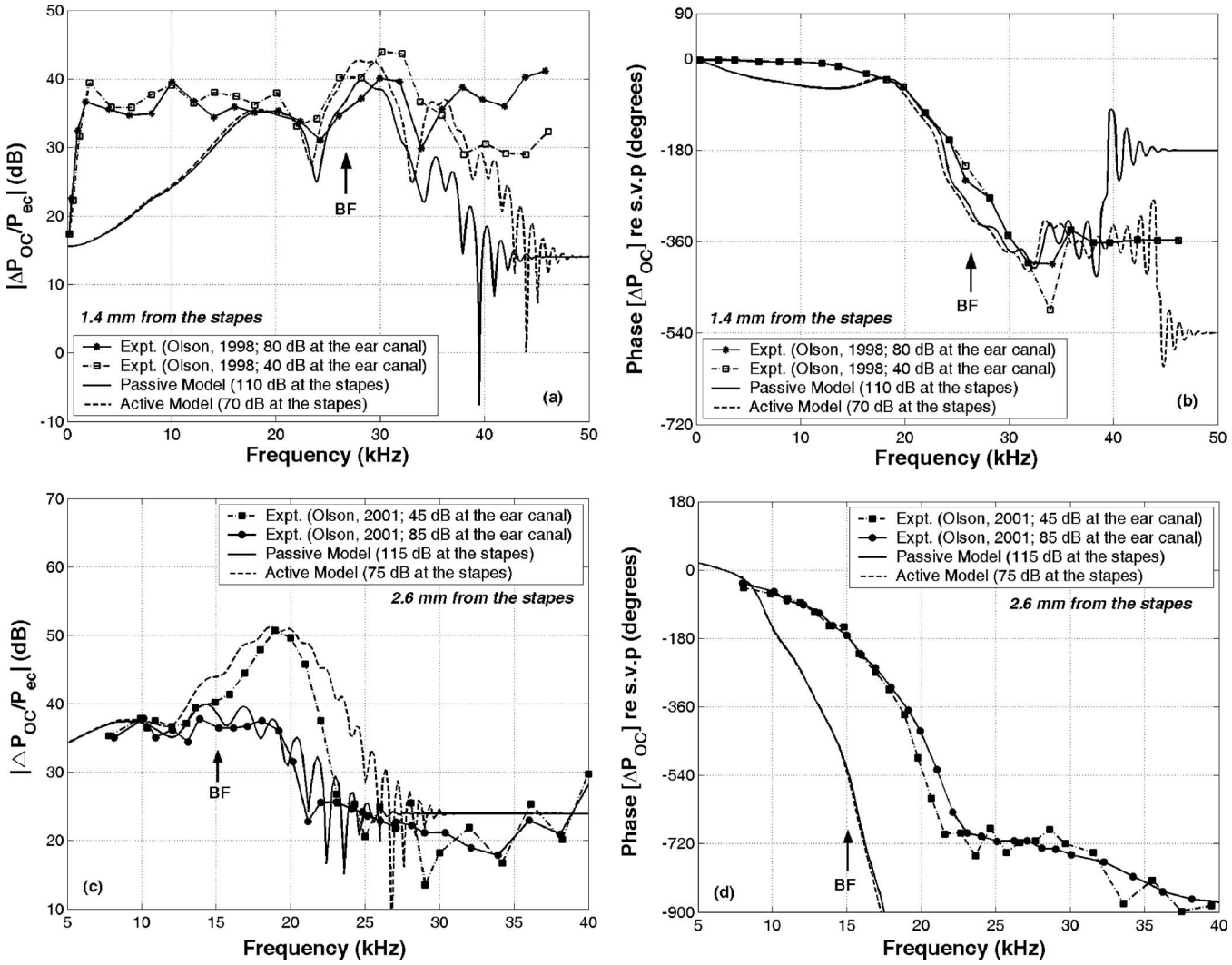


FIG. 9. Derived pressure across the OC complex, ΔP_{OC}, from the gerbil cochlear model and measurements, using the formulas in Olson [1998, Appendix 2, Eq. (A10)]. (a) Magnitude. (b) Phase relative to SV pressure at 1.4 mm from the stapes (BF=26 kHz). Data are from Olson (1998, Fig. 19) expt. 2-26 at the stimulus levels of 40 and 80 dB SPL at the ear canal. (c) Derived ΔP_{OC} magnitude and (d) corresponding phase at 2.6 mm from the stapes (BF=15 kHz) for the passive (85 dB SPL) and active case (45 dB SPL). Data are from Olson [2001, Figs. 15(a) and 15(c)] expt. 9-8-98-I-usual.

to the SV pressure are calculated and compared with measurements for the passive and active case [Fig. 9(d)]. Near the BF reason, phases show one cycle more excursion than measurements both in the active and passive case.

3. Derived OC impedance

Olson (1998) estimated the specific acoustic impedance of the OC (Z_{OC}) which is defined as ΔP_{OC} divided by v_{BM} [Eq. (34)]. The same approach in the gerbil model is conducted to allow quantitative comparison with Olson's Z_{OC} estimation. First, the derived Z_{OC} is studied by comparing the gerbil model to measurements. Next, the exact value of Z_{OC} is compared with the derived Z_{OC}. Last, the values of Z_{OC} of the passive and active models are discussed based on the exact calculations.

The magnitude [Fig. 10(a)] and phase [Fig. 10(b)] of the Z_{OC} is shown for the gerbil model and 2–26 animal measurements which were obtained near the stapes (Olson, 1998). The model results show good qualitative and quantitative agreement with measurements: (i) a tuning to frequencies between 22 and 26 kHz, (ii) a primary minimum [5 Pa/(mm/s)] at 24 kHz, with secondary minimum close to half an octave above at 32 kHz, (iii) constant slope of the magnitude in the low frequency region (−6 dB/oct) which represents stiffness dominated impedance, and (iv) phase fluctuation after the BF.

The estimated real and imaginary part of Z_{OC} for the passive and active case at 2.6 mm from the stapes are calculated and compared with Olson's 2001 9-8-98 measurements [Figs. 10(c) and 10(d)]. The real part of estimated Z_{OC} from the model stays near zero values [±10 Pa/(mm/s)] and the imaginary part of estimated Z_{OC} stays at a negative value over the whole frequency range (5–25 kHz). This shows that the estimated Z_{OC} phases for the passive and active model stay near −90° (stiffness dominated region). Magnitude of Z_{OC} can be calculated (not shown). Measured Z_{OC} magnitudes decrease up to 20 kHz then increase after 20 kHz in both the active and passive cases. On the other hand, magnitude from the passive model decreases by −15 dB/oct up to 13 kHz and increases after 23 kHz whereas magnitude

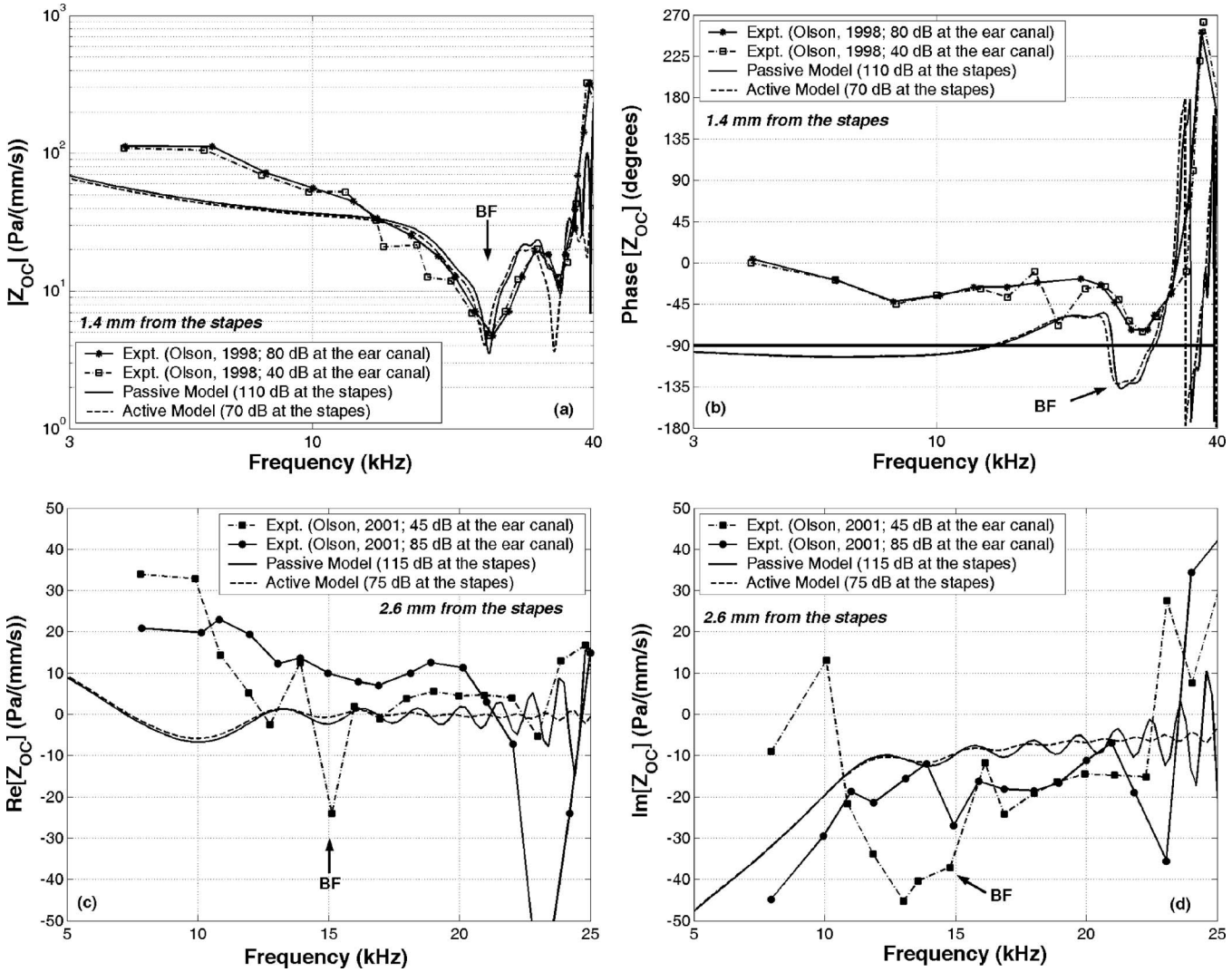


FIG. 10. Derived impedance of organ of Corti (Z_{OC}) from the gerbil cochlear model and measurements, using the formulas in Olson (1998). (a) Magnitude. (b) Phase of Z_{OC} for model and measurements at 1.4 mm from the stapes (BF=26 kHz). Data are from Olson (1998, Fig. 20) expt. 2-26. (c) Real part of Z_{OC} and (d) imaginary part of Z_{OC} at 2.6 mm from the stapes (BF=15 kHz) for the passive (85 dB SPL) and active case (45 dB SPL). Data are from Olson [2001, Figs. 15(d) and 15(e)] expt. 9-8-98-I-usual.

from the active model decreases over the whole frequency (5–25 kHz) with different decreasing rate: −15 dB/oct (5–13 kHz) and −3 dB/oct (13–25 kHz).

The impedance phase of a classical mechanical resonance begins at −90° (stiffness dominated region) at low frequencies, then increase through 0° at the resonance frequency, and ends up at +90° (mass dominated region). The phase of gerbil model stays near −90° up to 10 kHz, which indicates stiffness dominated response whereas measurement impedance phase shows a more complicated value which is considered as a combination of stiffness and damping. Most notably, in the phase results at 1.4 mm from the stapes [Fig. 10(b)], the derived Z_{OC} phase of the passive model and measurement (80 dB SPL) appears beyond the region between +90° and −90°; this indicates that the real part of the impedance is negative. Since the passive system always has positive real components of impedance, the Z_{OC} phase for the passive model and measurement (80 dB SPL) should stay between +90° and −90°. This discrepancy may come from the estimation approach, thus it can be resolved by finding the exact Z_{OC} from the model.

4. Comparison between estimated and exact theoretical OC impedance

Near the stapes (1.4 mm from the stapes), the physiologically based three-dimensional gerbil cochlear model shows the best agreement with measured derived quantities. With this approach, exact theoretical value of those quantities can be studied at this location. In this section, the difference between exact and estimated theoretical OC impedance is discussed.

The derived ΔP_{OC} includes the compressive fast wave because it is estimated by $P_{sv}-2P_{st}$ which contains the fast wave component. This fast wave in the derived ΔP_{OC} causes deep notches in the Z_{OC} magnitude [Fig. 10(a)] and Z_{OC} phase fluctuation [Fig. 10(b)]. However, the fast wave component should be canceled out in the exact ΔP_{OC} which is calculated from P_{sv}-$P_{st}|_{z=0}$.

In Fig. 11, the exact and estimated theoretical Z_{OC} are shown. Magnitude [Fig. 11(a)] and phase [Fig. 11(b)] of the exact and estimated theoretical Z_{OC} for the passive case are compared. The exact theoretical Z_{OC} was obtained from the

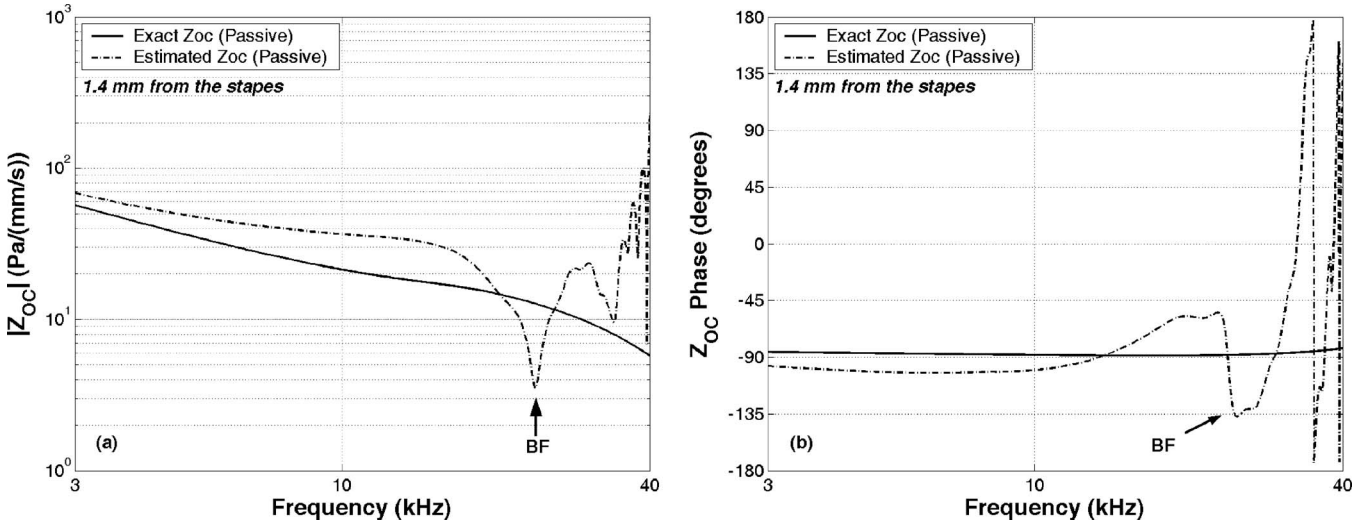


FIG. 11. Exact and derived theoretical impedance of organ of Corti (Z_{OC}) from the gerbil cochlear model (1.4 mm from the stapes, BF=26 kHz). The passive model is presented. (a) Magnitude. (b) Phase.

$v_{BM}|_{z=0}$ and $\Delta P_{OC}|_{z=0}$. The exact theoretical Z_{OC} shows: (i) smooth decrease with almost constant slope in the whole frequency region (−6 dB/oct) which represents stiffness dominated impedance, (ii) lower magnitude than estimated theoretical Z_{OC} except for the BF region, and (iii) the expected phase between −45° and −90°. The magnitude of the exact theoretical Z_{OC} decreases smoothly with increasing frequency without the fast wave mode [Fig. 11(a)]. Less exact theoretical Z_{OC} magnitude implies that larger v_{BM} and/or smaller ΔP_{OC} in the real case than derived theoretical quantities.

The exact theoretical Z_{OC} phase decrease from −50° at the low frequency region to −85° at the high frequency region. This corresponds to OC impedance that becomes more stiffness dominated near the BF. Also, the exact theoretical Z_{OC} phase remains between −90° and 90°, which represents passive mechanics [Fig. 11(b)].

5. Comparison between exact theoretical passive and active OC impedance

In Fig. 12, the magnitude [Fig. 12(a)] and the phase [Fig. 12(b)] of the exact theoretical Z_{OC} are shown for the passive and active cases. The active model (θ=60°, α =0.15) has a lower Z_{OC} magnitude than the passive case in Fig. 12(a). This difference is due to the lower slow wave pressure gain than BM velocity gain, as was discussed in Sec. III C 2. An octave or more below BF, in the tail-frequency region, the active model shows more compliance than the passive model with a magnitude difference of 2 dB. Similar differences were observed due to medial efferent activation on auditory-nerve responses in tail-frequency region by Stankovic and Guinan (1999, 2000). In Fig. 12(b), the phase of the exact theoretical Z_{OC} from the active model is below −90° in the BF region where that of the passive model

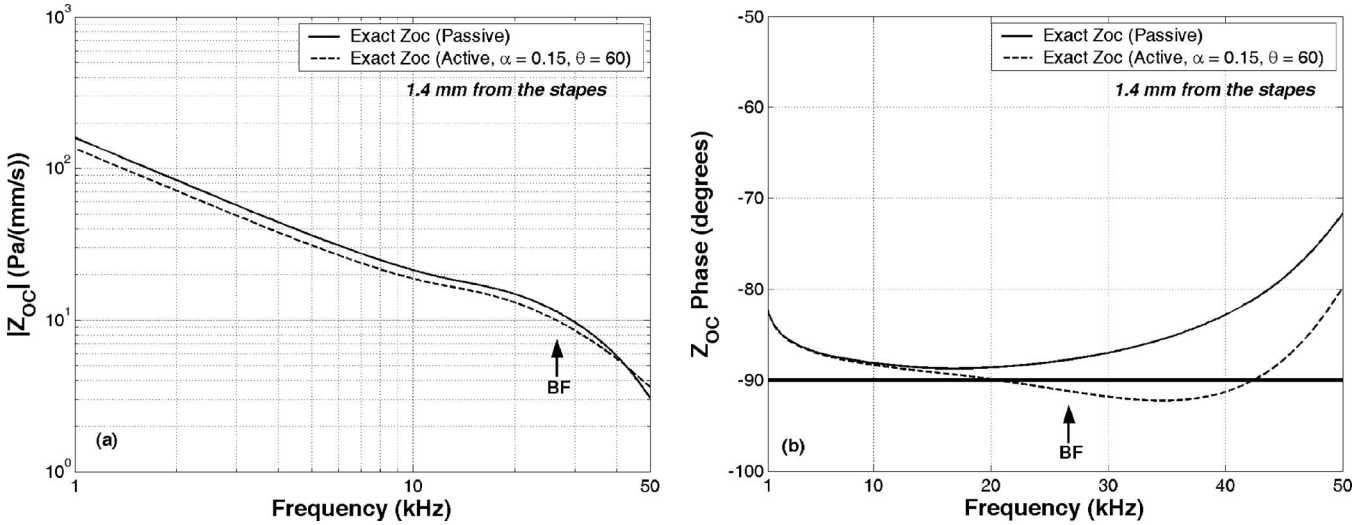


FIG. 12. Exact theoretical impedance of organ of Corti (Z_{OC}) from the gerbil cochlear passive and active model (1.4 mm from the stapes, BF=26 kHz). The 0.15 feed-forward gain factor is used in the active model. (a) Magnitude. (b) Phase.

is above −90°. Below −90° phase represents negative real component of the OC impedance which is from the force acting on the BM due to OHCs motility.

IV. DISCUSSION

The measurements of gerbil intracochlear pressure (Olson, 1998, 2001) offer an excellent opportunity to test model calculations. Presently, the macromechanical cochlear model for the chinchilla anatomy (Yoon *et al.*, 2006) is extended to the gerbil anatomy. The BM properties are physical, with orthotropic elastic properties and no fictitious mass or damping.

The comparison of results from the model and experiment is promising, but not fully satisfactory. Using the single set of anatomically based parameters, the model predicts several significant features of the cochlea. The BF-to-place map in the passive model and frequency responses of BM velocity and intracochlear pressure were in close agreement with those observed in animal measurement. The feed-forward linear active model, the most speculative feature of the framework presented, showed excellent agreement with experimental data in the BM relative velocity and intracochlear pressure magnitude. However, the calculated phases for the BM velocity and intracochlear pressure show a larger excursion at the BF by 2.5 cycles and 1 cycle at a fixed point; $x=4.2$ mm and $x=2.6$ mm, respectively. In contrast, the calculated amplitude and phase show excellent agreement for the fixed point ($x=1.4$ mm). This phase excursion issue in the current model should be improved. Preliminary results from the most recent model which has a modified BM plate and push-pull mechanism for the organ of Corti may provide better results in the large phase excursion phenomenon.

By virtue of the 3D cochlear model, intracochlear pressure in the ST was obtained by adding the fast wave to the traveling pressure slow wave. From the intracochlear pressure simulation, derived quantities: (1) *BM velocity*, (2) *pressure difference across OC*, and (3) *OC impedance* in the base, were calculated by following Olson's estimation (1998). These quantities were compared with animal measurements and showed excellent agreement. From the validated gerbil cochlear model, the exact theoretical OC impedance was obtained and compared with the estimated theoretical OC impedance. By comparing exact and estimated theoretical OC impedances, a fast wave component in the estimated theoretical OC impedance is found and it causes phase fluctuation out of the reasonable range (negative real part of impedance in the passive response) and notches in the estimated theoretical OC impedance. Finally, the exact theoretical OC impedances for the passive and active model were compared in the magnitude and phase. The exact theoretical OC impedance from the active model shows negative real components which represents active process from the OHCs' motility.

An important consideration of the feed-forward active mechanism is that BM impedance for the active case is 2 dB less than for the passive case below BF in the tail region. Near BF, the change in impedance is due to an apical shift in resonance for the low-level active case, but the magnitude and phase change is also small (Fig. 12). This indicates that the zero crossings of the time domain response for the high level passive case and the low level active case will be nearly invariant. This suggests that force generation by OHCs in the feed-forward formalism satisfies the near-invariance of fine time structure of the organ of Corti response predicted by Shera (2001). Further calculations of the model in the time domain will provide a more definitive test.

V. CONCLUSION

In the current work, the pressure in the cochlear fluid computed from the model is found to agree with the intracochlear pressure measurements (Olson, 1998, 2001). This gives support to our proposition that the present model is close to the actual behavior of the gerbil cochlea, and that the remaining discrepancies can be resolved. Extension of the cochlear model can be achieved by including more detailed structures of the OC to the current model (Steele and Puria, 2005).

ACKNOWLEDGMENTS

This work was funded by HFSP Grant No. RGP0051 and NIDCD of NIH Grant No. DC007910.

NOMENCLATURE

BM	=	basilar membrane
OC	=	organ of Corti
OHCs	=	outer hair cells
BF, EC	=	characteristic frequency, ear canal
F_{PZ}, F^f_{BM}	=	forces acting on the pectinate zone and fluid
F^C_{BM}	=	force exerted by OHC
α	=	feed-forward gain factor
l_{OHC}	=	length of OHC
θ	=	angle of tilt of OHC
n	=	wave number
ϕ	=	scalar potential for fluid displacement
Φ	=	coefficient of ϕ
T_j	=	Fourier coefficient for jth component of ϕ
τ_j	=	decay coefficient for jth component of ϕ
L_2, L_3	=	width and height of fluid chamber
p_t, p_c	=	fluid pressure associated with the slow traveling and compressive fast wave
u	=	fluid displacement in the x direction
q	=	fluid flux
A	=	cross-sectional area of ducts
ρ_f	=	fluid density
t	=	time
x, y, z	=	Cartesian coordinates
ω	=	angular velocity
Z_{oc}	=	mechanical impedance of the OC
ΔP_{OC}	=	pressure difference across the OC
v_{BM}	=	z component of BM velocity
P_{sv}	=	pressure near the stapes in the scala vestibuli
P_{st}	=	pressure near the BM in the scala tympani

Böhnke, F., and Arnold, W. (**1998**). "Nonlinear mechanics of the organ of Corti caused by Deiters cells," IEEE Trans. Biomed. Eng. **45**, 1227–1233.
Cabezudo, L. M. (**1978**). "The ultrastructure of the basilar membrane in the

cat," Acta Oto-Laryngol. **86**, 160–175.
Cohen, Y. E., Bacon, C. K., and Saunders, J. C. (**1992**). "Middle ear development III. Morphometric changes in the conducting apparatus of the Mongolian gerbil," Hear. Res. **62**, 187–193.
Cooper, N. P., and Rhode, W. S. (**1992a**). "Basilar membrane mechanics in the hook region of cat and guinea-pig cochlea: Sharp tuning and nonlinearity in the absence of baseline position shifts," Hear. Res. **63**, 163–190.
Cooper, N. P., and Rhode, W. S. (**1992b**). "Basilar membrane tonotopicity in the hook region of the cat cochlea," Hear. Res. **63**, 191–196.
Dannhof, B. J., Roth, B., and Bruns, V. (**1991**). "Length of hair cells as a measure of frequency representation in the mammalian inner ear," Naturwiss. **78**, 570–573.
de Boer, E. (**1983**). "On active and passive cochlear models: Towards a generalized analysis," J. Acoust. Soc. Am. **73**, 574–576.
Dong, W., and Olson, E. S. (**2007**). "Relating intracochlear pressure to cochlea emissions," 30th ARO Midwinter Research Meeting, Denver, CO.
Edge, R. M., Evans, B. N., Pearce, M., Richter, C. P., Hu, X., and Dallos, P. S. (**1998**). "Morphology of the unfixed cochlea," Hear. Res. **124**, 1–16.
Geisler, C. D., and Sang, C. (**1995**). "A cochlear model using feed-forward outer-hair-cell forces," Hear. Res. **86**, 132–146.
Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species-29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.
Kanis, L. J., and de Boer, E. (**1996**). "Comparing frequency-domain with time-domain solutions for a locally active nonlinear model of the cochlea," J. Acoust. Soc. Am. **100**, 2543–2546.
Kanis, L. J., and de Boer, E. (**1997**). "Frequency dependence of acoustic distortion products in a locally active model of the cochlea," J. Acoust. Soc. Am. **101**, 1527–1531.
Karavitaki, K. D. (2002). "Measurements and models of electrically-evoked motion in the gerbil organ of Corti," Ph.D. thesis, MIT, Cambridge.
Khanna, S. M., and Leonard, D. G. B. (**1986**). "Relationship between basilar membrane tuning and hair cell condition," Hear. Res. **23**, 55–70.
Kolston, P. J., and Ashmore, J. F. (**1996**). "Finite element micromechanical modeling of the cochlea in three dimensions," J. Acoust. Soc. Am. **99**, 455–467.
Lim, D. J. (**1980**). "Cochlear anatomy related to cochlear micromechanics A review," J. Acoust. Soc. Am. **67**, 1686–1695.
Lim, K. M., and Steele, C. R. (**2002**). "A three-dimensional nonlinear active cochlear model analyzed by the WKB-numeric method," Hear. Res. **170**, 190–205.
Loh, C. H. (**1983**). "Multiple scale analysis of the spirally coiled cochlea," J. Acoust. Soc. Am. **74**, 95–103.
Miller, C. E. (**1985**). "Structural implications of basilar membrane compliance measurements," J. Acoust. Soc. Am. **77**, 1465–1474.
Neely, S. T. (**1985**). "Mathematical modeling of cochlear mechanics," J. Acoust. Soc. Am. **78**, 345–352.
Neely, S. T. (**1993**). "A model of cochlear mechanics with outer hair cell motility," J. Acoust. Soc. Am. **94**, 137–146.
Nuttall, A. L., and Dolan, D. F. (**1996**). "Steady-state sinusoidal responses of the basilar membrane in guinea pig," J. Acoust. Soc. Am. **99**, 1556–1565.
Olson, E. S. (**1998**). "Observing middle and inner ear mechanics with novel intacochlea pressure sensors," J. Acoust. Soc. Am. **103**, 3445–3463.
Olson, E. S. (**2001**). "Intracochlear pressure measurements related to cochlear tuning," J. Acoust. Soc. Am. **110**, 349–367.
Overstreet, E. H., Temchin, A. N., and Ruggero, M. A. (**2002**). "Basilar membrane vibrations near the round window of the gerbil cochlea," J. Assoc. Res. Otolaryngol. **3**, 351–361.
Parthasarathi, A. A., Grosh, K., and Nuttall, A. L. (**2000**). "Three-dimensional numerical modeling for global cochlear dynamics," J. Acoust. Soc. Am. **107**, 474–485.
Peterson, L. C., and Bogert, B. P. (**1950**). "A dynamic theory of the cochlea," J. Acoust. Soc. Am. **22**, 369–381.
Ren, T., and Nuttall, A. L. (**2001**). "Basilar membrane vibration in the basal turn of the sensitive gerbil cochlea," Hear. Res. **151**, 48–60.
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., Robles, L., and Recio, A. (**1996**). "The effects of acoustic trauma, other cochlear injury, and death on basilar-membrane responses to sound," in *Scientific Basis of Noise-Induced Hearing Loss*, edited by A. Axelsson *et al.* (Thieme Medical, New York), pp. 23–35.
Schweitzer, L., Lutz, C., Hobbs, M., and Weaver, S. P. (**1996**). "Anatomical correlates of the passive properties underlying the developmental shift in the frequency map of the mammalian cochlea," Hear. Res. **97**, 84–94.
Shera, C. A. (**2001**). "Intensity-invariance of fine time structure in basilar-membrane click responses: Implications for cochlear mechanics," J. Acoust. Soc. Am. **110**, 332–348.
Smith, C. A. (**1968**). "Ultrastructure of the organ of Corti," Adv. Sci. **24**, 419–433.
Sokolich, W. G., Hamernik, R. P., Zwisiocki, J. J., and Schmiedt, R. A. (**1976**). "lnferred response polarities of cochlear hair cells," J. Acoust. Soc. Am. **59**, 963–979.
Stankovic, K. M., and Guinan, J. J. (**1999**). "Medial efferent effects on auditory-nerve responses to tail-frequency tones. I. Rate reduction," J. Acoust. Soc. Am. **106**, 857–869.
Stankovic, K. M., and Guinan, J. J. (**2000**). "Medial efferent effects on auditory-nerve responses to tail-frequency tones. II. Alteration of phase," J. Acoust. Soc. Am. **108**, 664–678.
Steele, C. R., Baker, G., Tolomeo, J., and Zetes, D. (**1993**). "Electromechanical models of the outer hair cell," in *Proceedings of the International Symposium on Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 207–215.
Steele, C. R., Baker, G., Tolomeo, J., and Zetes, D. (**1995**). "Cochlear mechanics," in *The Biomedical Engineering Handbook*, edited by Z. D. Bronzino (CRC Press, New York), pp. 505–516.
Steele, C. R., and Lim, K. M. (**1999**). "Cochlear model with three-dimensional fluid, inner sulcus and feed-forward mechanism," Audiol. Neuro-Otol. **4**, 197–203.
Steele, C. R., and Puria, S. (**2005**). "Force on inner hair cell cilia," Int. J. Solids Struct. **42**, 5887–5904.
Steele, C. R., and Taber, L. A. (**1979**) "Comparison of WKB calculations and experimental results for three-dimensional cochlear models," J. Acoust. Soc. Am. **65**, 1007–1018.
Steele, C. R., and Zais, J. G. (**1985**). "Comparison of WKB calculations and experimental results for three-dimensional cochlear models," J. Acoust. Soc. Am. **77**, 1849–1852.
Thorne, M., Salt, A. N., DeMott, J. E., Henson, M. M., Henson, O. W., Jr., and Gewalt, S. L. (**1999**). "Cochlear fluid space dimensions for six species derived from reconstructions of three-dimensional magnetic resonance images," Laryngoscope **109**, 1661–1668.
Xue, S., Mountain, D. C., and Hubbard, A. E. (**1995**). "Electrically evoked basilar membrane motion," J. Acoust. Soc. Am. **97**, 3030–3041.
Yoon, Y. J., Puria, S., and Steele, C. R. (**2006**). "Intracochlear pressure and organ of Corti impedance from a linear active three-dimensional model," ORL **68**, 365–372.

Loudness growth observed under partially tripolar stimulation: Model and data from cochlear implant listeners

Leonid M. Litvak[a)]
Advanced Bionics Corporation, 12740 San Fernando Road, Sylmar, California, 91342

Anthony J. Spahr
Arizona Biomedical Institute, Arizona State University, Tempe, Arizona 85287

Gulam Emadi
Advanced Bionics Corporation, 12740 San Fernando Road, Sylmar, California, 91342

(Received 5 September 2006; revised 8 May 2007; accepted 18 May 2007)

Most cochlear implant strategies utilize monopolar stimulation, likely inducing relatively broad activation of the auditory neurons. The spread of activity may be narrowed with a tripolar stimulation scheme, wherein compensating current of opposite polarity is simultaneously delivered to two adjacent electrodes. In this study, a model and cochlear implant subjects were used to examine loudness growth for varying amounts of tripolar compensation, parameterized by a coefficient σ, ranging from 0 (monopolar) to 1 (full tripolar). In both the model and the subjects, current required for threshold activation could be approximated by $I(\sigma)=I\mathrm{thr}(0)/(1-\sigma K)$, with fitted constants $I\mathrm{thr}(0)$ and K. Three of the subjects had a "positioner," intended to place their electrode arrays closer to their neural tissue. The values of K were smaller for the positioner users and for a "close" electrode-to-tissue distance in the model. Above threshold, equal-loudness contours for some subjects deviated significantly from a linear scale-up of the threshold approximations. The patterns of deviation were similar to those observed in the model for conditions in which most of the neurons near the center electrode were excited.
 [DOI: 10.1121/1.2749414]

PACS number(s): 43.64.Me, 43.66.Ba, 43.66.Cb [AJO] Pages: 967–981

I. INTRODUCTION

Electrical stimulation of the human cochlea via cochlear implant (CI) systems is typically implemented in a "monopolar" configuration, in which a relatively remote ground electrode provides the return path for the current delivered by the active stimulating electrodes. One consequence of using a remote ground is that the electric field generated in the cochlea is broader than in cases where the return path for stimulation current is positioned closer to the stimulating electrode. Examples of these latter cases include "bipolar" stimulation (wherein an adjacent electrode is used as the ground) and "common-mode" stimulation (wherein all nonstimulating electrodes are shorted together and used as a ground path for the stimulating current). For this paper, these latter stimulus configurations are referred to as "nearby-ground modes."

A number of studies have shown that current spread in the cochlea can be reduced by using nearby-ground modes (Jolly *et al.*, 1996). For example, animal studies indicate that such stimulation modes produce a more restricted region of activation in the auditory nerve (van den Honert and Stypulkowski, 1984), in the inferior colliculus (Rebscher *et al.*, 2001; Snyder *et al.*, 2004), and in the auditory cortex (Raggio and Schreiner, 1999; Bierer and Middlebrooks, 2002). Presumably, reduced current spread and narrower excitation regions would result in decreased channel interaction and improved spectral resolution for CI listeners. In turn, since spectral resolution has been shown to correlate with performance on speech tasks [e.g. (Shannon *et al.*, 1995; Henry and Turner, 2003; Henry *et al.*, 2005)], speech understanding is expected to improve with nearby-ground configurations. However, the speech understanding ability of CI users has been shown to be no better or even worse using these nearby-ground modes as compared to using a monopolar mode of stimulation (Pfingst *et al.*, 2001; Franck *et al.*, 2003). There seems to be a contradiction between the presumed benefits of using nearby-ground configurations (as predicted from animal studies) and the empirically measured performance results for these strategies in human listeners.

Several hypotheses have been proposed to resolve this apparent contradiction. One possibility is that monopolar stimulation is actually more selective in human CI users than is deduced from animal studies. For example, Bruce *et al.* (1999a) proposed that cochlear implants in humans operate at fairly low levels (i.e., in a "tip-of-the-iceberg" sort of scheme), whereas physiological recordings in animals are obtained at higher current levels. Consequently, the physiological animal experiments are not ideal for comparison to the human studies.

A second hypothesis is that the potential narrowing of current spread due to a nearby-ground configuration may be offset by an increase in the amount of current necessary to achieve equivalent loudness (Pfingst *et al.*, 2001). A second-

[a)]Author to whom correspondence should be addressed. Electronic mail: leonidl@advancedBionics.com

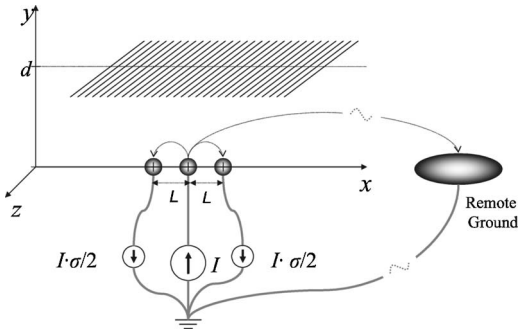


FIG. 1. Model geometry. The electrodes are assumed to be point current sources arrayed along the x axis separated by the inter-electrode spacing L. The bodies of the neuronal response elements are assumed to be located in discrete clusters on a line distance d away from the electrode array. The axons are oriented parallel to the z axis in the plane at $y=d$. A remote ground is assumed to be located infinitely far away from the source. All elements are assumed to be situated in a homogenous electrical medium. Various stimulation configurations are controlled by the compensation coefficient σ.

ary effect of increasing stimulus level with nearby-ground configurations can be the production of significant sidelobes [e.g. (Jolly *et al.*, 1996)]. As a result, the excitation patterns produced by nearby-ground configurations can effectively be broader than with a monopolar configuration.

The present manuscript explores a stimulation mode that is referred to here as "partially tripolar" (Fig. 1). In the partially tripolar stimulation mode, rather than using a nearby ground location (as in bipolar or common-ground mode), a remote ground is used while actively delivering compensation current through adjacent electrodes. The partially tripolar mode is a generalization of the full tripolar mode (Jolly *et al.*, 1996; Mens and Berenstein, 2005), sometimes referred to as the "quadrupolar" mode (three active electrodes plus one ground electrode). Using a simplified model, it will be shown that varying the relative level of the compensation current can allow control over the degree of current spread. With more precise control of current spread, it may be possible to achieve a narrowing of activation while minimizing the influence of sidelobes on the neuronal response. Loudness growth with varying levels of compensation current has been studied in the model and with CI users. It will be shown that the overall current required to achieve threshold varies with the relative level of compensation current in a way that is consistent with linear field summation. For stimuli that evoke a comfortably loud sensation (i.e., supra-threshold stimulation), it is shown that the relationship between the required current and the degree of tripolar compensation can differ from that near threshold and, in some cases, is not consistent with linear field summation. Finally, results from the model are combined with the psychophysical data from the CI users to make general inferences regarding the excitation patterns for various modes of stimulation.

II. METHODS: COMPUTATIONAL MODEL

Models of electrical stimulation of the cochlea, by necessity, involve a trade-off between complexity (in terms of incorporating an accurate representation of current spread and neural responses in the cochlea) and simplicity (in terms of being tractable from a computational standpoint). The approach taken in this paper is to investigate a very simple model which captures many, although not all, of the features of electrical stimulation of the neural tissue. The present model assumes the following: (1) there exists a finite number of discrete neuronal elements spread out over the cochlear space, (2) these elements have a range of thresholds drawn from a log-normal distribution, (3) the electric field at a given spatial location depends only on the amount of current injected and the distance to each injection site (typically an electrode contact), and (4) electric fields from simultaneously activated electrodes interact linearly. The simplified model will be used to make a set of quantitative and qualitative predictions of loudness growth for a range of currents and electrode configurations. These predictions will then be compared to equal loudness contours measured with CI users. For the purpose of making predictions of the loudness contours with the model, it will be assumed that equivalent loudness is achieved across different stimulus configurations by exciting the same total number of neuronal elements. The *selectivity* of a particular stimulation configuration will be assumed to be inversely related to the spatial extent of the neuronal activation.

The model consists of a group of stimulating elements (current sources) and a group of response elements (model neurons). The current sources are arrayed on a line at $y=0$ and spaced at intervals of L units along the x direction (Fig. 1). The neurons are arranged along a line at $y=d$ (i.e., parallel to the current sources) in 200 clusters spaced 0.05 units apart along the x axis. Each cluster contains 100 neurons. The axon of each neuron is assumed to be oriented parallel to the z axis in the plane at $y=d$.

The current sources and neurons are assumed to sit in a homogeneous medium with constant resistivity, and the voltage field generated by each current source is assumed to sum linearly with that generated by every other source. It is assumed that the resistivity of the medium is 4π; note that using a different resistivity value would change only the absolute values of the voltages, and would not affect the relative excitation patterns. The voltage generated at point (x,d,z) in the medium due to current I applied on an electrode at $(L,0,0)$ is given by (Cheng, 1993)

$$V_e = \frac{I}{\sqrt{(x-L)^2 + d^2 + z^2}}. \tag{1}$$

The effective stimulation to the array of neurons can be described by an *electrical activation function* $A(x)$ (Rattay, 1990), which describes the current that flows into the axon of a neuron at position x. Because the axons are oriented parallel to the z axis, the electrical activation function produced by stimulation with current I from an electrode located at $(L,0,0)$ is given by a second spatial derivative of the voltage

$$A_L(x,I) = \frac{\partial^2 V}{\partial z^2} = \frac{-I}{((x-L)^2 + d^2)^{3/2}}. \tag{2}$$

Under partially tripolar compensation, the net electrical activation function is defined by the linear sum of the electrical activation functions created by the center and two ad-

TABLE I. Cochlear implant users. Note that all electrode arrays here are in the general category of "HiFocus" type.

Subject initials	Device (ICS/electrode array type)	Positioner	CNC word score (%)	Center electrode	Pulse width (μs)
DB	CII/HFII	Y	74	6	215
JP	CII/HFII	Y	80	8	107
PG	CII/HFII	Y	92	5	107
SS	CII/HF	N	90	8	215
PH	HR90K/1J	N	76	6	215
VF	HR90K/1J	N	38	8	215
KH	HR90K/1J	N	72	7	215

joining electrodes. In this case, the net electrical activation depends on the amount of tripolar compensation σ

$$A(x,I,\sigma) = A_0(x,I) - A_L\left(x,\frac{\sigma}{2}I\right) - A_{-L}\left(x,\frac{\sigma}{2}I\right). \quad (3)$$

To compute $N(x)$, the number of neurons firing as a function of position x, each of the 100 neurons in a cluster at x is modeled independently. For a neuron j at location x, the firing probability is equal to

$$P(x,j) = \Phi\left[\frac{|A(x,I,\sigma)| - A_{\text{thr}}(x,j)}{A_{\text{thr}}(x,j)\cdot RS(x,j)}\right], \quad (4)$$

where Φ is the cumulative normal distribution function, $A_{\text{thr}}(x,j)$ is the electrical activation required to reach the neuron's threshold, and *RS* is the neuron's *relative spread* [a factor that characterizes the steepness of the function relating stimulus intensity to the neuron's firing probability (Bruce *et al.*, 1999b)]. Use of the absolute value (i.e., no dependence on polarity) of $A(x,I,\sigma)$ in computing firing probability is based on the observation that, for narrow biphasic pulses, the threshold to a cathodic-anodic pulse is similar to that for an anodic-cathodic pulse (Miller *et al.*, 1997).

For the simulations, thresholds $A_{\text{thr}}(x,j)$ were assigned randomly from a log-normal distribution with the ratio of standard deviation to mean set at a value of 0.3 (van den Honert and Stypulkowski, 1987). For computational convenience, the mean of the threshold distribution was arbitrarily set at 0 dB relative to units of $A(x,I,\sigma)$. This arbitrary assignment of the mean was justified as follows. The probability of firing for a given neuronal element depends [see Eq. (4)] on the *ratio* between the net suprathreshold activation $[|A(x,I,\sigma)| - A_{\text{thr}}(x,j)]$ and the threshold $[A_{\text{thr}}(x,j)]$. As a consequence, the choice of the mean threshold will affect the absolute level of neuronal activity for any given current, but not the relative levels of activity for changes in input current. Because it is the relative levels of activity that are of interest in the analyses here, the choice of the mean threshold can be arbitrary. The *RS* of each neuron was chosen from a normal distribution with a mean of 0.0635 and a standard deviation of 0.04; moreover, the distribution was "clipped" in the sense that *RS* was not allowed to go below 0.03 or above 0.10, based on physiological measurements in cat auditory nerve fibers (Miller *et al.*, 1999).

The model was used to compute the spatial distribution of neuronal activity $N(x)$ under various conditions. Note from the model derivation above that $N(x)$ depends implicitly on the electrode-to-tissue distance d, the center electrode current I, the inter-electrode spacing L, and the tripolar compensation σ. Summed neuronal activity $\Sigma N(x)$ (i.e., the total number of neurons firing) was computed for several configurations of electrical stimulation. The effective "perceptual loudness" of the electrical stimulation is assumed to be proportional to $\Sigma N(x)$. The stimulus configurations ranged from monopolar at one extreme (i.e., $\sigma=0$) to full tripolar at the other extreme (i.e., $\sigma=1$). In the monopolar case, only the center electrode was stimulated and the return path was provided by an infinitely remote ground. In the full tripolar case, in addition to having the remote ground, the two adjacent electrodes formed a combined current sink that was equal in magnitude to the current sourced by the center electrode. In order to compare the spatial profiles of the neuronal activity patterns at equivalent perceptual loudness levels under the different stimulus configurations, the model was run to find the center electrode current I that yielded a particular total neuronal activity $\Sigma N(x)$ across configurations.

III. METHODS: PSYCHOPHYSICS

Seven subjects, all users of CII or HR90K implants from Advanced Bionics Corporation, participated in the study (Table I); six of the seven subjects performed very well on tests of speech understanding, with CNC word scores (audio only, in quiet) ranging from 72% to 92%. The CII and HR90K implants were chosen because they have multiple independent current sources, which allow for the delivery of partially tripolar stimulation. Three of the four subjects with the CII implant had a silastic positioner as part of their electrode array. The intended surgical goal of the positioner was to help place the electrodes as close as possible to the neural tissue.

The stimuli utilized in this study were 500-ms biphasic pulse trains (500 Hz, 107 μs or 215 μs per phase; Table I) presented in several partially tripolar configurations. The following values were used for the compensation coefficient σ: 0, 0.25, 0.5, 0.75, 0.875, and 1.

A. "Comfort"-level contours

Equal loudness contours were measured as a function of σ at levels that were "most comfortable." Measurements were taken four times for each value of σ. Data were col-

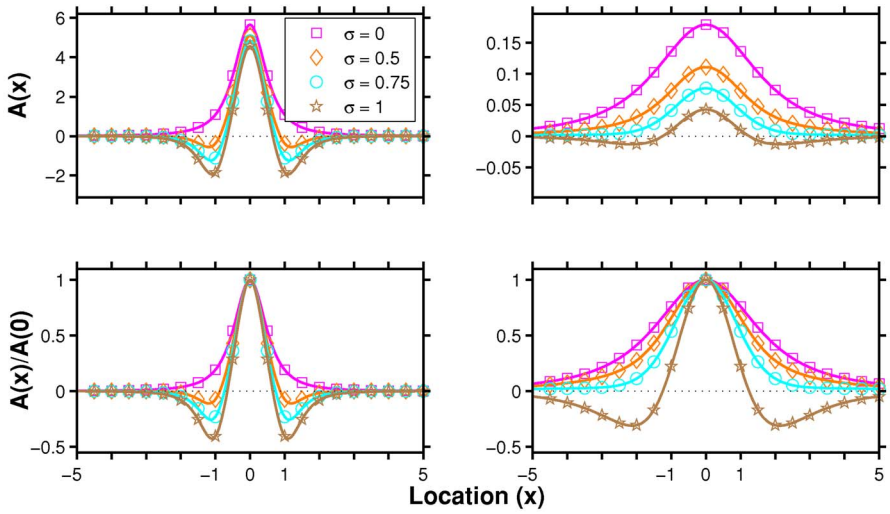


FIG. 2. (Color online) Activation functions during the cathodic phase of the pulse, plotted under assumption that $\rho_e/4\pi=1$. The location of the three electrodes is $L=0$, 1 and −1. The left panel plots the situation where the neurons are relatively close to the electrodes, at $d=0.7$. The right panel shows electrodes relatively far from the target electrodes at $d=2.24$. The bottom panels show the "normalized" activation functions such that the peak of each function is set to 1.

lected for two different inter-electrode spacings (see Fig. 1): (1) compensation current was presented on the electrodes immediately adjacent to the center electrode (inter-electrode spacing $L=1$), and (2) compensation current was presented on electrodes 4 contacts away from the center electrode (inter-electrode spacing $L=4$).

To construct the equal loudness contours, the subject used a scroll wheel to adjust the current I from the center electrode in order to match the loudness of each test stimulus to that of a "standard" stimulus. The standard stimulus was set by increasing the current level for the fully tripolar configuration ($\sigma=1$) until the subject indicated that the most comfortable loudness was achieved. If the loudness was below most comfortable at the maximum current that could be delivered from that electrode (see below), then the next lower σ value was used, and the procedure for finding the standard was repeated. On each test trial, the subject heard the test stimulus continuously alternated with the standard stimulus and was tasked with adjusting the scroll wheel until both stimuli were the same loudness. With few exceptions, the highest values of σ required the highest currents in order to achieve an equivalent loudness (see Fig. 7), and so using the highest attainable σ for the standard stimulus allowed for the greatest available range of current (and corresponding range for loudness adjustments) during the test sequences (at lower σ values); maximizing the overall range available for loudness adjustments was intended to improve the reliability of the data collection. The loudness balancing procedure was repeated four times for each value of σ: two times with the initial test loudness set higher than the loudness of the standard, and two times with the initial test loudness set lower than that of the standard. Loudness balancing was conducted separately for each inter-electrode spacing condition.

Previous studies [e.g. (Mens and Berenstein, 2005)] have found that, for stimulus configurations with large tripolar compensation coefficients, relatively high currents are required to achieve comfortable loudness percepts. The CII and HR90K implants have current sources with a compliance voltage of 7.5 V, and so the maximum current that can be delivered is the ratio of 7.5 V to the electrode impedance. Therefore, for the present experiments, it was particularly important to take additional steps to remain below the maximum current that could actually be delivered by the current sources (without running into compliance issues) for each electrode in each subject. To do so, the impedance was measured for each electrode using a custom software tool [EFIM, (Vanpoucke *et al.*, 2004)], and the testing software automatically ensured that the current level on any stimulus did not exceed the maximum supported by the implant system. In order to maximize the available range of current for testing with a given subject, the center electrode for the partially tripolar stimulation was chosen as the one near the center of the subject's array with the lowest impedance (Table I).

B. Threshold contours

Threshold contours as a function of compensation coefficient σ were measured for two inter-electrode spacings ($L=1$ and $L=4$) with a two-interval forced-choice procedure using a three-down, one-up method (Levitt, 1971). The stimulus timing parameters were the same as for the comfort level measurements. Threshold was computed by averaging the current levels at the last six reversals. For each inter-electrode spacing, two thresholds were acquired for each value of σ. The 99% confidence intervals were estimated by bootstrapping the reversal data, which typically consisted of 12 reversals for the two runs combined (Efron and Tibshirani, 1993). In particular, during every bootstrap iteration, the 12 reversals were re-sampled without replacement, and threshold was computed as the mean of the re-sampled data. The 99% confidence intervals for the threshold data were extracted from 10,000 simulated threshold estimates.

IV. MODEL PREDICTIONS

A. Spatial extent of activation

Figure 2 shows the electrical activation function $A(x,I,\sigma)$ for a "close" electrode-to-tissue distance ($d=0.7$; left panels) and for a "far" electrode-to-tissue distance ($d=2.24$; right panels), each with monopolar stimulation ($\sigma=0$), partially tripolar stimulation ($\sigma=0.5$ and $\sigma=0.75$), and full tripolar stimulation ($\sigma=1$). The inter-electrode spacing L

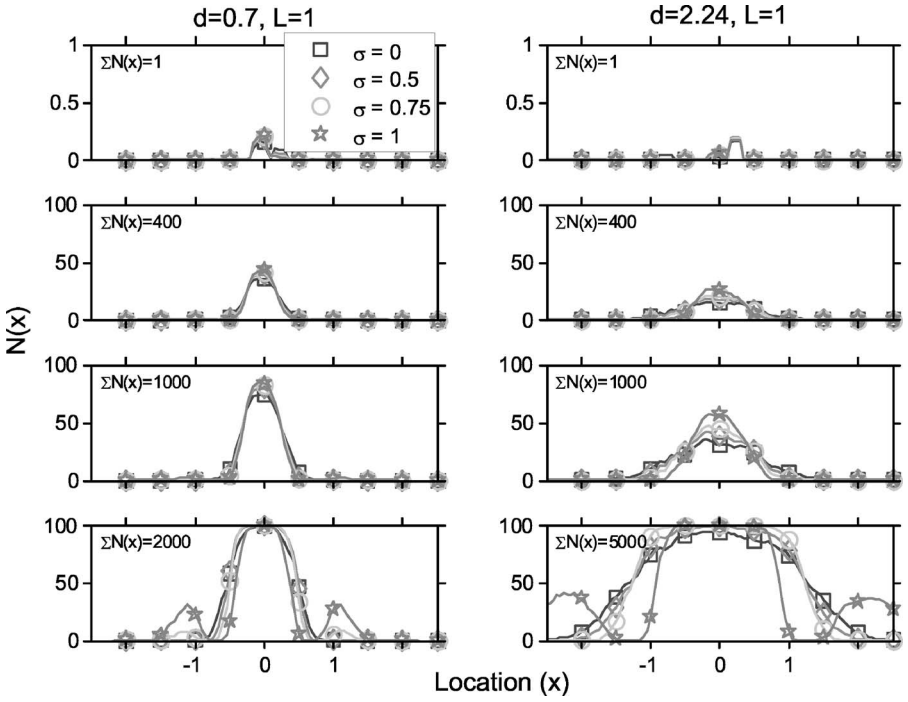


FIG. 3. Neuronal activity patterns for equal total neuronal activity [$\Sigma N(x)$] for inter-electrode spacing $L=1$. Within each set of axes, different symbols indicate different compensation coefficients (ranging from 0 to 1). The left column shows results from the model with a close electrode-to-tissue distance ($d=0.7$), and the right column shows results for a far electrode-to-tissue distance ($d=2.24$). Note the change of scale between the top panel in each column and the remaining three panels. The text in the upper left corner of each panel indicates total neuronal activity [$\Sigma N(x)$].

and current I are set to 1 in all cases. For monopolar stimulation ($\sigma=0$), the electrical activation function for both electrode-to-tissue distances exhibits a single peak aligned with the position of the central electrode ($x=0$). The peak is broader for the far electrode-to-tissue distance as compared to the close distance. Applying tripolar compensation narrows the width of the main peak of the electrical activation function (which becomes especially apparent in the normalized activation functions, shown in the bottom panels) and also decreases its amplitude. Both of these effects of tripolar compensation are more pronounced for the far electrode-to-tissue distance. Sidelobes can appear in the electrical activation functions as tripolar compensation is added. These sidelobes are particularly pronounced for the close electrode-to-tissue distance.

Although the electrical activation function for partially tripolar stimulation can be narrower than the activation function for monopolar stimulation for a given current I from the center electrode, the activation function does not necessarily remain narrower once the current level of the partially tripolar stimulus is increased to achieve an equivalent loudness to the monopolar stimulus. One can approximate "loudness" in the model by making the assumption that loudness is proportional to the total neuronal activity [i.e., $\Sigma N(x)$]. Figures 3

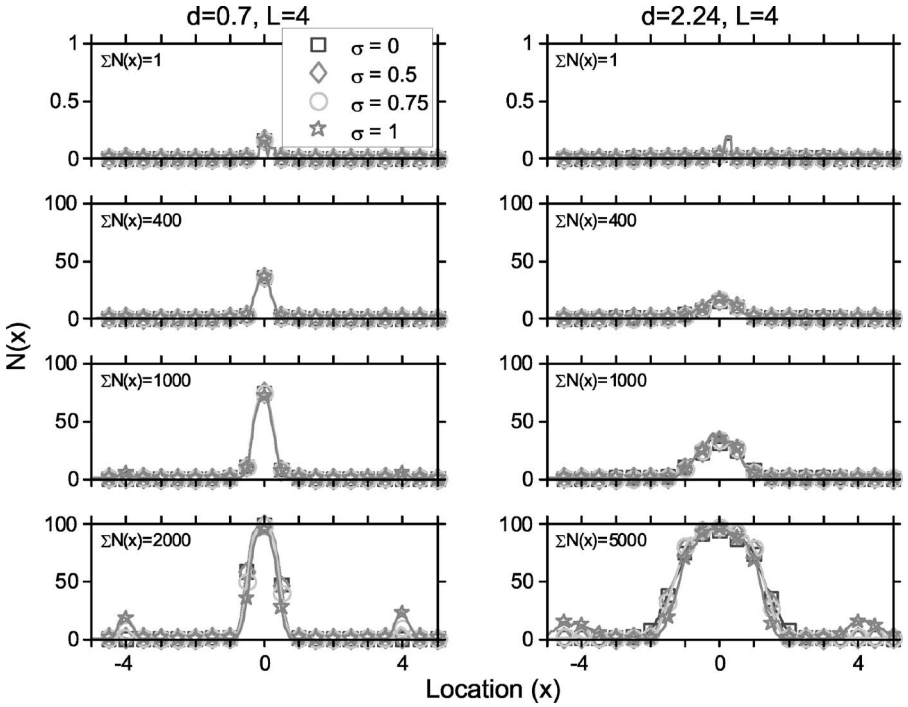


FIG. 4. Same as Fig. 3, but for inter-electrode spacing $L=4$.

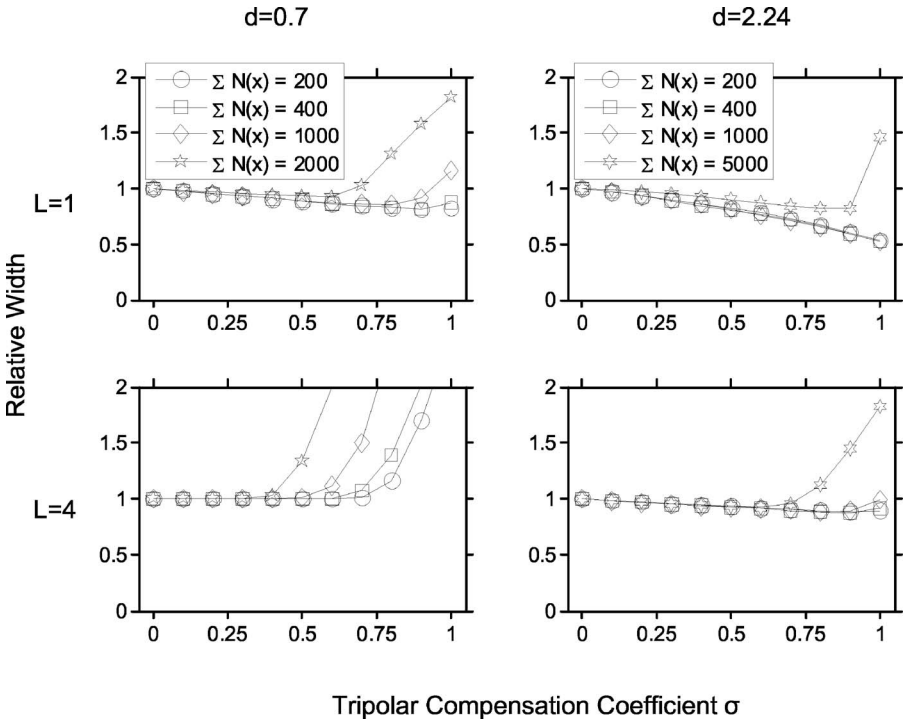


FIG. 5. Metric of relative width of neuronal activity as a function of tripolar compensation coefficient σ. Within each panel, data are shown for four different levels of total neuronal activity [$\Sigma N(x)$]. The left panels in the figure represent a close electrode-to-tissue distance, and the right panels represent a far distance. The upper panels represent a narrow inter-electrode spacing, and the lower panels represent a wide spacing. The different symbols correspond to different levels of total neuronal activity.

and 4 plot the spatial distribution of neuronal activity $N(x)$ for different electrode configurations at different levels of total activity. Figure 3 corresponds to a "narrow" inter-electrode spacing ($L=1$), and Fig. 4 corresponds to a "wide" inter-electrode spacing ($L=4$). The panels in the left column of each of these two figures correspond to a close electrode-to-tissue distance ($d=0.7$), and the panels in the right column of each figure correspond to a far electrode-to-tissue distance ($d=2.24$). The top row of panels shows the spatial distribution for near-threshold neuronal activity ($\Sigma N(x)=1$). The second and third rows show the spatial distribution for cases of "moderate" neuronal activity ($\Sigma N(x)=400$ and $\Sigma N(x)=1000$), defined here as a regime of operation in which none of the neuronal clusters are saturated (i.e., $N(x)<100$ for all x). The bottom panels in each figure show the spatial distribution of activity for stimulus levels that have just begun to induce "saturation" in the monopolar configuration. Saturation is defined here as a regime of operation in which at least one of the neuronal clusters (typically, in the immediate vicinity of the center electrode) is saturated (i.e., where $N(x)=100$ for some x). Note that the total activity at which saturation begins is greater for the far electrode-to-tissue distance ($\Sigma N(x)=5000$) than for the close electrode-to-tissue distance ($\Sigma N(x)=2000$). The reason for this difference is that the relatively broader peak region for the far condition in the subsaturation regime necessitates an overall higher level of total activity to reach saturation in the first place as compared to the close condition.

For near-threshold activity, nearly all of the activity is concentrated in the region near the center electrode ($x=0$) for all electrode configurations. For moderate activity, all neuronal activity is localized to a single-peaked region centered near the stimulating electrodes. The width of the active region decreases as tripolar compensation is increased. The decrease in width with increasing compensation is especially apparent for the far electrode-to-tissue distance combined with the narrow inter-electrode spacing (Fig. 3, right panels). In general, the model predicts that, for near-threshold and moderate levels of activity, tripolar compensation (under conditions that evoke equivalent loudness percepts) narrows the spatial extent of neuronal activity as compared to monopolar stimulation.

A different behavior occurs for levels of activity that include saturation. In this condition, further increases in total activity can be achieved only by activating neurons in clusters beyond the regions that are already saturated. Consequently, a narrowing in the main lobe of the electrical activation function due to tripolar compensation is offset by the need to recruit neurons further away from the center electrode. Saturation near the center of the stimulation region is not the only phenomenon that can result in a wider than expected spatial spread of neuronal activity. The presence of sidelobes in the electrical activation functions for nonmonopolar stimuli also can result in an effectively wider region of neuronal activity, especially for higher stimulation levels. As shown in Figs. 3 and 4 (bottom panels), the neuronal activity shows a single peak for monopolar stimulation ($\sigma=0$) but exhibits three distinct peaks for full tripolar compensation ($\sigma=1$). The multi-peaked neuronal activity is due to the prominent sidelobes that appear in the electrical activation functions $A(x,I,\sigma)$, as shown in Fig. 2. Note that, based on the underlying activation functions, multi-peaked neuronal activity could, in theory, occur for total activity levels below those including any saturation. For the model parameters used here, however, the neuronal activity pattern was primarily single peaked for moderate (i.e., subsaturation) activity levels. For the close electrode-to-tissue distance ($d=0.7$), there was some noticeable contribution from the sidelobes, even at moderate activity levels. This contribution will become more apparent in the discussion of Fig. 5 (below).

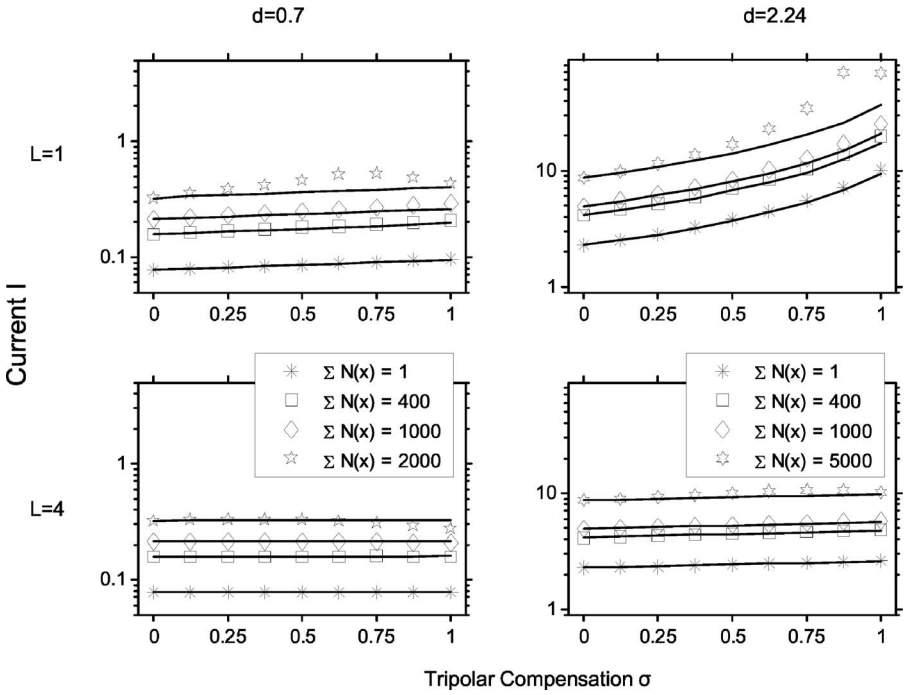


FIG. 6. Predicted current levels needed to achieve equal loudness contours for near-threshold and suprathreshold conditions. The panel layout is the same as for Fig. 5. The total neuronal activity levels $[\Sigma N(x)]$ are the same as those in Figs. 3 and 4. In each panel, the points indicate current levels derived from the model simulations, and the lines through the points indicate the equal loudness contours predicted from the constant peak approximation for the given set of stimulus conditions. The near-threshold contour in each panel is the one for which $\Sigma N(x)=1$. The remaining contours in each panel represent suprathreshold stimulation levels (i.e., $\Sigma N(x)>1$).

Plotted in Fig. 5 is a metric of "relative width" of neuronal activity as a function of tripolar compensation σ. Each panel represents a different model geometry (encompassing the electrode-to-tissue distances and inter-electrode spacings represented in Figs. 3 and 4). Within each panel are shown four equal-loudness contours, including a condition that includes saturation. The width of excitation was quantified using the following equation:

$$W = \sqrt{\frac{\sum_x (x^2 \cdot N(x))}{\sum_x N(x)}}. \tag{5}$$

To focus specifically on the effects of increasing tripolar compensation, the widths for a given contour were normalized to the monopolar configuration ($\sigma=0$) to yield "relative width."

It becomes apparent from Fig. 5 that initial increases of tripolar compensation (from $\sigma=0$) result in a decrease in the spatial extent of neuronal activity for all model configurations. The decrease is greatest for the narrow inter-electrode spacing ($L=1$) in conjunction with the far electrode-to-tissue distance ($d=2.24$) (upper right panel). For the narrow inter-electrode spacing (upper panels), the activity level that includes saturation (starred points) exhibits the smallest decrease (as compared to the other activity levels) in relative width as the tripolar compensation is increased. Thus, even before there is a contribution from sidelobes (see below), the benefit of tripolar compensation with regards to narrowing the spatial extent of neuronal activity can be limited for an overall activity level that includes some saturation.

As tripolar compensation is further increased, at some point the relative width of neuronal activity begins to increase again. This increase in width is consistent with the introduction of sidelobes into the neuronal activity distributions. Consistent with the distributions shown in Figs. 3 and 4, the effect of sidelobes on the relative width is most pronounced for higher activity levels and for the wider inter-electrode spacing (in particular, see the bottom panels of Fig. 4).

In summary, the application of tripolar compensation can theoretically decrease the relative width of neuronal activity while maintaining an equivalent loudness percept. The benefit tends to be greatest under the following conditions: (1) relatively narrow inter-electrode spacing, (2) relatively far electrode-to-tissue distance, and (3) levels of neuronal activity that do not entail any saturation. In general, the amount of tripolar compensation that minimizes the relative width of activity may be less than that which results in full tripolar stimulation (i.e., σ may be less than 1).

B. Current required to maintain equal loudness

Figure 6 shows the center electrode current I required to maintain constant neuronal activity (i.e., equal loudness) as a function of tripolar compensation σ for various model geometries (same panel layout as for Fig. 5). Within each panel are shown equal-loudness contours for four different overall activity levels: (1) near-threshold activity (squares), (2) moderate activity (circles and diamonds), and (3) an activity level that includes saturation (stars). In general, for any given model configuration and value of tripolar compensation, an increase in overall loudness requires an increase in current, as expected. For all subsaturation activity levels, the current required to maintain equal loudness increases monotonically with tripolar compensation. For activity levels that include saturation, the change in current required with increasing tripolar compensation is typically nonmonotonic. These observations are discussed below in greater detail.

C. Near-threshold behavior: "Constant peak approximation"

For near-threshold activity (squares in Fig. 6), the model simulations predict that, in all cases, I increases monotonically with increasing σ. This result makes sense in light of the fact that, near threshold, the response is composed almost entirely of neurons located in the single cluster aligned with the center electrode. Because tripolar compensation causes the peak of the electrical activation function $A(x,I,\sigma)$ to decrease in amplitude (see Fig. 2, upper panels), I must be increased to maintain a constant loudness percept [i.e., a constant $\Sigma N(x)$]. The amount by which I needs to be increased closely parallels the decrease in the magnitude of the peak of the electrical activation function. This amount can be determined mathematically from Eq. (3). Since $I_{\text{thr}}(0)$ is the current necessary to reach threshold level neuronal activity for the monopolar configuration ($\sigma=0$), the electrical activation at the peak location ($x=0$) is equal to $A(0,I_{\text{thr}}(0),0)$. For a configuration with a nonzero tripolar compensation coefficient σ, the current $I_{\text{thr}}(\sigma)$ necessary to achieve threshold level activity must satisfy

$$A(0,I_{\text{thr}}(\sigma),\sigma)=A(0,I_{\text{thr}}(0),0). \tag{6}$$

The right side of Eq. (3) can be substituted into the left and right sides of Eq. (6). Solving for $I_{\text{thr}}(\sigma)$ yields the following:

$$I_{\text{thr}}(\sigma)=\frac{I_{\text{thr}}(0)}{1-K\cdot\sigma}. \tag{7}$$

In Eq. (7), which shall be referred to as the *constant peak approximation*, $I_{\text{thr}}(0)$ is simply the current required to achieve threshold activity in the monopolar condition (i.e., $\sigma=0$), and K is an "interaction coefficient" that incorporates the contribution from the electric fields generated by the side electrodes to the peak of the electrical activation function [i.e., to $A(0,I,\sigma)$] when nonzero tripolar compensation is applied

$$K=\frac{A_L(0,I)+A_{-L}(0,I)}{2\cdot A_0(0,I)}. \tag{8}$$

Note that this derivation for K requires only an assumption of linearity in electric field summation and is independent of any specific model geometry. For the particular model geometry described in Fig. 1, it is possible to solve for K as a function of the electrode-to-tissue distance d and the interelectrode spacing L. Using Eq. (2), K can be reduced to the following:

$$K(d,L)=\left(\frac{d^2}{d^2+L^2}\right)^{3/2}. \tag{9}$$

The accuracy of the constant peak approximation in predicting the current required for near-threshold activity ($\Sigma N(x)=1$) is shown in Fig. 6. The squares in each panel show the results of the model simulations for near-threshold neuronal activity, and the solid curves through these points have been computed using the constant peak approximation [Eq. (7)].

Note from Eq. (9) that, as the electrode-to-tissue distance d increases from 0 to ∞, K goes from 0 to 1. Placed into the context of Eq. (7), a value of $K=0$ yields no dependence of $I_{\text{thr}}(\sigma)$ on σ, and a value of $K=1$ yields the maximum possible dependence of $I_{\text{thr}}(\sigma)$ on σ. This behavior can be understood in the context of a "far field" limit as $d\rightarrow\infty$. Namely, as the electrodes are moved further and further away from the tissue, they begin to "merge" (from the perspective of the neurons) to a single point location containing one current source (the center electrode) and two current sinks (the side electrodes). For the neuronal cluster at position $x=0$, this effective merging of the source and the sinks results in a decrease of electrical activation with increasing tripolar compensation that goes beyond that expected solely from increasing the distance from a monopolar source. In practical terms, it is predicted that, with nonzero tripolar compensation, implant recipients with relatively larger electrode-to-tissue distances (see right panels of Fig. 6) will have larger interaction coefficients and subsequently greater increases in threshold as σ is increased.

D. Supra-threshold behavior: Deviations from constant peak approximation

As demonstrated above, the constant peak approximation is expected to provide a fairly accurate prediction of the current required to maintain near-threshold activity in the model as the tripolar compensation is increased. A linearly scaled up version of this approximation turns out to be less accurate in predicting equal-loudness contours for supra-threshold levels, particularly for levels that include saturation (stars in Fig. 6). At higher levels in general, neuronal clusters in addition to the one located at $x=0$ will be active (in particular for the monopolar condition), and so the detailed spatial profile of the electrical activation function, rather than simply the amplitude at its peak, will contribute to the net neuronal activity. As the tripolar compensation is increased, the relative width of the activation function decreases (Fig. 2, bottom panels), fewer neuronal clusters contribute to the response, and a greater increase in current (than is expected from a simple linear scaling up of the constant peak approximation) is needed to maintain a constant given level of activity.

The deviations from the constant peak approximation become even more apparent as the level of total neuronal activity is increased from moderate to a level that includes saturation (stars in all panels). In the saturation regime, only a few neurons in the neuronal clusters near the center electrode are available for further excitation. As tripolar compensation is increased, the amplitude of the electrical activation function drops more quickly in regions away from the centrally located neuronal clusters than it does near the central clusters (see Fig. 2). In order to maintain a constant overall neuronal activity level, the current has to be increased so that the neuronal activity in these side regions does not change. The required increase at the sides results in a net increase at the central peak of the electrical activation function without a concomitant increase of neuronal activity at the peak location. The consequence is a deviation from the constant peak approximation.

The initial increase of current required (above and be-

TABLE II. Qualitative predictions of the model.

Config. number	Model geometry: *Electrode-to-tissue distance*	Model geometry: *Inter-electrode spacing*	Threshold contours: $I_{thr}(0)$	Threshold contours: K	Comfort contours: Contour shape relative to constant peak approx., small σ	Comfort contours: Contour shape relative to constant peak approx. large σ
1	d=2.24 (far)	L=1 (narrow)	Higher than configs. 3 and 4	*Largest of all four configs.*	*Similar*	*Similar*
2	d=2.24 (far)	L=4 (wide)	Higher than configs. 3 and 4	*Smaller than config. 1*	*Similar*	*Similar*
3	d=0.7 (near)	L=1 (narrow)	Lower than configs. 1 and 2	*Smaller than config. 1*	*More steep*	*Similar/less steep*
4	d=0.7 (near)	L=4 (wide)	Lower than configs. 1 and 2	*Smallest of all config.*	*More steep*	*Similar/less steep*

yond that predicted from the constant peak approximation) to maintain supra-threshold activity levels with increasing tripolar compensation σ can be followed by a drop in the required current for the largest values of σ. At these levels of compensation, the sidelobes in the electrical activation functions become large enough to elicit responses from neuronal elements located in clusters that are positioned away from the center electrode region (see lower panels of Figs. 3 and 4). Consequently, the current required to maintain an equivalent loudness begins to drop. This behavior is most pronounced for neuronal activity levels that include some saturation.

E. Effects of electrode-to-tissue distance and inter-electrode spacing

Comparing the left and right panels of Fig. 6 allows for a qualitative analysis of the effect of electrode-to-tissue distance. The left panels show predicted equal loudness contours for a "close" electrode-to-tissue distance (d=0.7), and the right panels show contours for a "far" distance (d=2.24) The most obvious difference between the results for the two distances is that, for the far condition, the equal-loudness contours are generally steeper as a function of tripolar compensation. In effect, there is a greater interaction between the center and side electrodes for the far condition. This result is consistent with the previous demonstration that the value of the interaction coefficient K (used in the constant peak approximation) increases with increasing electrode-to-tissue distance d.

Similarly, a comparison between the top and bottom panels of Fig. 6 allows for an analysis of the effect of inter-electrode spacing. The top panels show equal-loudness contours for a "narrow" inter-electrode spacing (L=1), and the bottom panels show the contours for a "wide" spacing (L=4). It can be seen that the narrow spacing yields steeper contours than the wide spacing. The narrower inter-electrode spacing effectively results in greater interaction between the center and side electrodes, as would be expected.

Table II provides a general summary of the predictions of the model under the various geometrical configurations. The least amount of interaction between the center and side electrodes is expected for the close electrode-to-tissue distance and wide inter-electrode spacing (bottom left panel of Fig. 6). In this case, the center and side electrodes can elicit responses from completely independent clusters of neurons (see bottom left panel of Fig. 4). As a consequence, for the largest tripolar compensations in conjunction with overall neuronal activity levels that include the contribution of sidelobes, the current required to maintain constant loudness can actually drop below the current required in the monopolar condition (see points for $\Sigma N(x)$=2000 in bottom left panel of Fig. 6).

V. PSYCHOPHYSICAL RESULTS

Figure 7 shows threshold and comfort-level equal-loudness contours as a function of tripolar compensation for the seven cochlear implant subjects in this study. The top panels show data for a narrow inter-electrode spacing (L=1), and the bottom panels show data for a wide inter-electrode spacing (L=4). There were two sets of gaps in the data: (1) for subject JP, time constraints allowed data collection only for the narrow spacing, and (2) for subject KH, threshold-level data were not acquired for tripolar compensation coefficients of 0.875 and 1. Upward-pointing triangles indicate cases where threshold or comfort levels could not be reached without exceeding the maximum current supported by the implant (see Methods). Because two different pulse durations were used across the subject population (see Table I), the data are shown in units of charge delivered per phase rather than in units of current, for ease of comparison.

A. Threshold contours

The threshold data (squares) for each subject were fit with the constant peak approximation [Eq. (7)] to yield values for the interaction coefficient K and the monopolar current $I_{thr}(0)$. The fits (lower solid curve in each panel) show that the biophysical model does, in fact, provide a good approximation of threshold current as a function of tripolar compensation for both the narrow and wide inter-electrode spacings. The difference between the model fit and the raw subject data exceeded the 99% confidence intervals of the fit in only 1 out of 76 measurements (subject PH, L=1, σ=0). Recalling that the fundamental basis for the constant peak

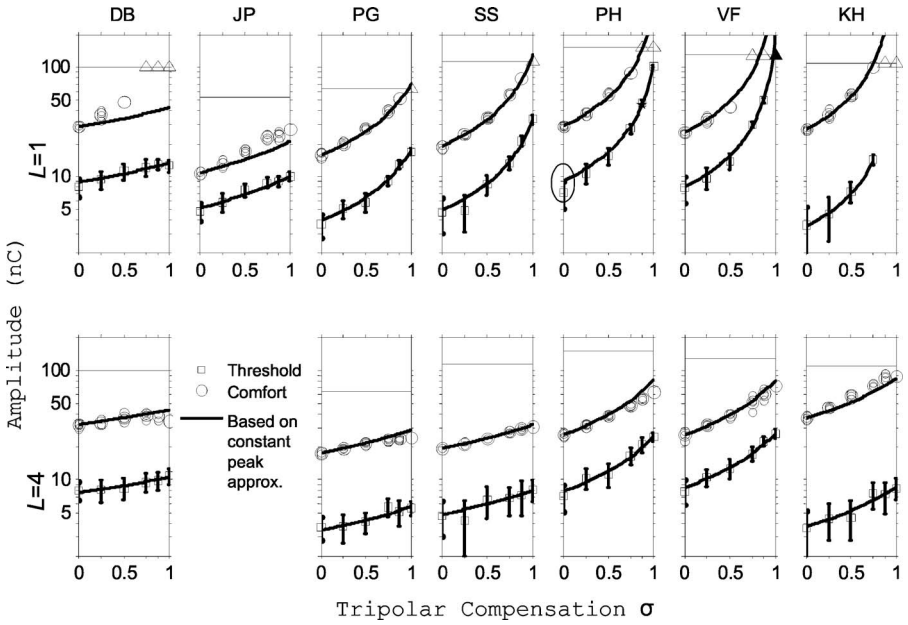


FIG. 7. Summary of the loudness balancing and threshold data for each subject. The top panel shows data for spacing of 1 electrode, and bottom spacing for 4 electrodes. The squares are the thresholds while the circles are measurements at comfort level. The horizontal lines indicate absolute maximum levels that could be achieved by the implant. The triangles indicate tripolar compensation coefficients σ for which comfort (empty triangle) or threshold (filled triangle) exceeded maximum current. The solid black lines represent constant peak approximation functions and will be discussed in the text.

approximation is that all of the neuronal activity is clustered at the location of the center electrode, the good fit between the subject data near threshold and the constant peak approximation supports the idea that threshold responses for the CI subjects are governed primarily by neuronal activity in a relatively narrow region, presumably near the center stimulation electrode.

The left panel of Fig. 8 shows the estimates of the interaction coefficient K for each implant user. In general, for a given user, the interaction coefficient is greater for the narrow inter-electrode spacing ($L=1$) than for the wide spacing ($L=4$), as predicted by the model [see Table II and Eq. (9)]. Also consistent with the model is the fact that the subjects with a positioner, who presumably have a closer electrode-to-tissue distance than subjects without a positioner, exhibit the smallest interaction coefficients for a given inter-electrode spacing. Within the users with a positioner, subject PG had the largest interaction coefficient for a given inter-electrode spacing. This finding might be explained by the fact that the center electrode chosen for this particular subject (electrode 5) happened to be the most apical of the subject group (Table I). Since the apical end of the positioner ends around electrode 5, the measurements taken with this particular subject may have been more akin to the measurements taken with the nonpositioner subjects, in terms of the electrode-to-tissue distance.

The right panel of Fig. 8 shows the monopolar threshold for each subject; note that inter-electrode spacing is irrelevant in this panel because only the center electrode is stimulated. While the model predicts that the presence of a positioner (i.e., close electrode-to-tissue distance) should correlate with a lower monopolar threshold, the empirical results in Fig. 8 do not show such a correlation. The fact that monopolar thresholds do not correlate with the interaction coefficients in the subjects can be attributed to possible differences in the spatial extent of neural survival.

B. Comfort contours

To characterize the extent to which the constant peak approximation could account for supra-threshold measurements, the comfort level data were approximated with a

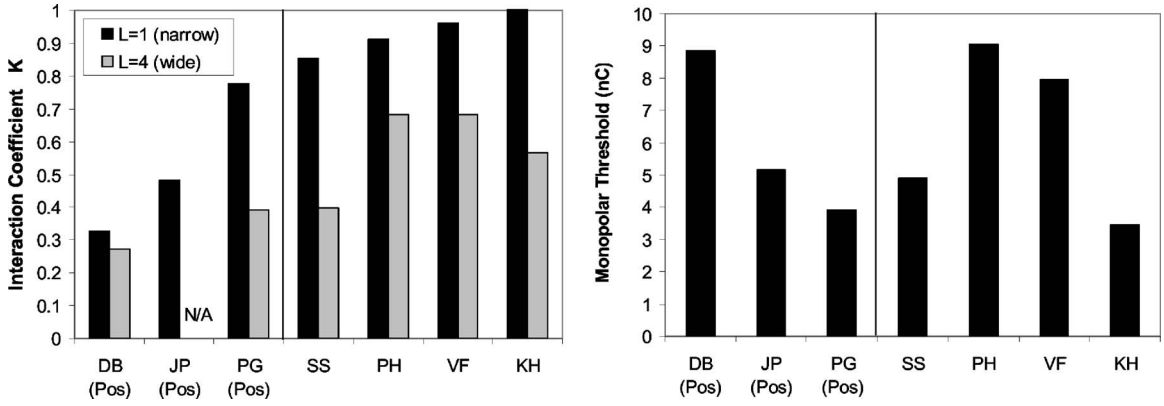


FIG. 8. The left panel shows the relationship between interaction coefficient and electrode type for spacing of 1 and 4 electrodes. The three subjects on the left of the vertical line have a positioner in their array, while the data on the right are for nonpositioner subjects. The right panel shows the monopolar threshold in nC in the same format.

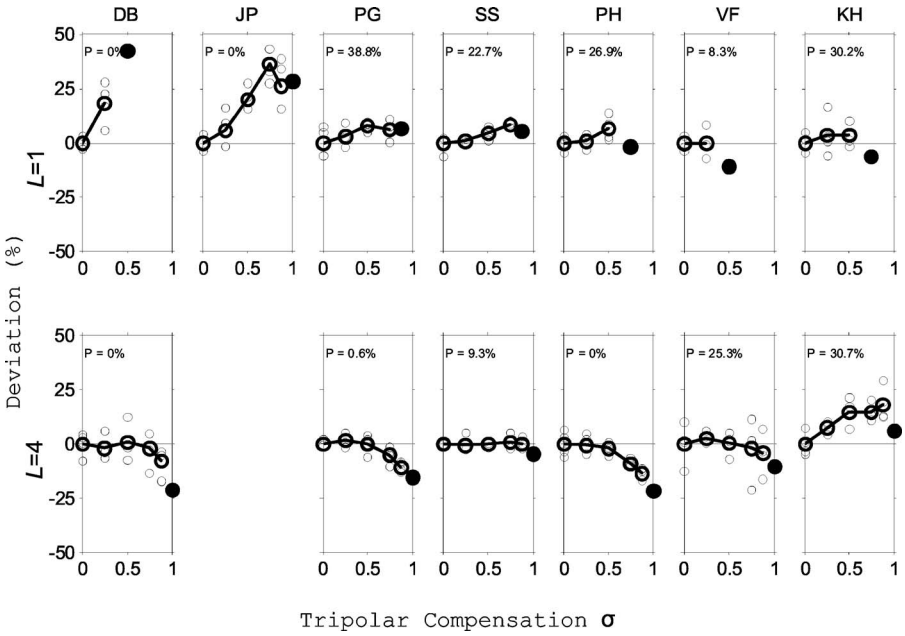


FIG. 9. Differences between equal-loudness data for comfortable loudness and the scaled up constant peak approximation. The empty circles show the differences for the individual balancing results (smaller circles) and the mean differences (larger circles). The filled circle in each panel indicates the standard stimulus, against which the test stimuli were balanced. Note that, because of the scaling procedure, the mean difference is by definition zero for $\sigma=0$. The P value shown in each panel is the probability that the deviations from the constant peak approximation are within the expected confidence limits. Low P values correspond to significant deviations.

scaled up version of the fit to the threshold data. Namely, the monopolar current required to achieve comfort was substituted for $I_{\text{thr}}(0)$ in Eq. (7), and the value for the interaction coefficient K was taken directly from the fit to the threshold data; the direct substitution of K was justified by the fact that the interaction coefficient should be independent of activity level. The upper solid line in each panel of Fig. 7 shows the constant peak approximation scaled to the comfort level data. Figure 9 shows the difference (in percent) between the comfort level data and the constant peak approximation. The significance of the difference between the psychophysical data and the constant peak approximation was computed by a bootstrap technique (Efron and Tibshirani, 1993). In particular, for each subject, simulated threshold data were generated for each tripolar compensation by drawing from a normal distribution with a mean equal to the measured threshold and a variance estimated from the reversal data of the threshold-tracking procedure. Similarly, simulated comfort-level data were generated by drawing from a normal distribution with a mean equal to the measured comfort data and a variance estimated from the multiple balancing trials. The simulated threshold data were then fit with a constant peak approximation, the approximation was scaled up to the level of the simulated comfort-level data, and the difference between the approximation and the simulated data was computed. One thousand runs of the bootstrap simulation allowed for an estimate of confidence intervals for the differences shown in Fig. 9, given the variability in the original real data measurements.

For the narrow inter-electrode spacing ($L=1$), there were no significant deviations (at a 5% criterion) between the constant peak approximation and the measured data for five subjects (PG, SS, PH, KH, and VF). Based on the results from Fig. 6, these data are consistent with low to moderate levels of neuronal activity, with no regions of saturation. Statistically significant deviations from the constant peak approximation were observed for two of the subjects (DB and JP) for the narrow inter-electrode spacing. Of interest is that both of these subjects had a positioner with their electrode array. The pattern of the deviations from the constant peak approximation is similar to that observed in the model for a small electrode-to-tissue distance and "a total neuronal activity level that includes saturation" (see Fig. 6, upper left panel, stars). Moreover, for subject JP, a decrease in the separation between the data and the constant peak approximation was observed between $\sigma=0.875$ and $\sigma=1$, suggesting the possible influence of sidelobes at the largest tripolar compensation level. For subject DB, measurements could not be taken at the larger tripolar compensations because of compliance limits of the current sources.

For the wide inter-electrode spacing ($L=4$), significant deviations between the data and the constant peak approximation were observed for three of the six subjects (PG, PH, and DB). For subject PG (one of the positioner patients) and subject PH, the pattern of deviations was similar to that observed for the close electrode-to-tissue distance in conjunction with neuronal activity levels that included saturation (see Fig. 6, lower left panel, stars), in that the current required to achieve comfort for the higher compensations was actually lower than that predicted by the constant peak approximation. It was suggested earlier that, under these conditions, the electrical activation includes sidelobes that stimulate clusters of neurons that are independent of those at the position of the central electrode. For the other three subjects (SS, KH, and VF), the data were similar to the "moderate" neuronal activity condition, in that deviations from the constant peak approximation were statistically negligible; note that, although visual inspection of the data for subject KH appear to show a significant deviation from the constant peak approximation, the variances in the individual threshold measurements for this subject were very large.

Finally, some of the subjects reported that the pitch evoked by the stimuli changed as a function of tripolar compensation. In principle, it is possible that such pitch changes could have affected the loudness comparisons across different conditions. To assess whether changes in pitch were a

TABLE III. Predicted electrode-to-tissue distances for linear and circular model geometries. The predicted distances are in units of millimeters.

	Interaction coefficient K		Predicted distance, linear model		Predicted distance, circular model	
Subject	$L=1$	$L=4$	$L=1$	$L=4$	$L=1$	$L=4$
DB	0.33	0.27	0.95	3.4	0.77	2.09
JP	0.48		1.27		0.98	
PG	0.78	0.39	2.33	4.29	1.61	2.51
SS	0.85	0.4	3.01	4.34	1.97	2.54
PH	0.91	0.68	3.97	7.42	2.44	3.84
VF	0.96	0.68	6.15	7.43	3.4	3.85
KH	0.99	0.57	8.59	5.9	4.35	3.23

factor in the present study, subjects were asked to indicate the direction as well as the degree of any pitch changes. The pitch changes were generally small for the narrow inter-electrode spacing ($L=1$). For the wide inter-electrode spacing ($L=4$), some subjects reported strong pitch changes, with some subjects reporting that the tripolar stimulus was consistently higher in pitch than the monopolar stimulus and others reporting that the tripolar stimulus was lower in pitch. While it is acknowledged that the effects of pitch changes on loudness judgments were not directly controlled in the present experiments, the data suggest that there was no systematic effect of pitch. Generally speaking, the direction and magnitude of pitch changes did not correlate with the patterns of deviation from the constant peak approximation. For example, some of the largest deviations between the comfort level data and the scaled up constant peak approximation were observed for subject DB (see Fig. 9), who reported that all stimuli evoked the same pitch.

In summary, the comfort-level data from the CI subjects are consistent with the qualitative predictions of the model, as summarized in Table II.

VI. DISCUSSION

A. Model assumptions

The present manuscript considers a simplified biophysical model that allows for relatively straightforward predictions in terms of current spread and the spatial extent of neuronal activity. It was shown that the model is qualitatively consistent with threshold and comfort-level equal-loudness contours measured with CI subjects. It is of interest to know to what extent the simplified model quantitatively agrees with the observed data and whether specific predictions of the model are consistent with the known geometry and size of the human cochlea. The model can be used to predict the distance between the central electrode and the neuronal tissue for a given CI subject. This distance can be estimated from the fitted interaction coefficient K by solving Eq. (8) for the electrode-to-tissue distance d and plugging in the known value for the inter-electrode spacing L; the physical spacing between the individual electrode contacts was 1 mm for the electrode arrays for all the subjects used in this study. The results of the analysis are tabulated in Table III (left and middle sets of columns).

Examination of the table (see columns for "Predicted distance, linear model") shows that, for some subjects, the model (as laid out in Fig. 1) predicts electrode-to-tissue distances that are too large compared to the known dimensions of the human cochlea. For example, the predicted distance is greater than 7 mm for subjects VF, PH, and KH, whereas the distance between the furthest point on the scala tympani and the neural tissue at the base of the human cochlea is below 4 mm. In addition, the model is not consistent in that different electrode-to-tissue distances are predicted for the two inter-electrode spacings ($r=0.65$). To examine whether predictions of the electrode-to-tissue distance would be improved if more realistic model geometries were considered, the model geometry was changed such that the cell bodies of the neurons were arranged in a circle of 3 mm diameter [the approximate diameter of the human modiolus (Snyder, personal communication)]. In turn, the electrode contacts were arrayed on a concentric circle with a diameter of $(3+d)$ mm. Simulations of the model under this circular geometry were run in order to determine values for d which best fit the observed interaction coefficient K computed from Eq. (7). The predicted distances for the circular model are shown in the last two columns of Table III. These distances are smaller than those derived from the linearly arrayed model and are more realistic with respect to an electrode array sitting in the human cochlea. In addition, the predictions are more self-consistent between the two inter-electrode spacings ($r=0.71$). It is generally expected that inclusion of more realistic assumptions (e.g., incorporating the nonhomogenous tissue in order to account for the bony wall of the modiolus) will lead to further improvements in the ability of the model to predict the electrode-to-tissue distance.

The model also assumes that the spatial distribution of the spiral ganglion neurons in humans is similar to that reported for normal-hearing cats (van den Honert and Stypulkowski, 1987). The assumed distribution most directly affects the predicted dynamic ranges in the model, and to a lesser extent the current levels at which "saturation" and "sidelobe" phenomena are observed. Because human spiral ganglion counts vary drastically across individuals with severe hearing loss (Nadol, 1997; Fayad and Linthicum, 2006) and generally differ from counts in animal models of hearing loss, it is unlikely that the biophysical model used in this study will be able to provide an accurate quantitative predic-

tion of the dynamic ranges observed in the CI listeners. For example, the present model predicts that, with a narrow inter-electrode spacing ($L=1$) and a value for the interaction coefficient K derived from the threshold data, the constant-peak approximation should hold over a range of no more than approximately 11 dB for the close electrode-to-tissue distance ($d=0.7$) and over a range of no more than 7 dB for the far electrode-to-tissue distance ($d=2.24$). In general, the model predicts that, with increasing d, the dynamic range over which the constant peak approximation holds will decrease. The actual CI subject data show that the constant peak approximation holds over a larger than expected dynamic range for a given estimated electrode-to-tissue distance. For example, the approximation holds over a dynamic range of at least 10 dB for nonpositioner subject VF ($d=6.15$ mm) and at least 17 dB for nonpositioner subject KH ($d=8.59$ mm). This greater than expected dynamic range suggests that the spatial distribution of neurons assumed in the model is inconsistent with that for the CI subjects. In addition, differences in the dynamic range observed across the CI subjects in this study may be rooted in differences in neural survival across the subjects. Because present visualization techniques are not capable of estimating spiral ganglion cell counts in living CI recipients, it is generally not possible to extend the biophysical model to incorporate more realistic spatial distributions of the neuronal elements for individual subjects.

B. Loudness model

The present model assumes that constant loudness corresponds to equal total neuronal activity. However, total neuronal activity has been suggested to be an incomplete model of loudness for the normal ear (Relkin and Doucet, 1997). McKay *et al.* have suggested that a more accurate model for loudness may be one in which the neural activity at each point is first transformed by a specific loudness power function with an exponent greater than 1 (McKay *et al.*, 2001; McKay *et al.*, 2003). Because such a function would emphasize the contribution of peaks in the neuronal activity pattern to the effective loudness, incorporation of the theory of McKay *et al.* into the present model would predict a smaller contribution of the sidelobes to the loudness and, except at levels that include saturation, would predict smaller deviations from the constant peak approximation. In general, extending the present model to include the power function suggested by McKay *et al.* would increase differences between the model predictions and the CI subject data, especially in cases where the data suggest a significant influence of sidelobes. This influence seems to be pronounced for the wide inter-electrode spacing ($L=4$), where the equal-loudness contours drop more quickly than predicted by the constant peak approximation (see bottom panels in Fig. 9).

C. Absolute thresholds

A notable exception to the agreement between the model and the subject data was that the model predicts a lower absolute threshold for close electrode-to-tissue distances, whereas the absolute thresholds to monopolar stimulation were not consistently lower for the positioner subjects (see Fig. 8, right panel). An implicit assumption here is that subjects with a positioner have a closer electrode-to-tissue distance, but independent data (e.g., x-ray images) were not available to confirm this assumption for the subjects in this study. Moreover, because absolute threshold is ultimately determined by the most sensitive of the remaining neurons, subtle differences in the neuronal distributions across CI recipients can confound cross-subject comparisons of absolute thresholds.

D. Variability across the electrode array

The present study quantified both the interaction coefficient and the absolute thresholds near the middle of the electrode array. It is unclear, however, whether both measurements extend over the entire cochlea. While the monopolar thresholds are reported to be somewhat constant across the array, thresholds for more near-ground configurations are more variable than those for monopolar (Pfingst and Xu, 2004). Since the interaction coefficient is determined by the ratio between the monopolar and the tripolar thresholds, it is likely that the interaction coefficient also varies across the cochlea. In addition, there is some indication from subject PG that the positioner may be less effective in placing the electrode near the target tissue at the apical part of the array.

E. Comfort-level contours

Figure 10 shows the comfort-level equal-loudness data already shown in Fig. 7. Superimposed on these data are three contours derived from various simulations of the model, as follows. The first contour is the prediction from the model with linearly arrayed elements (as in Fig. 1) with total neuronal activity $\Sigma N(x)$ set to match the measured dynamic range between threshold and comfort for monopolar stimulation for each subject (dashed lines). The second contour is the prediction from the model with linearly arrayed elements with a fixed total neuronal activity of 2000 (dotted lines). The third contour is derived from the model with circularly arrayed elements with a fixed total neuronal activity of 1500 (solid lines). In each case, the electrode-to-tissue distance d used in the model simulations was derived directly from the interaction coefficient K measured at threshold for each subject (see Table III). The total neuronal activity for the last two cases was chosen to optimize the fit across all subjects between the model and the observed data; a further refinement of the analysis would entail an optimization for each individual subject. For the analysis that was performed, in all cases, the model predictions were scaled such that the predicted comfort levels matched the observed comfort levels for the monopolar condition ($\sigma=0$).

Generally speaking, the linearly arrayed model with matched dynamic range (dashed lines in Fig. 10) was the least consistent with the observed data. As discussed earlier, the dynamic range over which the constant peak approximation holds was larger in some of the CI subjects than was derived from the model. With total neuronal activity $\Sigma N(x)=2000$ (dotted lines in Fig. 10), the linearly arrayed model was closer to the observed data for all subjects, except PH,

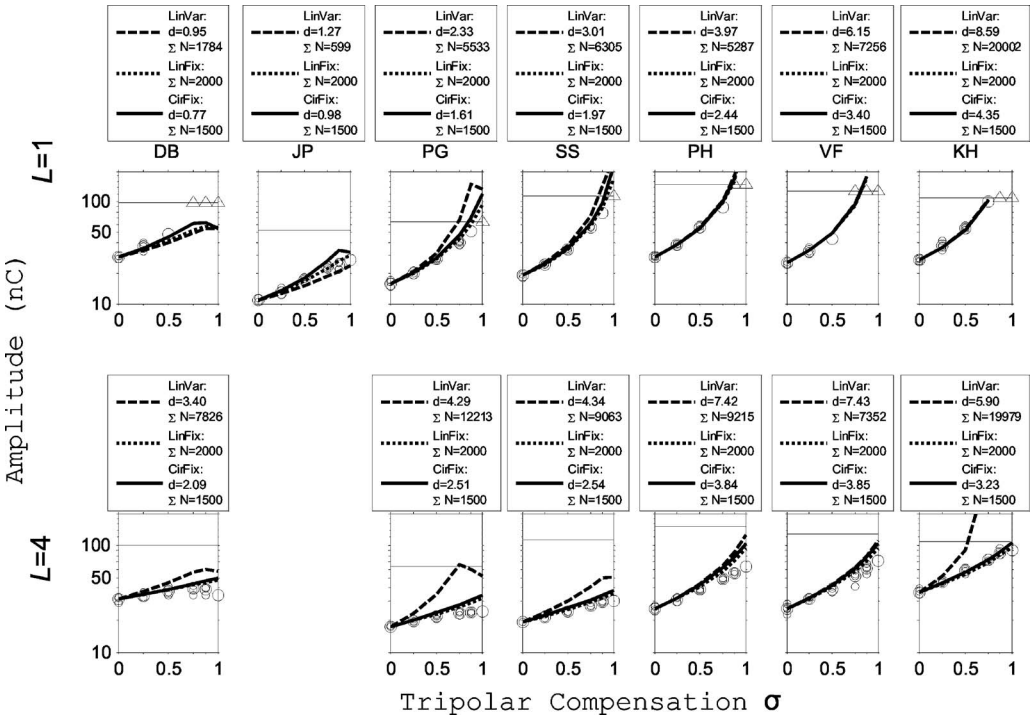


FIG. 10. This figure shows the equal-loudness contours for the seven subjects in a format similar to that of Fig. 7. The lines indicate the predictions of three different model configurations: (1) the linearly arrayed model with total neuronal activity set to match the dynamic range between threshold and comfort (dashed line), (2) the linearly arrayed model with total neuronal activity set to 2000 in all cases (dotted line), and (3) the circularly arrayed model with total neuronal activity set to 1500 in all cases (solid line).

KH, and VF at $L=1$, where the two fits were similar. Good fits were also obtained with the circularly arrayed model utilizing total neuronal activity $\Sigma N(x)=1500$ (solid lines in Fig. 10). This last model was especially successful in predicting the data from the positioner users with inter-electrode spacing $L=1$. However, the last two model configurations were not able to predict the slowdown in the growth of loudness observed for the inter-electrode spacing of $L=4$ for five of the subjects (PG, SS, PH, DB, and VF). Because slower growth in the loudness contour for larger tripolar compensation coefficients σ correlates with the contribution of the sidelobes, more complex models of the cochlea that incorporate more accurate cochlear geometry and nonhomogenous media may be necessary to predict the shape of the loudness contours for the larger inter-electrode spacing.

F. Would partially tripolar excitation be advantageous under all conditions?

The simplified model suggests that it is possible to achieve focusing of neuronal activity patterns under conditions of equal total neuronal activity, especially when using a narrow inter-electrode spacing (Figs. 3 and 5). However, under conditions that include "saturation" of activity in the region of the central electrode or conditions where sidelobes contribute to the activity pattern, larger tripolar compensation coefficients may not necessarily lead to a more narrow spatial extent of neuronal activity. In such cases, the optimal compensation coefficient σ may be less than 1. For all but two of the subjects in this study (both of whom had a positioner), the results indicate that at "comfortable" levels it is possible to narrow the region of neuronal activity with relatively large compensation coefficients. It should be noted that the inferences with regards to neuronal activity patterns made in this paper are indirect, primarily by way of comparing the shapes of equal-loudness contours between the model and CI subjects. More direct measurements of excitation patterns with CI subjects (e.g., using forward masking paradigms) may be necessary to confirm the narrower activity patterns for the partially tripolar configurations (Shannon, 1983; Cohen *et al.*, 2001; Boex *et al.*, 2003; Kwon and van den Honert, 2006).

In some cases, it was not possible to implement the largest tripolar compensations simply because of compliance limits of the current sources on the implant system. For example, as shown in Fig. 7, with the narrow inter-electrode spacing ($L=1$), sensations of comfortable loudness could be achieved with full tripolar compensation ($\sigma=1$) for only one of the seven subjects. Accordingly, a careful assessment of implant limitations should be performed prior to application of a partially tripolar stimulation scheme with a given CI user.

Recently, Mens and Berenstein (Mens and Berenstein, 2005) investigated the effect of tripolar and partially tripolar electrode configurations on speech performance. Two configurations were considered: (1) a partially tripolar configuration with tripolar compensation coefficient σ of 0.5 (using the same definition for σ as in the present model), and (2) a configuration akin to fully tripolar ($\sigma=1$) with the inter-electrode spacing between the center electrode and the compensating electrodes between 1 and 3, depending on the electrode pair. Although poorer performance was reported for the latter configuration, it is unclear whether stimulation was

maintained below the compliance limits of the implant system. As shown earlier, the present study suggests that a fully tripolar configuration with sufficient loudness growth may not be physically achievable in most users. In contrast, the partially tripolar configuration resulted in modest, although nonstatistically significant, improvements in speech understanding in the nonstationary noise condition, suggesting that the use of a nonzero tripolar compensation coefficient σ has the potential to improve performance over a monopolar condition.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Michael Dorman for providing comments on an earlier version of the manuscript. This research was supported by Advanced Bionics Corporation.

Bierer, J. A., and Middlebrooks, J. C. (**2002**). "Auditory cortical images of cochlear-implant stimuli: Dependence on electrode configuration," J. Neurophysiol. **87**, 478–492.

Boex, C., Kos, M. I., and Pelizzone, M. (**2003**). "Forward masking in different cochlear implant systems," J. Acoust. Soc. Am. **114**, 2058–2065.

Bruce, I. C., White, M. W., Irlicht, L. S., O'Leary, S. J., and Clark, G. M. (**1999a**). "The effects of stochastic neural activity in a model predicting intensity perception with cochlear implants: Low-rate stimulation," IEEE Trans. Biomed. Eng. **46**, 1393–1404.

Bruce, I. C., White, M. W., Irlicht, L. S., O'Leary, S. J., Dynes, S., Javel, E., and Clark, G. M. (**1999b**). "A stochastic model of the electrically stimulated auditory nerve: Single-pulse response," IEEE Trans. Biomed. Eng. **46**, 617–629.

Cheng, D. K. (**1993**). *Fundamentals of Engineering Electromagnetics* (Addison-Wesley, Reading, MA).

Cohen, L. T., Saunders, E., and Clark, G. M. (**2001**). "Psychophysics of a prototype peri-modiolar cochlear implant electrode array," Hear. Res. **155**, 63–81.

Efron, B., and Tibshirani, R. (**1993**). *An Introduction to the Bootstrap* (Chapman and Hall, New York).

Fayad, J. N., and Linthicum, F. H., Jr. (**2006**). "Multichannel cochlear implants: Relation of histopathology to performance," Laryngoscope **116**, 1310–1320.

Franck, K. H., Xu, L., and Pfingst, B. E. (**2003**). "Effects of stimulus level on speech perception with cochlear prostheses," J. Assoc. Res. Otolaryngol. **4**, 49–59.

Henry, B. A., and Turner, C. W. (**2003**). "The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 2861–2873.

Henry, B. A., Turner, C. W., and Behrens, A. (**2005**). "Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners," J. Acoust. Soc. Am. **118**, 1111–1121.

Jolly, C. N., Spelman, F. A., and Clopton, B. M. (**1996**). "Quadrupolar stimulation for cochlear prostheses: Modeling and experimental data," IEEE Trans. Biomed. Eng. **43**, 857–865.

Kwon, B. J., and van den Honert, C. (**2006**). "Effect of electrode configuration on psychophysical forward masking in cochlear implant listeners," J. Acoust. Soc. Am. **119**, 2994–3002.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**(2), 467–477.

McKay, C. M., Henshall, K. R., Farrell, R. J., and McDermott, H. J. (**2003**). "A practical method of predicting the loudness of complex electrical stimuli," J. Acoust. Soc. Am. **113**, 2054–2063.

McKay, C. M., Remine, M. D., and McDermott, H. J. (**2001**). "Loudness summation for pulsatile electrical stimulation of the cochlea: Effects of rate, electrode separation, level, and mode of stimulation," J. Acoust. Soc. Am. **110**, 1514–1524.

Mens, L. H., and Berenstein, C. K. (**2005**). "Speech perception with mono- and quadrupolar electrode configurations: A crossover study," Otol. Neurotol. **26**, 957–964.

Miller, A. L., Morris, D. J., and Pfingst, B. E. (**1997**). "Interactions between pulse separation and pulse polarity order in cochlear implants," Hear. Res. **109**, 21–33.

Miller, C. A., Abbas, P. J., Robinson, B. K., Rubinstein, J. T., and Matsuoka, A. J. (**1999**). "Electrically evoked single-fiber action potentials from cat: Responses to monopolar, monophasic stimulation," Hear. Res. **130**, 197–218.

Nadol, J. B., Jr. (**1997**). "Patterns of neural degeneration in the human cochlea and auditory nerve: Implications for cochlear implantation," Otolaryngol.-Head Neck Surg. **117**, 220–228.

Pfingst, B. E., Franck, K. H., Xu, L., Bauer, E. M., and Zwolan, T. A. (**2001**). "Effects of electrode configuration and place of stimulation on speech perception with cochlear prostheses," J. Assoc. Res. Otolaryngol. **2**, 87–103.

Pfingst, B. E., and Xu, L. (**2004**). "Across-site variation in detection thresholds and maximum comfortable loudness levels for cochlear implants," J. Assoc. Res. Otolaryngol. **5**, 11–24.

Raggio, M. W., and Schreiner, C. E. (**1999**). "Neuronal responses in cat primary auditory cortex to electrical cochlear stimulation. III. Activation patterns in short- and long-term deafness," J. Neurophysiol. **82**, 3506–3526.

Rattay, F. (**1990**). *Electrical Nerve Stimulation: Theory, Experiments, and Applications* (Springer-Verlag/Wien, Vienna, Austria).

Rebscher, S. J., Snyder, R. L., and Leake, P. A. (**2001**). "The effect of electrode configuration and duration of deafness on threshold and selectivity of responses to intracochlear electrical stimulation," J. Acoust. Soc. Am. **109**, 2035–2048.

Relkin, E. M., and Doucet, J. R. (**1997**). "Is loudness simply proportional to the auditory nerve spike count?," J. Acoust. Soc. Am. **101**, 2735–2740.

Shannon, R. V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction," Hear. Res. **12**, 1–16.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Snyder, R. L., Bierer, J. A., and Middlebrooks, J. C. (**2004**). "Topographic spread of inferior colliculus activation in response to acoustic and intracochlear electric stimulation," J. Assoc. Res. Otolaryngol. **5**, 305–322.

van den Honert, C., and Stypulkowski, P. H. (**1984**). "Physiological properties of the electrically stimulated auditory nerve. II. Single fiber recordings," Hear. Res. **14**, 225–243.

van den Honert, C., and Stypulkowski, P. H. (**1987**). "Single fiber mapping of spatial excitation patterns in the electrically stimulated auditory nerve," Hear. Res. **29**, 195–206.

Vanpoucke, F. J., Zarowski, A. J., and Peeters, S. A. (**2004**). "Identification of the impedance model of an implanted cochlear prosthesis from intracochlear potential measurements," IEEE Trans. Biomed. Eng. **51**, 2174–2183.

Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners

Leonid M. Litvak[a)]
Advanced Bionics Corporation, 12740 San Fernando Road, Sylmar, California, 91342

Anthony J. Spahr
Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287

Aniket A. Saoji
Advanced Bionics Corporation, 12740 San Fernando Road, Sylmar, California, 91342

Gene Y. Fridman
Department of Biomedical Engineering, University of California Los Angeles, Los Angeles, California, 90095

(Received 11 December 2006; revised 16 May 2007; accepted 18 May 2007)

Spectral resolution has been reported to be closely related to vowel and consonant recognition in cochlear implant (CI) listeners. One measure of spectral resolution is spectral modulation threshold (SMT), which is defined as the smallest detectable spectral contrast in the spectral ripple stimulus. SMT may be determined by the activation pattern associated with electrical stimulation. In the present study, broad activation patterns were simulated using a multi-band vocoder to determine if similar impairments in speech understanding scores could be produced in normal-hearing listeners. Tokens were first decomposed into 15 logarithmically spaced bands and then re-synthesized by multiplying the envelope of each band by matched filtered noise. Various amounts of current spread were simulated by adjusting the drop-off of the noise spectrum away from the peak (40–5 dB/octave). The average SMT (0.25 and 0.5 cycles/octave) increased from 6.3 to 22.5 dB, while average vowel identification scores dropped from 86% to 19% and consonant identification scores dropped from 93% to 59%. In each condition, the impairments in speech understanding were generally similar to those found in CI listeners with similar SMTs, suggesting that variability in spread of neural activation largely accounts for the variability in speech perception of CI listeners. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2749413]

PACS number(s): 43.64.Me, 43.71.An, 43.71.Es, 43.71.Ky [AJO] Pages: 982–991

I. INTRODUCTION

Although the average outcomes of cochlear implant (CI) procedures have improved over the past decade, large variability in the ability to understand speech by CI listeners is commonly reported (e.g., Firszt *et al.*, 2004). The variability in outcomes is not adequately explained by pre-operative factors such as age and duration of hearing loss. Although several hypotheses have been proposed in order to explain the variability among the patient population, one likely possibility is that differences in performance may be accounted for by differences in spatial selectivity in CI listeners. The evidence for this hypothesis is twofold. First, several psychophysical measures thought to be related to spatial selectivity have been found to correlate to speech perception in CI listeners (Eddington *et al.*, 1997a; Henry and Turner, 2003; Henry *et al.*, 2005; Saoji *et al.*, 2005). Second, normal-hearing (NH) subjects listening to simulations of reduced spectral resolution exhibit a range of performance that matches that observed in best CI patients (Shannon *et al.*, 1995; Friesen *et al.*, 2001). The goal of the present work is to determine whether the "spectral selectivity" hypothesis is sufficient to entirely account for the range of speech performance observed in CI listeners. In particular, the performances of NH listeners on a nonspeech spectral resolution task are matched with those of CI listeners, by introducing varying amounts of spectral smearing (Baer *et al.*, 1993; Baer and Moore, 1994) for the NH listeners. The performance of CI listeners on speech identification tasks is then compared to that of NH listeners, with respect to their performance on the nonspeech spectral resolution task.

Several metrics have been proposed as measures of spectral selectivity, including (1) forward masking patterns, which may be assessed either psychophysically (Boex *et al.*, 2003; Cohen *et al.*, 2003), or using evoked potentials (Abbas *et al.*, 2004), (2) simultaneous threshold interaction (Eddington *et al.*, 1997b), or (3) spectral shape perception (Henry and Turner, 2003; Henry *et al.*, 2005). All of the psychophysical measures have been reported to correlate in varying degrees to speech perception, with spectral shape perception measures producing the best correlations, possibly because those measures assess spectral resolution across the whole cochlea. In the spectral shape task, subjects are asked to discriminate between locations of spectral peaks. For example, Henry *et al.* (2005) reported on a task where the subjects are

[a)]Author to whom correspondence should be addressed. Electronic mail: leonidl@advancedBionics.com

required to discriminate between two spectra, each of which resembles a full-wave rectified sinusoid in the frequency domain. Each subject's spectral resolution was quantified by the highest ripple frequency (in cycles/octave) with 30 dB contrast that the subject could identify as different from one where the spectral peaks versus valleys have been reversed in frequency location. More recently, Saoji *et al.* (2005) reported on spectral modulation transfer functions of cochlear implant recipients using methods that are similar to those previously utilized in normal-hearing listeners (Bernstein and Green, 1987). In that study, the authors quantified the necessary peak-to-valley contrast to differentiate between a spectral ripple and white noise as a function of spectral ripple frequency. They showed that spectral modulation thresholds (SMTs) corresponding to the lowest spectral ripple frequencies (0.25 and 0.5 cycles/octave) best related to speech recognition. Because ability to understand speech is of greatest relevance to the present study, SMT at these ripple frequencies was chosen as the measure of spectral resolution.

Studies of acoustic hearing have demonstrated that speech recognition of NH listeners is decreased with spectral smearing (Baer *et al.*, 1993; Baer and Moore, 1994; Shannon *et al.*, 1995; Dorman *et al.*, 1997; Friesen *et al.*, 2001). Several studies [e.g. (Shannon *et al.*, 1995; Dorman *et al.*, 1997; Friesen *et al.*, 2001)] simulated performance of CI patients by exposing NH listeners to "noise vocoders," which reduced spectral information into a limited number of "channels." Friesen *et al.* (2001) demonstrated that for 7–8 channels, NH listeners using an equivalent number of channels performed better than the average CI performer, but comparably to the best CI performers. However, it is unclear from Friesen *et al.* whether performance of poorer listeners relates specifically to loss of spectral resolution in CI listeners. Fu and Nogaki (2005) showed that performance of NH listeners can be reduced further by spectrally smearing the output of the vocoder bands by using overlapping noise carriers. However, unlike in the present study, no attempt was made to match the degree of smearing to specific CI subjects.

II. METHODS

A. Subjects

Ten normal hearing subjects ranging in age from 22 to 26 years participated in these experiments. Each subject had normal hearing based upon pure-tone thresholds (<20 dB Hearing Level (HL) from 250 to 8000 Hz, ANSI, 1996) and screening tympanograms (*Y*, 226 Hz). All studies have been approved by a private Institutional Review Board (IRB), as well as by the institutional IRB at the Arizona State University.

The speech perception and SMT data for CI subjects reported here are taken from Saoji *et al.* (2005) and will be published in a separate article. Twenty five Advanced Bionics CII or HR/90k cochlear implants (ranging in age from 38 to 65 years) listeners with varying speech perception abilities participated in that experiment. All subjects had at least one year of experience with their cochlear implants.

B. Stimuli

All stimuli were generated using MATLAB software (Mathworks, 2006). The stimuli were generated in the frequency domain assuming a sampling rate of 44,100 Hz. First, the desired spectral shape was generated using the equation

$$|F(f)| = \begin{cases} 10^{\frac{C}{2}\sin(2\pi(\log_2(f/350))\cdot f_c+\theta_0)/20} & 350 < f < 5600 \\ 0 & otherwise \end{cases}, \quad (1)$$

where $F(f)$ is the amplitude of a bin with center frequency f Hz, f_c is the spectral modulation frequency (in cycles/octave), and θ_0 is the starting phase. The desired noise band was synthesized by adding a random phase to each bin, and taking an inverse Fourier transform. The flat noise stimuli were generated using a similar technique, except that spectral contrast C was set to 0. The amplitude of each stimulus was adjusted to an overall level of 60 dB sound pressure level (SPL). Independent noise stimuli were presented on each observation interval. The stimulus duration was 400 ms.

Speech understanding was assessed using a vowel and consonant identification task. Vowel stimuli consisted of 13 vowels created with the use of KLATT software (Klatt, 1980) in /bVt/ format ("bait, bart, bat, beet, bert, bet, bit, bite, boat, boot, bought, bout, but"). The vowels were brief (90 ms) and of equal duration so that vowel length would not be a cue to identity (Dorman *et al.*, 1989). The stimuli for the tests of consonant identification were 16 male-voice consonants in the /aCa/ context, originally taken from the Iowa laser video disk (Tyler *et al.*, 1986).

C. Spectral modulation thresholds

Normal-hearing listeners were tested in a double-walled sound treated room using Sennheiser HD 25-SP1 circumaural headphones and all stimuli were presented at 60 dB SPL.

As reported by Saoji *et al.* (2005), the CI listeners used their everyday program in all test conditions. The stimuli were output from a standard PC to an Audiophile soundcard. The output of the soundcard was fed to the body worn Platinum Series Processor through the Advanced Bionics Direct Connect® system. Sound card output was attenuated using an inline attenuator such that for all stimuli, the electric input to the DirectConnect® system was equivalent to a 60 dB SPL acoustic input to the microphone of the speech processor. A cued two interval, two-alternative, forced choice procedure was used to collect data. Prior to data collection subjects received a few sample trials for any new condition to familiarize them with the stimuli. On each trial, the three observation intervals were separated by 400 ms silent intervals. In the first interval the standard stimulus was always presented. This cuing or reminder interval is helpful in cases where listeners can hear a difference between the signal and the standard stimulus but cannot identify which is which. The standard stimulus had a flat spectrum with bandwidth extending from 350 to 5600 Hz. The signal and the second standard were randomly presented in the other two intervals. Thresholds were estimated using an adaptive psychophysical

procedure employing 60 trials. The signal contrast level was reduced after three consecutive correct responses and increased after a single incorrect response. Initially the contrast was varied in a step size of 2 dB, which was reduced to 0.5 dB after three reversals in the adaptive track (Levitt, 1971). Threshold for the run was computed as the average modulation depth corresponding to the last even number of reversals, excluding the first three. The equilibrium point of such a procedure is 79.4% correct. Using the above procedure, modulation detection thresholds were obtained for the modulation frequencies of 0.25 and 0.5 cycles/octave. A threshold was defined as the average of three 60 trial runs. The average of the thresholds for the two modulation frequencies was used as the measure of spectral resolution, as this measure was found to best correlate to consonant and vowel recognition (Saoji *et al.*, 2005).

D. Vowel and consonant recognition

During both the vowel and consonant identification tasks, listeners completed a practice session, in which they heard each vowel presented twice while the text representation of the token was visually displayed on the computer screen. Subjects then completed two repetitions of the test procedure, with feedback, as a final practice condition. In the test condition, vowel identification performance was measured in two blocks of 78 trials in which each of the 13 vowels was presented six times in random order. Thus, vowel identification and the resulting confusion matrix for each subject were based on 12 presentations of each vowel stimulus. Likewise, consonant identification performance was measured in two blocks of 80 trials in which each of the 16 consonants was presented five times in random order. Consonant identification and the resulting confusion matrix for each subject were based on a total of ten presentations per consonant stimulus. The vowel and consonants were presented at an overall level of 60 dB SPL and the overall level was randomized by 3 dB (±1.5 dB) in 0.5 dB steps. The randomization would discourage the listeners from using loudness cues while performing the vowel and consonant recognition task.

As reported by Saoji *et al.* (2005), due to time constraints, consonant identification was not measured in one CI subject. In addition, a loss of data occurred on the experimental computer, whereby the confusion matrices were not stored for three CI subjects on the vowel task, and for four CI subjects on the consonant task.

E. Vocoder simulations

Vocoder simulations utilized in this study were designed to model both the processing typically performed in a cochlear implant, and spread of excitation that may occur in electrically stimulated cochlea. Each token was digitally sampled at 17,400 Hz. Short-time Fourier transform was computed with resolution of 256 bins, and temporal overlap of 192 samples (Oppenheim and Schafer, 1975). Next, individual bins were grouped into 15 nonoverlapping, logarithmically spaced analysis channels. The envelope of each channel was computed on a frame-by-frame basis by computing the square root of the total energy in the channel. The energy computation implies an implicit envelope detector with a low-pass filter whose cutoff equals the bandwidth of each bin, or 68 Hz. The output of each channel was used to modulate a noise band. The noise band was similarly synthesized in the frequency domain. The center frequency of the noise band was identical to the center frequency of the corresponding analysis channel. The rate of the drop-off of the noise spectrum away from the center frequency was varied from 5 to 40 dB/octave, to simulate various amounts of spread of excitation that may occur in an electrically stimulated cochlea. The desired time-frequency output pattern for each channel was computed by multiplying its instantaneous energy by the corresponding spectral envelope of the noise band. The time-spectral patterns corresponding to each channel were then added to compute the total time-spectral pattern. The overall bandwidth of the signal was limited to 8700 Hz. Finally, temporal output wave form was computed by first adding random phase to each bin, and then computing an inverse short-time Fourier transform (Oppenheim and Schafer, 1975).

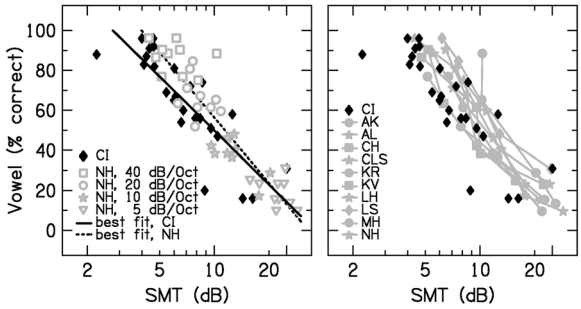


FIG. 1. (Color online) Vowel speech performance of the CI and NH listeners plotted against their respective SMT scores. Panels A and B contain the same data. In each panel, the data from CI listeners are indicated with black diamonds. Panel A shows the performance of the NH listeners with open symbols, where each symbol corresponds to the noise slope of the vocoder used in the simulation. Panel B shows that the individual performance for each NH listener follows the same trend as the average data. The lines in panel A are the best-fit lines for the CI patients and the NH listeners.

Ten normal-hearing (NH) listeners participated in this study. Each NH listener listened to four different vocoder simulations, which differed only in the slopes of the noise bands. Based on a small pilot study, slopes of 5, 10, 20, and 40 dB/octave were chosen. For each simulation condition, the SMT at 0.25 and 0.5 cycles/octave, as well as performance on vowel and consonant recognition, was measured.

III. RESULTS

A. Vowel recognition

Figure 1 panels show the results of vowel recognition as a function of SMT for the CI patients and NH listeners. Panels A and B demonstrate different aspects of the same data. The data collected from the CI listeners are shown in black diamonds in all two panels.

Large CI subject variability was observed in both vowel recognition and SMT tasks. As in previously reported studies (Henry *et al.*, 2005; Saoji *et al.*, 2005), the SMT is strongly correlated with vowel recognition ($r=-0.84$). For the NH

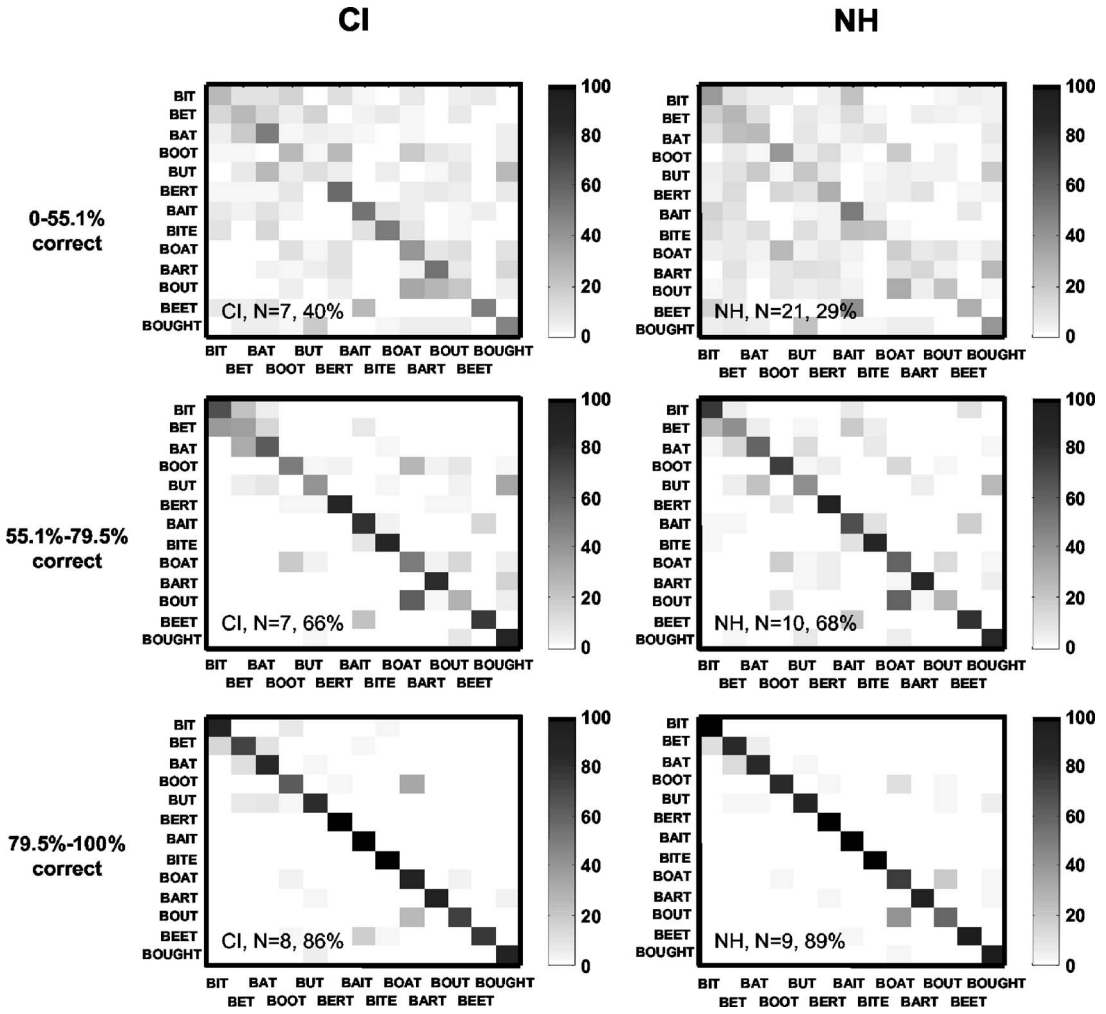


FIG. 2. Patterns of errors which occurred during the vowel speech tests. Confusion matrices for the vowel speech scores for the CI patients are presented in the left panels and those for the NH listeners in the right panels. The subjects were separated based on their vowel test scores (indicated in percent correct on the left of the matrices), so that each group contained the data from approximately the same number of CI subjects. The *y* axis corresponds to the stimulus, and the *x* axis corresponds to the response.

listeners using vocoder simulations, the vowel scores decreased as the noise-band slopes were broadened. The average as well as individual SMTs also tended to deteriorate from 6.3 dB for 40 dB/octave (narrow noise band) condition to 22.5 dB for 5 dB/octave (broad noise band) condition. Thus, use of progressively shallower noise-band slopes had an effect of decreasing spectral resolution abilities of NH listeners. The range of SMTs observed in NH subjects listening to simulations was similar to the range observed in CI listeners (2.25–25 dB). In addition, as shown in panel B, the individual scores on the vowel recognition task decreased with shallower noise-band slopes. Comparing the NH data to the corresponding data for CI recipients, a good quantitative agreement is reached in that the NH data overlays the CI data. For quantitative comparison, panel A shows the data from all of the simulation conditions, pooled together so that vowel identification abilities of CI patients can be directly compared to that of NH subjects under condition of similar spectral resolution. The pooling of the data is justified because the relationship between the SMT and vowel recognition within each group (different open symbols in Panel A) appears to be similar to the relationship across groups. The lines indicate the best linear fits of the CI and NH data. Although one can observe slight differences in the fits, both the differences between the two lines in the range from 4 to 30 dB are not statistically different [$p=0.57$, permutation test (Good, 1994)]. Thus, both statistical analysis and visual inspection suggest that a modified vocoder that models spectral resolution of cochlear implant listeners also models vowel identification performance.

Figure 2 compares the pattern of errors made by NH subjects listening in various simulation conditions and matched CI subjects. Each of the rows of panels shows the average confusion matrices for CI subjects (left) and NH listeners (right) with overall performance of 0–55.1% correct (top), 55.1–79.5% correct (middle), and 79.5–100% correct (bottom). These performance ranges were chosen such that there is an approximately equal number of CI subjects in each range. The data for NH group were combined across simulation conditions. The similarity in scores was chosen as the basis for assigning subjects to a group so that the results in this section can be readily compared to those for consonant recognition task, where the equivalent performance is not achieved at the same spectral resolution between the two

TABLE I. Correlation coefficients between the confusion matrices for the vowel data computed either with or without the main diagonal. The probability that the two populations have the same confusion patterns is indicated by the p level, and was computed using a nonparametric method described in the text.

	Number		Correlation between NH and CI data	
Category	NH group	CI group	With diagonal	Without diagonal
0.0–55.1	(N=21)	(N=7)	0.772(P=26.8%)	0.695(P=22.0%)
55.1–79.5	(N=10)	(N=7)	0.971(P=57.9%)	0.858(P=45.6%)
79.5–100.0	(N=9)	(N=8)	0.985(P=12.9%)	0.716(P=4.4%)

groups. The results for vowels would be similar if spectral resolution would be chosen as the basis for the grouping.

Visual inspection of the two confusion patterns in each row of Fig. 2 indicates many similarities, but also some differences, between the two plots. Table I shows the correlations between the confusion matrices for NH and CI listeners. The correlation coefficient between the two matrices (computed by treating each matrix as a set of ordered numbers) is higher than 0.75 for all groups, and is especially high for the two higher-performing groups (0.97 and 0.99, respectively). The correlations remain reasonably high (close to or above 0.7) even if the entries on the main diagonal of both the NH and CI matrices are set to zero prior to computing the correlation. The latter manipulation is of interest because only the errors contribute to the off diagonal entries; hence correlation without the off diagonal allows comparison of the error pattern almost independently of the overall score.

A version of a significance test based on the bootstrap method (Efron and Tibshirani, 1993) was undertaken to determine whether the differences between the corresponding matrices are due to the variability seen between subjects and between tests, or whether the variability reflects true differences in the errors made by CI and NH listeners. The null hypothesis was that both sets of confusion matrices (i.e., those from CI and NH listeners) correspond to the same population. Under the null hypothesis, expected range of correlations between the average NH and CI matrices could be computed by mixing up the individual data, and re-dividing the subjects arbitrarily into the two groups corresponding to the original NH and CI groups, and computing the correlation between the average matrices for the two new groups. The "re-drawing" procedure was repeated 1000 times, and the p values were computed as proportions of the time that the correlations obtained from the bootstrap procedure were higher than the observed correlation between NH and CI matrices. Note that if the observed correlation is significantly lower than the distributions under the null hypothesis, then the null hypothesis is invalidated, which means that the two sets of matrices come from significantly different populations. For all of the three performance groups, the correlation coefficients observed in the original data were not significantly different from those observed under the null hypothesis ($P>4.4\%$), suggesting that the two populations are not significantly different.

B. Consonant recognition

The panels of Fig. 3 plot consonant recognition as a function of SMT. As in Fig. 1, the panels show different aspects of the same data. As in Fig. 1, the data collected from the CI listeners are shown in black diamonds in all three panels.

In panel A, the data for NH listeners are shown with open symbols. The symbol shapes correspond to the particular noise-band slopes applied in the vocoder simulations. Panel B shows performance of individual NH subjects. As for vowels, SMT was strongly correlated with consonant recognition of both CI subjects and NH subjects listening to simulations ($r=-0.82$ and -0.88, respectively). However, as compared to the vowel scores shown in Fig. 1, the consonant scores dropped off less with decreased spectral resolution (16% per doubling versus 26% per doubling). The lower dependence of consonant scores on spectral resolution is consistent with the observation that consonant recognition is less dependent on spectral cues as compared to vowel recognition. As with the vowel identification task, the highest scores and the lowest SMTs correspond to the condition with the largest noise-band slope (40 dB/octave). Panel B shows similar performance trends of individual NH subjects for each condition.

As with the vowel scores, the consonant scores of NH listeners decreased with shallower noise-band slopes. However, whereas the vowel data for CI listeners matched closely to the NH listeners with the same SMT, this was not the case for the consonant data. In particular, at the same spectral resolution, performance on the consonant task was generally higher for the NH vocoder listeners. The difference between the two is quantified in panel A where the data from NH listeners are pooled across different simulation conditions, and the best-fit lines were fit to both sets of data. Permutation test indicated significant ($p=0.006$) differences between the two line fits in the region of SMT from 3.5 to 30 dB (Good, 1994). The slope of the best-fit line was not different in the

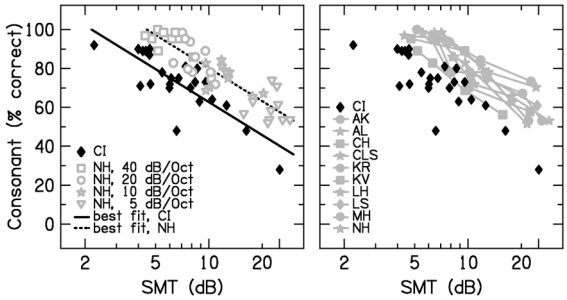


FIG. 3. (Color online) Consonant speech performance of the CI and NH listeners plotted against their respective SMT scores. The format of the plots is identical to that in Fig. 1.

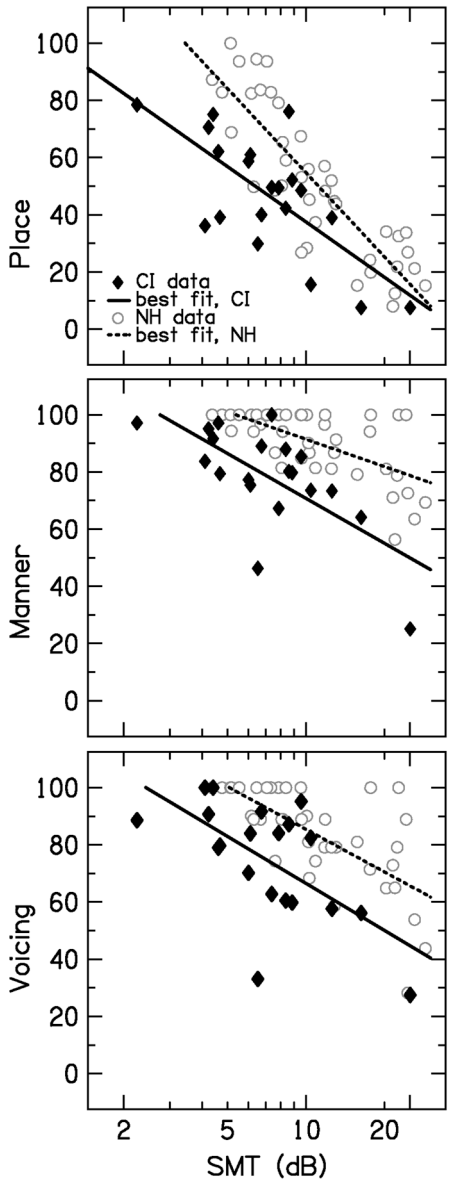


FIG. 4. (Color online) Identification abilities of place, manner, and voicing cues versus the SMT for CI subjects and NH listeners listening through vocoder simulations. The data for NH listeners is pooled across simulation conditions. The best-fit lines show the similarity in the data trends for each group.

two conditions ($p=0.65$), indicating that the degree of degradation of the consonant scores of CI and NH listeners with equivalent decrease of SMT was similar. As shown in Fig. 4, similar difference was observed in all classical production-based features of voicing, manner, and place of articulation (Miller and Nicely, 1955). Although NH listeners tended to perform better on all features, the differences across individual features were not significant ($p>0.09$).

Figure 5 and Table II repeat the error pattern analysis for the consonant recognition. The subjects were split into three groups based on the consonant test performance. The statistical analysis for the correlations between the matrices obtained for the CI listeners and those obtained for the NH listeners is described in the previous section. The correlations between the NH and CI confusion matrices were comparable to those obtained on vowel matrix comparison. However, the p-value analysis of the confidence of the correlation measures revealed that these correlation measures were not highly significant. This statistical analysis suggests that, despite the high degree of similarity, there were also significant differences between the confusion matrices of the NH compared to the CI listeners.

IV. DISCUSSION AND CONCLUSIONS

A. Vowels versus consonants

When matched in spectral resolution abilities, performance of NH subjects listening through vocoder simulations was similar to that of CI patients. The match between the two populations was greatest for the vowel recognition task, where the average performance, variability in performance, and confusion patterns were all similar between the two groups. In addition, the decrease in the consonant scores observed in NH listeners as a function of decreased spectral resolution was similar to that of the CI patients as reported by Saoji *et al.* (2005). However, even at equivalent spectral resolution abilities, somewhat lower overall scores were observed for the CI patients in the consonant recognition task. While many studies have reported that CI listeners perform more poorly than NH subjects on spectral tasks, these results suggest that CI patients have also deficits in perception of nonspectral (amplitude/temporal) cues as compared to NH listeners. Such deficits should only affect the consonant scores because the "synthetic" vowel recognition task relies almost exclusively on spectral cues. For example, Fu, 2002 found the highest correlation between consonants and temporal modulation and a smaller correlation between vowels and temporal modulation. Alternatively, these temporal deficits may be partially explained by the age differences between the NH and CI groups (Ohde and Abou-Khalil, 2001).

While this observation would suggest that CI patients have deficits in the temporal/amplitude domain as compared with NH listeners, this assertion is not fully supported by the data in Fig. 4. As this figure indicates, roughly equivalent differences between the NH and CI listeners are observed for features with primarily spectral cues (such as place), and primarily temporal/amplitude cues (such as voicing or manner). More data will be required to ascertain whether deviations in the features which are primarily based on the temporal and amplitude cues are more significant than deviations of the features with the spectral cues.

B. Factors affecting speech perception abilities of CI patients

The ability to recognize speech varies substantially across CI users. Several hypotheses have been advanced as to the underlying causes of this variability. First, it has been suggested that variability may be due to variability in electrode placement relative to the natural frequency-to-place alignment in the cochlea. While such frequency-to-place misalignment might exist in some subjects, and while acute changes to the frequency-to-place alignment in the vocoder simulations and normal-hearing listeners can lead to large decreases in performance (for a recent review see Baskent

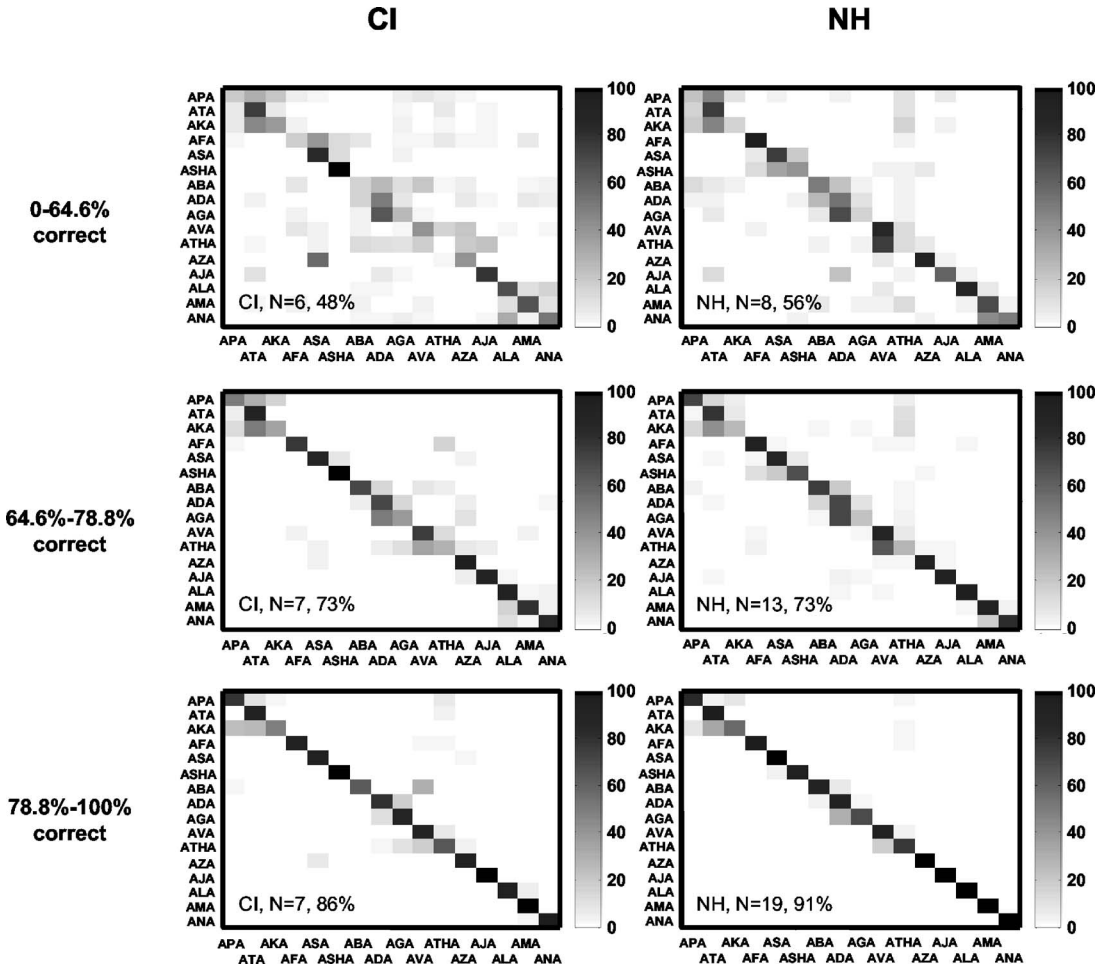


FIG. 5. Patterns of errors which occurred during the consonant speech tests. Confusion matrices for the consonant speech scores for the CI patients are presented in the left panels and those for the NH listeners in the right panels. The subjects were separated based on their consonant test scores (indicated in percent correct on the left of the matrices), so that each group contained the data from approximately the same number of CI subjects.

and Shannon, 2003), several reports suggest that both CI as well as NH listeners are able to partially adapt to such misalignment over time (Rosen *et al.*, 1999; Harnsberger *et al.*, 2001; Fu *et al.*, 2002). For example, Harnsberger *et al.* (2001) examined perceptual "vowel spaces" of synthetic vowels which differed only by first two formants. They found that with one exception, there were no systematic shifts observed in the perceptual vowel spaces of CI listeners relative to those established for NH subjects. Since the subjects that participated in the present study had substantial experience with their CI (on average more than 2 years), it is unlikely that variability observed in speech scores can be accounted by such frequency misalignments.

CI subjects also differ in their ability to perceive temporal modulations that are essential for proper perception of speech. Perceptions of temporal modulations (above 400 Hz) have been shown to not be correlated to speech scores (Cazals *et al.*, 1994). However, several investigators reported strong correlation between measures of low frequency ($<$400 Hz) temporal modulation and speech recognition tasks (Cazals *et al.*, 1994; Fu, 2002), suggesting a causal relationship between the two measures. Low frequency information provides periodicity and envelope cues, indicative of voicing and manner, respectively (Rosen, 1992). The relationship between low frequencies and manner and voicing is consistent with the results obtained in (Fu, 2002) that indicates a stronger correlation between temporal modulation thresholds and consonants as opposed to vowels.

However, a causal relationship between temporal modu-

TABLE II. Comparison of the confusion matrices for the consonant data. The format is similar to that of Table I.

Category	Number		Correlation between NH and CI data	
	NH group	CI group	With diagonal	Without diagonal
0.0–64.6	(N=8)	(N=6)	0.752 (P=0.0%)	0.545 (P=0.0%)
64.6–78.8	(N=13)	(N=7)	0.964 (P=9.0%)	0.820 (P=2.1%)
78.8–100.0	(N=19)	(N=7)	0.985 (P=4.5%)	0.627 (P=0.2%)

lation perception and speech recognition is not consistent with results from vocoder simulations, which demonstrate that speech performance is only slightly affected by severe degradations in temporal information. Xu *et al.* (2005) examined reductions in temporal information in a vocoder simulation by low-pass filtering the envelope down to as much as 1 Hz. Since Xu *et al.* (2005) utilized the second order Butterworth low-pass filter, which decreases the response by 12 dB for every doubling in frequency beyond the low-pass cutoff, such filtering would translate to an 80 dB attenuation of temporal modulations in the 100 Hz range (Cazals *et al.*, 1994; Fu, 2002). The effect of reduction in temporal information on vowel perception was approximately 20%. Deficits in processing of temporal modulations alone are therefore not sufficient to account for almost an 80% difference in scores observed between the best and the worst performers on the vowel identification task.

Across-subject differences in stimulation selectivity of electrically stimulated cochlea have been observed using a variety of techniques (Boex *et al.*, 2003; Cohen *et al.*, 2003, 2004, 2005). Recently, several measures of spatial selectivity that simultaneously assess selectivity across the whole array have been found to moderately correlate to speech perception (Henry and Turner, 2003; Henry *et al.*, 2005). These correlations have been improved using techniques invoked in the present and previous manuscripts which rely on measuring spectral perception specifically at the lowest modulation frequencies which are apparently most important for speech perception of CI listeners (Saoji *et al.*, 2005). The relatively high correlations between spectral resolution and speech perception are consistent with a causal relationship between speech perception and spatial selectivity. The present study with NH subjects listening through the modified vocoder provides further support for the causal relationship, because it shows that performance of NH subjects with similar reductions in spectral selectivity is similar to that observed in CI patients. The variability in performance observed in NH listeners under these conditions is similar to variability observed in CI patients. In the case of consonant recognition, the data suggest a uniform temporal deficit in CI listeners.

C. Is the SMT a measure of peripheral spatial selectivity?

SMT can be interpreted as a measure of spectral selectivity (Dubno and Dorman, 1987; Horst, 1987). Poor peripheral spatial selectivity can smear the spectral contrast and thereby increase the SMT. Alternatively, differences in SMTs across patients may also reflect differences in sensitivity to internal contrast. This "efficiency" hypothesis would suggest that across-subject differences in SMT should be roughly independent of the spectral modulation frequency. Saoji *et al.* (2005) reported on SMT to spectral modulation frequencies of 0.25, 0.5, 1 and 2 cycles/octave. Subjects that were sensitive to spectral modulations of 0.25 and 0.5 cycles/octave tended to have thresholds that were least dependent on spectral modulation frequency, while subjects with the elevated SMTs at 0.25 and 0.5 cycles/octave tended to have greater increases in SMT for the frequencies of 1 and 2 cycles/octave. Thus, Saoji *et al.* (2005) supports the notion that the SMTs at 0.25 and 0.5 cycles/octave reflect spectral selectivity of CI listeners.

If SMT of CI listeners is partially determined by spatial selectivity at the periphery, then these results would suggest that more selective arrays or stimulation paradigms will lead to an improvement in speech recognition. Although the simpler of the approaches, a "peripheral selectivity" hypothesis is not the only possibility. Because central processes may play a role in recognition of spectral patterns, it is also possible that loss of spectral resolution may result without peripheral loss of selectivity. For example, with long duration of deafness, the structure of the auditory cortex changes profoundly. Such changes may negatively affect the ability of the listeners to discriminate spectral patterns (Irvine *et al.*, 2000).

It has been well documented that stimulation strategies employed in CIs may lead to activation patterns in the cochlea which may be very different from normal. For example, electrical stimulation may not properly preserve the crucial phase relationships observed in the firing patterns of healthy auditory nerves to realistic sounds (Loeb *et al.*, 1983; Carney, 1994; Loeb, 2005). In addition, electrical stimulation may lead to activation patterns that are unnaturally synchronized to the individual electrical pulses (Dynes and Delgutte, 1992; Litvak *et al.*, 2001; Ferguson *et al.*, 2003). The differences in individual performance on speech perception as well as nonspeech spectral tasks may therefore reflect the varying ability of subjects to adapt to such unnatural activation patterns.

Finally, it is possible that a deficit in more central processes that are unrelated to specifics of electrical stimulation patterns may be responsible for effective loss of spectral resolution. One argument in favor of the last possibility is that similar spectral and temporal deficits are encountered in hearing impaired (HI) listeners, although to a lesser degree than in the CI patients (Bacon and Viemeister, 1985; Glasberg and Moore, 1986; Summers and Leek, 1994; Henry *et al.*, 2005).

D. Comparison with other vocoder simulations

The results of the present study as well as those of Fu and Nogaki (2005) and Shannon *et al.* (1998) suggest that speech perception abilities of normal hearing listeners listening to CI simulations can be degraded by varying the overlap of the noise carriers. In contrast, several earlier reports accomplished the same task by varying the simulated number of analysis channels (Shannon *et al.*, 1995; Friesen *et al.*, 2001; Xu *et al.*, 2005). At face value, the "overlap" approach has greater resemblance to the clinical processors of most CI patients, who are almost always fit with strategies that utilize a greater number of analysis channels and electrodes than equivalent "effective channels" (Shannon *et al.*, 1995). It is possible therefore that the overlap approach may be responsible for the close match between the vowel confusion patterns between CI subjects and NH listeners listening to simulations. If similar correspondence cannot be achieved with the "effective channel" simulations, then the "overlap" simu-

lations, possibly extended to include some deficits in temporal processing, may be a more effective tool for modeling performance of CI listeners.

The "overlap" approach advocated in this paper provides a rich space in which to explore effects of spatial selectivity on performance. For example, if tools can be established that effectively measure spatial selectivity in separate regions of the cochlea, then "overlap" simulations can be modified accordingly to include various overlaps in various bands. Another productive direction may be to manipulate the "speech processor" part of the simulation. If strategies can be proposed that attempt to overcome poor spectral resolution of some CI listeners, then these strategies can be incorporated into the "speech processor" part of the simulation, and thus evaluated in NH subjects in parallel with CI patients. Comparison of effects between the two groups may lead to better understanding of limitations of CI performance.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Robert Shannon and Dr. Michael Dorman for their encouragement throughout this research, and useful feedback on an earlier version of the manuscript. This research was sponsored by Advanced Bionics Corporation.

Abbas, P. J., Hughes, M. L., Brown, C. J., Miller, C. A., and South, H. (**2004**). "Channel interaction in cochlear implant users evaluated using the electrically evoked compound action potential," Audiol. Neuro-Otol. **9**, 203–213.

American National Standards Institute (ANSI) (**1996**). "American standard specification for audiometers" (American National Standards Institute, New York).

Bacon, S. P., and Viemeister, N. F. (**1985**). "Temporal-modulation transfer functions in normal-hearing and hearing-impaired listeners," Audiology **24**, 117–134.

Baer, T., and Moore, B. C. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277–2280.

Baer, T., Moore, B. C., 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.

Baskent, D., and Shannon, R. V. (**2003**). "Speech recognition under conditions of frequency-place compression and expansion," J. Acoust. Soc. Am. **113**, 2064–2076.

Bernstein, L. R., and Green, D. M. (**1987**). "Detection of simple and complex changes of spectral shape," J. Acoust. Soc. Am. **82**, 1587–1592.

Boex, C., Kos, M. I., and Pelizzone, M. (**2003**). "Forward masking in different cochlear implant systems," J. Acoust. Soc. Am. **114**, 2058–2065.

Carney, L. H. (**1994**). "Spatio-temporal encoding of sound level: Models for normal encoding and recruitment of loudness," Hear. Res. **76**, 31–44.

Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C. (**1994**). "Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants," J. Acoust. Soc. Am. **96**, 2048–2054.

Cohen, L. T., Lenarz, T., Battmer, R. D., Bender von Saebelkampf, C., Busby, P. A., and Cowan, R. S. (**2005**). "A psychophysical forward masking comparison of longitudinal spread of neural excitation in the contour and straight nucleus electrode arrays," Int. J. Audiol. **44**, 559–566.

Cohen, L. T., Richardson, L. M., Saunders, E., and Cowan, R. S. (**2003**). "Spatial spread of neural excitation in cochlear implant recipients: Comparison of improved ECAP method and psychophysical forward masking," Hear. Res. **179**, 72–87.

Cohen, L. T., Saunders, E., and Richardson, L. M. (**2004**). "Spatial spread of neural excitation: Comparison of compound action potential and forward-masking data in cochlear implant recipients," Int. J. Audiol. **43**, 346–355.

Dorman, M. F., Dankowski, K., McCandless, G., and Smith, L. (**1989**). "Identification of synthetic vowels by patients using the Symbion multichannel cochlear implant," Ear Hear. **10**, 40–43.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**, 2403–2411.

Dubno, J. R., and Dorman, M. F. (**1987**). "Effects of spectral flattening on vowel identification," J. Acoust. Soc. Am. **82**, 1503–1511.

Dynes, S. B., and Delgutte, B. (**1992**). "Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea," Hear. Res. **58**, 79–90.

Eddington, D., Tierney, J., and Long, C. (**1997a**). *Cochlear Implants* (RLE, Cambridge, MA).

Eddington, D. K., Rabinowitz, W. R., Tierney, J., Noel, V., and Whearty, M. (**1997b**). "Speech processors for auditory prostheses," 8th Quarterly Progress Report, NIH Contract No. N01–DC-6–2100."

Efron, B., and Tibshirani, R. (**1993**). *An Introduction to the Bootstrap* (Chapman and Hall, New York).

Ferguson, W. D., Collins, L. M., and Smith, D. W. (**2003**). "Psychophysical threshold variability in cochlear implant subjects," Hear. Res. **180**, 101–113.

Firszt, J. B., Holden, L. K., Skinner, M. W., Tobey, E. A., Peterson, A., Gaggl, W., Runge-Samuelson, C. L., and Wackym, P. A. (**2004**). "Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems," Hear. Res. **25**, 375–387.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q. J. (**2002**). "Temporal processing and speech recognition in cochlear implant users," NeuroReport **13**, 1635–1639.

Fu, Q. J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Fu, Q. J., Shannon, R. V., and Galvin, J. J., III. (**2002**). "Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant," J. Acoust. Soc. Am. **112**, 1664–1674.

Glasberg, B. R., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.

Good, P. I. (**1994**). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses* (Springer-Verlag, New York).

Harnsberger, J. D., Svirsky, M. A., Kaiser, A. R., Pisoni, D. B., Wright, R., and Meyer, T. A. (**2001**). "Perceptual 'vowel spaces' of cochlear implant users: Implications for the study of auditory adaptation to spectral shift," J. Acoust. Soc. Am. **109**, 2135–2145.

Henry, B. A., and Turner, C. W. (**2003**). "The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 2861–2873.

Henry, B. A., Turner, C. W., and Behrens, A. (**2005**). "Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners," J. Acoust. Soc. Am. **118**, 1111–1121.

Horst, J. W. (**1987**). "Frequency discrimination of complex signals, frequency selectivity, and speech perception in hearing-impaired subjects," J. Acoust. Soc. Am. **82**, 874–885.

Irvine, D. R. F., Rajan, R., and McDermott, H. J. (**2000**). "Injury-induced reorganization in adult auditory cortex and its perceptual consequences," Hear. Res. **147**, 188–199.

Klatt, D. H. (**1980**). "Software for cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**(2), 467–477.

Litvak, L. M., Delgutte, B., and Eddington, D. K. (**2001**). "Auditory nerve fiber responses to electric stimulation: Modulated and unmodulated pulse trains," J. Acoust. Soc. Am. **110**, 368–379.

Loeb, G. E. (**2005**). "Are cochlear implant patients suffering from perceptual dissonance?," Ear Hear. **26**, 435–450.

Loeb, G. E., White, M. W., and Merzenich, M. M. (**1983**). "Spatial cross-correlation. A proposed mechanism for acoustic pitch perception," Biol. Cybern. **47**, 149–163.

Miller, G., and Nicely, P. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Ohde, R. N., and Abou-Khalil, R. (**2001**). "Age differences for stop conso-

nant and vowel perception in adults," J. Acoust. Soc. Am. **110**, 2156–2166.
Oppenheim, A. V., and Schafer, R. W. (**1975**). *Digital Signal Processing* (Prentice–Hall, Englewood Cliffs, N.J.).
Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.
Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). "Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants," J. Acoust. Soc. Am. **106**, 3629–3636.
Saoji, A., Litvak, L., Emadi, G., and Spahr, A. (**2005**). "Spectral modulation transfer function in cochlear implant listeners," in *Conference on Implantable Auditory Prostheses*, Asilomar, California.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Shannon, R. V., Zeng, F. G., and Wygonski, J. (**1998**). "Speech recognition with altered spectral distribution of envelope cues," J. Acoust. Soc. Am. **104**, 2467–2476.
Summers, V., and Leek, M. R. (**1994**). "The internal representation of spectral contrast in hearing-impaired listeners," J. Acoust. Soc. Am. **95**, 3518–3528.
Tyler, R. S., Preece, J. P., Lansing, C. R., Otto, S. R., and Gantz, B. J. (**1986**). "Previous experience as a confounding factor in comparing cochlear-implant processing schemes," J. Speech Hear. Res. **29**, 282–287.
Xu, L., Thompson, C. S., and Pfingst, B. E. (**2005**). "Relative contributions of spectral and temporal cues for phoneme recognition," J. Acoust. Soc. Am. **117**, 3255–3267.

Cortical responses to the $2f_1$-f_2 combination tone measured indirectly using magnetoencephalography

David W. Purcell,[a] Bernhard Ross (ß), Terence W. Picton, and Christo Pantev[b]
Rotman Research Institute at Baycrest, 3560 Bathurst Street, Toronto, Ontario, M6A 2E1, Canada

(Received 1 March 2007; revised 26 May 2007; accepted 29 May 2007)

The simultaneous presentation of two tones with frequencies f_1 and f_2 causes the perception of several combination tones in addition to the original tones. The most prominent of these are at frequencies f_2-f_1 and $2f_1$-f_2. This study measured human physiological responses to the $2f_1$-f_2 combination tone at 500 Hz caused by tones of 750 and 1000 Hz with intensities of 65 and 55 dB SPL, respectively. Responses were measured from the cochlea using the distortion product otoacoustic emission (DPOAE), and from the auditory cortex using the 40-Hz steady-state magnetoencephalographic (MEG) response. The perceptual response was assessed by having the participant adjust a probe tone to cause maximal beating ("best-beats") with the perceived combination tone. The cortical response to the combination tone was evaluated in two ways: first by presenting a probe tone with a frequency of 460 Hz at the perceptual best-beats level, resulting in a 40-Hz response because of interaction with the combination tone at 500 Hz, and second by simultaneously presenting two f_1 and f_2 pairs that caused combination tones that would themselves beat at 40 Hz. The $2f_1$-f_2 DPOAE in the external auditory canal had a level of 2.6 (s.d. 12.1) dB SPL. The 40-Hz MEG response in the contralateral cortex had a magnitude of 0.39 (s.d. 0.1) nA m. The perceived level of the combination tone was 44.8 (s.d. 11.3) dB SPL. There were no significant correlations between these measurements. These results indicate that physiological responses to the $2f_1$-f_2 combination tone occur in the human auditory system all the way from the cochlea to the primary auditory cortex. The perceived magnitude of the combination tone is not determined by the measured physiological response at either the cochlea or the cortex. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2751250]

PACS number(s): 43.64.Ri, 43.64.Jb, 43.64.Qh, 43.66.Ki [BLM] Pages: 992–1003

I. INTRODUCTION

The mammalian cochlea contains nonlinearities that generate frequencies not present in the stimulus. The simplest acoustic stimulus that can be used to study this phenomena is composed of two pure tones with frequencies f_1 and f_2, where $f_2 > f_1$. In response, the cochlea produces a set of combination tones, or distortion products, of which the cubic $2f_1$-f_2 and the quadratic f_2-f_1 are the most prominent and most studied.

With favorable stimulus conditions, combination tones are audible. The Italian violinist Tartini knew of their existence as early as 1714 (Jones, 1935). As Jones (1935) reports, by 1801 it was known that the tone produced by a 32-ft pipe in a cathedral organ (about 16 Hz) could be generated in the ears of listeners by simultaneously activating a 16-ft pipe (about 32 Hz) and an even shorter pipe producing its fifth (about 48 Hz) through the $2f_1$-f_2 combination tone, thus dispensing with the need for 32-ft pipes.

The $2f_1$-f_2 combination tone has been extensively studied psychoacoustically (e.g., Zwicker, 1955; Plomp, 1965; Goldstein, 1966; Smoorenburg, 1972a, b). The $2f_1$-f_2 nonlinearity is termed "essential" because it occurs at low stimulus levels (Goldstein, 1966). This is unlike a mainly linear system distorting only at higher levels (von Helmholtz, 1877/1954; Goldstein, 1966). The level and phase of the $2f_1$-f_2 tone within the cochlea can be estimated by presenting an additional "cancellation" tone of the same frequency (Zwicker, 1955; Goldstein, 1966, 1969). Perception of the combination tone can be eliminated by carefully adjusting the magnitude and phase of the cancellation tone until nothing is heard. Goldstein (1966) facilitated the subject's ability to hear the cancellation effect by adding an additional probe tone slightly mistuned from the cancellation tone (e.g., 4 Hz), of sufficient level to beat with the combination tone. The cancellation tone was then adjusted to eliminate the beating sensation by canceling the combination tone. Using the cancellation technique, the cochlear $2f_1$-f_2 combination tone has an apparent level of an externally supplied tone that is 20–40 dB below that of equal level primaries in the external ear canal (where tone pair stimuli f_2 and f_1 have equal levels $L_2 = L_1$; Goldstein, 1966; Smoorenburg, 1972b; Hall, 1975). Cancellation methods may overestimate the level of the combination tone for small f_2/f_1 due to suppression effects of the stimulus f_1 (Smoorenburg, 1972b; Shannon and Houtgast, 1980; Zwicker, 1981).

The psychoacoustically estimated levels of the combination tones with respect to the stimuli (i.e., 20–40 dB below the primaries) are similar to the levels of distortion products

[a]Author to whom correspondence should be addressed; Electronic mail: purcelld@nca.uwo.ca. Currently affiliated with the National Centre for Audiology at the University of Western Ontario, London, Ontario, N6G 1H1, Canada.
[b]Currently affiliated with the Institute for Biomagnetism and Biosignalanalysis at the University of Münster, Münster, Germany.

measured in the mechanical motion of the basilar membrane (Robles *et al.*, 1997), the cochlear microphonic (Gibian and Kim, 1982), receptor potentials of IHCs (Nuttall and Dolan, 1990), and single fibers of the auditory nerve (Goldstein and Kiang, 1968; Buunen and Rhode, 1978; Kim *et al.*, 1980). The response can be eliminated from recordings of nerve fibers using cancellation techniques (Goldstein and Kiang, 1968; Goldstein *et al.*, 1978). These results indicated that the $2f_1$-f_2 combination tone is present in the cochlea, and suggested that it was processed at the same characteristic place as an externally supplied tone of equal frequency. The $2f_1$-f_2 combination tone was later observed directly in the mechanical motion of the basilar membrane at its characteristic place (Robles *et al.*, 1991, 1997).

The distortion product otoacoustic emission (DPOAE) at frequency $2f_1$-f_2 is initially generated at the f_2 characteristic place where the two stimulus tones interact most prominently (Talmadge *et al.*, 1999; Shera and Guinan, 1999; Knight and Kemp, 2000). Signals at the DPOAE frequency then propagate on the basilar membrane both basally and apically. The apical propagation will reach the characteristic place for frequencies equal to $2f_1$-f_2. The DPOAE measured in the ear canal can be dominated by energy from either the f_2 or the $2f_1$-f_2 place depending on the stimulus parameters (Shera and Guinan, 1999; Knight and Kemp, 2000, 2001; Dhar *et al.*, 2005). Wilson (1980) found significant discrepancies between the levels of the ear canal DPOAEs and psychoacoustic cancellation levels in the same subjects. These may have been in part due to changes in dominant DPOAE sources across stimulus conditions. However, Furst *et al.* (1988) and Zwicker and Harris (1990) also reported that DPOAE levels in the canal were low compared to those estimated psychoacoustically in the same subjects. Furthermore, the finding that psychoacoustic cancellation does not eliminate the ear canal DPOAE highlights the fact that different processes are involved in the generation of the DPOAE and the perception of the combination tone (Furst *et al.*, 1988; Zwicker and Harris, 1990). The net DPOAE signal reaching the canal is a function of the relative contributions of its two sources and the reverse transmission characteristics of the basilar membrane and middle ear. The perceptual response to the combination tone likely relates to a different mixture of the two sources.

Correlates of the $2f_1$-f_2 combination tone have been measured at most levels of the auditory system, including the ear canal (e.g., Kemp and Brown, 1984; Probst *et al.*, 1991), basilar membrane (Robles *et al.*, 1991, 1997), cochlear microphonic (Gibian and Kim, 1982), inner hair cell [(IHC); Nuttall and Dolan, 1990; Cheatham and Dallos, 1997], auditory nerve (Kim *et al.*, 1980), and auditory brainstem (Rickman *et al.*, 1991; Pandya and Krishnan, 2004). The purpose of this study was to determine whether correlates of the $2f_1$-f_2 combination tone could be measured in the human auditory cortex using magnetoencephalography (MEG).

The most serious difficulty in obtaining the cortical response to a $2f_1$-f_2 combination tone is separation of the $2f_1$-f_2 response from the response to the stimulus tones that generate it. Since the stimulus tones have 20–40 dB higher intensity, and are present simultaneously with the response, it is impossible to use time-domain measures such as the evoked N1 or sustained response. In contrast, steady-state frequency-domain measurements of auditory evoked potentials and magnetic fields can separate the responses to the stimuli and to the combination tones because they are at different frequencies. Unfortunately, steady-state responses above 100 Hz derive mainly from brainstem sources (Herdman *et al.*, 2002; Purcell *et al.*, 2004), and combination tones are most relevant at frequencies higher than 100 Hz. Steady-state measurements cannot therefore directly measure the response to the $2f_1$-f_2 combination tone in cortex. To overcome these constraints, an approach based on beating used in psychoacoustics was adopted (Goldstein, 1966; Furst *et al.*, 1988, unpublished).. Two pure tones that are similar in frequency and magnitude will cause beating at the difference frequency. Therefore, cortical correlates of the $2f_1$-f_2 combination tone could be measured indirectly by causing a 40-Hz beat either between a combination tone and an externally presented probe tone, or between two combination tones. The sensation of beating between a $2f_1$-f_2 combination tone and an externally supplied tone can be perceived relatively easily by naïve participants with little training. The 40-Hz steady-state magnetic fields generated in the auditory cortices (Mäkelä and Hari, 1987; Pantev *et al.*, 1996; Gutschalk *et al.*, 1999; Schoonhoven *et al.*, 2003) by these beats can be measured using MEG.

This indirect method of obtaining correlates of the $2f_1$-f_2 combination tone measures physiological responses to the difference-frequency type combination tone or "beat." Cochlear microphonic (Gibian and Kim, 1982; Cheatham and Dallos, 1997), IHC (Cheatham and Dallos, 1997), and single neuron studies (Smoorenburg *et al.*, 1976; Kim *et al.*, 1980) have shown that both cubic and quadratic combination tones travel as waves on the basilar membrane from their initiation place where the stimuli interact at their own characteristic places. They are also present in nerve fiber recordings both near the stimulus interaction site and at their own characteristic places (Kim *et al.*, 1980). Dolphin and Mountain (1993; Dolphin, 1997) have shown however that the quadratic combination tone recorded from the scalp as an auditory evoked potential is predominantly from the place of the interaction of the stimuli. Robles *et al.* (1997; comment in Cheatham and Dallos, 1997) also found no evidence to support a basilar membrane origin of the quadratic combination tone. Half-wave rectification in the IHC and auditory nerve can produce the quadratic evoked potential (Lins *et al.*, 1995).

The present paper reports two experiments: the first used an external tone to beat with a $2f_1$-f_2 combination tone, and the second studied the beating together of two separate $2f_1$-f_2 combination tones. The beating frequency was set near 40 Hz since the cortical MEG response is easy to measure at this frequency (Ross *et al.*, 2000). In both experiments, the $2f_1$-f_2 combination tone is probably generated near the f_2 region of the basilar membrane. The $2f_1$-f_2 combination tone then propagates to its characteristic place where it could interact in Experiment 1 with the externally supplied probe tone, or in Experiment 2 with a second $2f_1$-f_2 combination tone of different frequency. However, interac-

tions between the combination tone and the probe tone (or between the two combination tones) could also take place at higher levels of the auditory pathway to produce the 40-Hz beat. The place of interaction was not the focus of this study. Rather the purpose was to observe the MEG correlates of the combination tone $2f_1$-f_2 in the human auditory cortex.

II. METHODS

A. Subjects

Subjects reported normal hearing and were screened for thresholds below 20 dB HL at frequencies 500, 750, and 1000 Hz, which are in the range of stimuli used in this experiment. For the first experiment (n=11, 2 males), the participants varied in age from 18 to 47 years. In the second experiment (n=2, 2 males), the ages were 24 and 33 years. The experiments were approved by the Research Ethics Board of the Baycrest Centre and written informed consent was obtained from each participant after the nature of the study was explained in accordance with the principles of the Declaration of Helsinki.

B. Experimental conditions

In Experiment 1, a stimulus tone pair of f_2=1000 Hz and f_1=750 Hz was used to elicit a combination tone of 500 Hz. An additional probe tone of similar frequency was used to produce beating with the combination tone. During the first part of the experiment, the participant adjusted the level of a probe tone at approximately 495.5 Hz to produce the strongest psychoacoustic perception of a slow beat at 4.5 Hz (Yost, 2000, 167 pp.). This beating rate was much easier to hear than at rates near 40 Hz, where it is perceived as "roughness" rather than "beats" (Fastl, 1977, 1990). This level of probe tone was subsequently used in the MEG measurement, but the frequency was lowered to 460 Hz to produce a 40-Hz beat [see Fig. 1(a)]. Frequencies near 40 Hz elicit a larger and stable MEG response, and correspondingly larger signal to noise ratio (SNR), than frequencies near 5 Hz (Ross *et al.*, 2000). A significant 40-Hz response in the MEG signal would show that the probe tone produced beats with the $2f_1$-f_2 combination tone. In pilot MEG measurements, a 4-Hz beat was detectable in one participant, but the SNR was substantially less favorable than at 40 Hz. We therefore decided to employ the beat frequency that was best for each measurement type (4.5 Hz for psychoacoustics and 40 Hz for MEG).

The psychoacoustic measurement of best-beats was used to estimate the approximate magnitude of the $2f_1$-f_2 combination tone so that the depth of modulation and corresponding response could be maximized both perceptually (Zwislocki, 1953) and in the MEG record. After some brief training, the participants were able to perceive the beats and to adjust the level of the probe tone to give a sensation of best-beats. The change in beat frequency between the psychoacoustic and MEG measurements (from 4.5 to 40 Hz) should not have affected the modulation depth of the beating.

Experiment 2 also used beating to show the presence of a $2f_1$-f_2 combination tone, but did not use an external probe tone. Rather, the stimuli were two tone pairs with frequencies

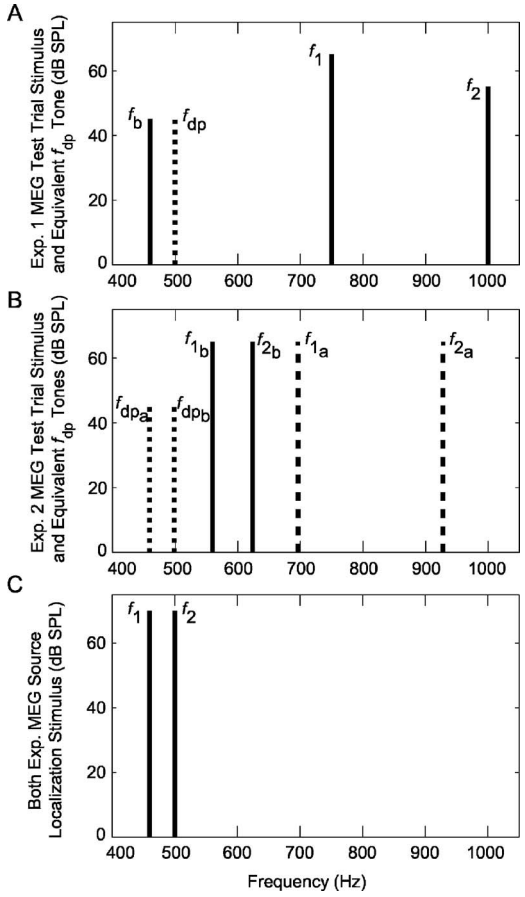


FIG. 1. Magnitude spectra of the stimuli used in the experiments. (A) The stimulus tone pair (f_1 and f_2) and the probe tone f_b used in Experiment 1. A hypothetical combination tone within the auditory system is shown with the dash line labeled f_{dp}. (B) The two tone pairs used in Experiment 2. The higher frequency tone pair is marked with dash lines labeled f_{1a} and f_{2a}. The lower frequency tone pair is marked with solid lines labeled f_{1b} and f_{2b}. The leftmost two dashed lines represent hypothetical combination tones elicited by the two tone pairs. The line labeled f_{dpa} would be elicited by the higher frequency tone pair, and the line labeled f_{dpb} by the lower frequency tone pair. (C) The ECD stimulus tone pair used in both experiments to elicit robust responses for the purpose of dipole fitting.

of f_{2a}=931 Hz and f_{1a}=696 Hz, and f_{2b}=626 Hz and f_{1b}=563 Hz [see Fig. 1(b)]. These stimuli were designed to elicit two combination tones at frequencies f_{dpa}=460 Hz and f_{dpb}=500 Hz. These two combination tones would beat at 40 Hz, provided their relative levels were similar. The beating could then elicit an MEG response if the absolute levels of the two combination tones were sufficiently high.

C. Stimulus generation

1. Psychoacoustic and DPOAE stimuli

Auditory stimuli for the psychoacoustic and DPOAE measurements were generated with 16-bit precision at a rate of 32 kHz using a National Instruments 6052E input/output board. A Grason Stadler Model 16 audiometer was used to adjust the magnitude of the electrical stimulus signals f_2 and f_1 prior to transduction into acoustic stimuli by a pair of Etymotic ER-2 transducers coupled to an Etymotic ER-10B+OAE probe. Stimulus tone f_2 was generated with one channel, and f_1 with the other to minimize acoustic distortion within the stimulus system. The probe tone near the

DPOAE frequency f_{dp} was generated electrically with a Stanford Research Systems signal generator, and its level was controlled by the participant using a Madsen Micro5 audiometer. The probe tone was produced acoustically using an Etymotic ER-3A transducer, and mixed acoustically with the f_2 stimulus tone in a small coupler just prior to one of the OAE probe's inlet tubes. Tones were calibrated at every frequency of interest with the OAE probe sealed in a Knowles DB-100 Zwislocki coupler (Siegel, 1994; Whitehead *et al.* 1995). The stimuli were presented continuously to the left ear at levels $L_2=55$ and $L_1=65$ dB SPL.

2. MEG stimuli

Auditory stimuli were generated with 16-bit precision at 11.025 kHz by a Sound Blaster AWE 64 card in a control computer linked to the MEG acquisition system. Magnitude of the electrical signals f_2 and f_1 were controlled with a Madsen OB822 audiometer prior to transduction to acoustic signals with a pair of Etymotic ER-2 transducers. These transducers delivered the sound to a small coupler at the left ear through approximately 2.92 m of PVC tubing. The left ear was chosen for the experiment because the 40-Hz response shows a right hemispheric and contralateral dominance (Ross *et al.*, 2005).

In the first experiment, f_2 was generated on its own channel, whereas the f_1 channel also produced the probe tone during test trials. For the second experiment, the higher tones of each pair (f_{2a} and f_{2b}) were produced on one channel and the lower tones on another (f_{1a} and f_{1b}). Signals were calibrated using the Zwislocki coupler, and no significant distortions were present at the response frequencies.

In both MEG experiments tones were presented to the left ear for trials of duration 2.3 s. Presenting limited-duration trials during MEG measurements allowed intermingling of control and test trials. This was intended to minimize time and motion confounds from the MEG data within individuals. These trials were sufficiently long to obtain a steady-state response and to perform spectral analysis.

Both MEG experiments had three stimulus conditions: *test trials, control trials*, and *reference trials*. The test trials in Experiment 1 used the acoustic stimulus shown in Fig. 1(a), where the stimuli, f_2 and f_1, as well as the probe tone f_b were presented (the dotted line labeled f_{dp} is an example of a potential combination tone generated in the cochlea). Control trials included the stimuli f_2 and f_1, but no probe tone. The 40-Hz beat was expected in the MEG test trial measurements since the $2f_1$-f_2 combination tone should generate beats with the probe tone. This response will be referred to as probe beating (combination tone with probe tone). In the control trials, there was no signal present to beat at 40 Hz with the $2f_1$-f_2 combination tone.

In Experiment 2, the test trial stimulus included two tone pairs shown in Fig. 1(b), where stimulus tones f_{2a} and f_{1a} were expected to produce the combination tone f_{dpa} in the cochlea. Similarly, the stimulus tones f_{2b} and f_{1b} produce the combination tone f_{dpb}. During test trials, if the combination tones f_{dpa} and f_{dpb} were of similar magnitude, they could beat with one another to generate a 40-Hz response in the MEG record. This response will be referred to as CT beating (combination tone with combination tone). In control trials, only the stimulus tones f_{2a} and f_{1a} were presented, so no MEG response at 40 Hz was expected.

The reference trials, used in both experiments, were designed to produce a robust 40-Hz MEG response to determine an equivalent current dipole [(ECD), Ross *et al.*, 2000, 2002] in each hemisphere for the 40-Hz steady-state responses. As shown in Fig. 1(c), two tones of $f_2=500$ and $f_1=460$ Hz were presented at the relatively high level of 70 dB SPL. These tones were the same frequencies as the combination and probe tones used in the test trials, but they were about 25 dB higher in level than the average level of the combination tones estimated psychoacoustically. This stimulus condition produced a relatively large 40-Hz magnetic field for the purpose of finding the equivalent cortical source (a single ECD) in each hemisphere. Henceforth this stimulus will be referred to as the ECD stimulus.

D. Recording and analysis

1. DPOAE

Sound pressure in the ear canal was measured for 61.44 s using the microphones integrated into the Etymotic acoustic probe. The microphone signal was amplified using the +40-dB setting on the supplied Etymotic amplifier, and bandpass filtered 40–4000 Hz using a Krohn-Hite Model 3322 filter with a 24-dB/octave slope. The conditioned microphone signal was digitized at 16 kHz with 16-bit precision. A discrete Fourier transform (DFT) was performed on the 61.44-s window, and the frequency bin containing f_{dp} was evaluated to determine whether the emission magnitude was significantly larger than the noise in six frequency bins on either side using an F-test with 2 and 24 degrees of freedom (John and Picton, 2000).

2. MEG recording and analysis

The MEG was measured by means of a whole head device with 151 sensors (Omega-151a, VSM Medtech) sampled at 625 Hz with a third-order gradient noise reduction. Trials with muscle artifacts on any channel due to events such as eye or jaw movements (approximately 20% of total) were removed from further analysis. The signal in each trial was referred to a baseline value equal to the mean in a 50-ms window immediately prior to the stimulus onset. For each participant and stimulus condition, trials were averaged together and the average trials were bandpass filtered within the range 20–60 Hz.

The estimated ECDs were localized in a Cartesian coordinate system with the y axis between the left and right external ear canals, the x axis from the center of this axis to the nasion, and the vertical z axis orthogonal to the others. A spherical head model of radius 7.5 cm with the head center at coordinates $x=0$, $y=0$, and $z=5$ cm was employed to determine the locations of the ECDs.

The average trial elicited with the ECD stimulus [Fig. 1(c)] was subsequently fit with a model sinusoid of 40 Hz on all channels in each hemisphere. The field of a single dipole was then fit in each hemisphere for this 40-Hz model data using the spherical head model. Although in most partici-

pants the dipole in each hemisphere was independently located, for two individuals the smaller dipole in the left hemisphere had to be fixed symmetrical to the one in the right hemisphere to avoid an unrealistic solution. Both dipoles were then fixed in position and orientation, and assumed to be the location of the sources of any 40-Hz steady-state responses in the test and control conditions.

Average trial data from the test and control conditions [elicited with the stimuli in Fig. 1(a) for Experiment 1, and Fig. 1(b) for Experiment 2] were used to determine current dipole moments over time at the single dipole fixed in each hemisphere using the method of source space projection (Ross *et al.*, 2000, 2002). These dipole moments were then evaluated for the presence of a 40-Hz beat in the frequency domain using a DFT. Spectral analysis was performed from 250-ms poststimulus onset to the end of the trial such that there were an integer number of cycles of 40 Hz in the analysis window (total length 2.05 s). The initial part of the source wave form was excluded because it takes about 250 ms for the steady-state response to become well established (Ross *et al.*, 2002). Signal magnitude at 40 Hz was compared, as for the DPOAE, to noise in six bins above and six below 40 Hz using an F-test with 2 and 24 degrees of freedom.

E. Experimental protocol

1. Psychoacoustic and DPOAE

During the first part of Experiment 1 when psychoacoustic and DPOAE measurements were performed, participants were seated in an Industrial Acoustics Company (IAC) sound insulated room. An acoustic probe was sealed in the left ear canal using an appropriately sized rubber tip. Participants were familiarized with controlling the level of the 495.5-Hz probe tone using an audiometer, and the sensation characteristics of a 4.5-Hz beat were demonstrated using example tones. During the actual psychoacoustic adjustment, stimulus tones f_2 and f_1 were presented to the left ear at $L_2=55$ and $L_1=65$ dB SPL, respectively. These relative levels were chosen to maximize the magnitude of DPOAEs (Gaskill and Brown, 1990). Each participant started with a low probe tone level, and incremented it in 5-dB steps until the level for best-beats had been surpassed. The participant then lowered the probe tone level and fine-tuned it in 1-dB steps to maximize the sensation of beating. Finally, the participant interrupted the probe tone to demonstrate that the beating sensation disappeared in the absence of the probe. The best-beats level of the probe tone was recorded for subsequent use in the MEG protocol. After the psychoacoustic assessment was finished, physiological measurements of the individual's DPOAE at 500 Hz were made for 61.44 s each, alone and with probe tones of either 495.5 and 460 Hz.

2. MEG

The second part of Experiment 1 took place in the magnetically shielded room of the MEG. The subject sat upright with their head touching the top of the dewar's cavity. Head localization coils were located on inserts placed in both ears, as well as at the nasion. Participants watched a silent subtitled movie during the 1-h measurement session and were instructed to remain as still as possible during the presentation of tones. During each of five blocks lasting 10 min, 230 randomly intermingled control and test trials were collected, each of 2.3 s with 150-ms interstimulus intervals. Between blocks the subject could relax, but was instructed to do their best not to change their head position for the course of the experiment. The stimulus tone pair from the first part of the experiment (same frequencies and levels) was presented to the left ear for all trials. For test trials, the 460-Hz probe tone was also presented at the best-beat level. A final block of 230 trials was collected with the ECD stimulus at 70 dB SPL for the purpose of dipole fitting. Head localization was performed at the start and end of each block of trials.

Experiment 2 was similar to Experiment 1 except that the test trials used two tone pairs rather than a tone pair and a probe. The first five blocks contained randomized control and test trials, where either the upper tone pair (control), or both tone pairs were presented (test). The tones were each presented at 65 dB SPL. The same final block used the ECD stimulus again at 70 dB SPL, as in Experiment 1.

TABLE I. Mean right and left hemisphere dipole sources determined from the MEG stimulus designed to elicit a robust 40-Hz steady-state response where tones were $f_2=500$ Hz and $f_1=460$ Hz presented at 70 dB SPL. Data are from (eight) participants in experiment 1 and the two individuals from Experiment 2. Dipole source coordinates are with respect to the origin of the spherical head model. In this model the center of the head was presumed at $x=0$, $y=0$, $z=5$ cm with sphere radius 7.5 cm. The x-z plane is sagittal with positive x toward the nose and positive z upwards. The y-z plane is coronal with positive y toward the individual's left ear. The x-y plane is axial. Standard deviation given in parentheses.

	Left hemisphere 40-Hz steady-state dipole	Right hemisphere 40-Hz steady-state dipole
x (cm)	1.4 (0.9)	2.0 (0.8)
y (cm)	4.1 (1.0)	−3.8 (0.5)
z (cm)	7.0 (0.7)	6.9 (0.6)
Magnitude Q (nA m)	1.5 (1.2)	2.3 (1.2)

III. RESULTS

A. Experiment 1: Beats between combination tone and probe tone

One of the 11 participants tested did not show a significant 40-Hz beat response in the test condition of the MEG measurement for either hemisphere. This person did not have atypical DPOAE magnitude (−6.9 dB SPL), or did he select an unusual best-beat probe tone level (42.7 dB SPL). In response to the ECD stimulus, he had dipole moment magnitudes larger than the average values in Table I, but well within the range found for other participants (left hemisphere $Q=2.6$ nA m, right hemisphere $Q=2.8$ nA m). However, MEG noise levels during the test condition were the highest of any participant (beyond the mean plus 2 s.d. of the noise levels estimated for the others), and this was likely the reason for his lack of a detectable MEG response. He was excluded from further analysis.

With two other participants, the experimental protocol had to be altered to suit their needs. In one person, the best-

TABLE II. Mean responses across eight participants from individual average test and control trials in Experiment 1. The three entries for DPOAE magnitude are without probe tone, with 495.5-Hz probe, and with 460-Hz probe, respectively. The three entries for best-beat probe level are as measured in the Zwislocki coupler with the 495.5-Hz probe tone, as measured in the ear canal with the 495.5-Hz probe, and as measured in the ear canal with the 460-Hz probe, respectively.

	Sound booth			MEG left hemisphere		MEG right hemisphere	
	Probe tone frequency (Hz)	DPOAE magnitude (dB SPL)	Best-beat probe level (dB SPL)	Source space projection magnitude (nA m)	Noise estimate (nA m)	source space projection magnitude (nA m)	Noise estimate (nA m)
Test trial mean (s.d.)	No probe 495.5 460	2.6 (12.1) 1.1 (12.3) 1.7 (12.1)	44.8 (11.3) 46.2 (12.8) 44.7 (13.0)	0.28 (0.14)	0.09 (0.03)	0.39 (0.10)	0.07 (0.02)
Control trial mean (s.d.)	⋯	⋯	⋯	0.05 (0.03)	0.07 (0.03)	0.07 (0.04)	0.06 (0.01)

beat probe tone level of 74 dB SPL was too high to be used with the MEG stimulus presentation system. A second person objected to the loudness of the tones in the MEG measurement, despite the same level having been used to measure DPOAEs. For both of these individuals, the MEG measurement proceeded with lower stimulus levels (10 and 5 dB lower, respectively). Significant 40-Hz beat responses were found for both individuals during the test condition (only the right hemisphere for the first, and both hemispheres for the second person). However, since the protocol was changed for these individuals, they were also excluded from further analysis.

All eight of the remaining participants in Experiment 1 had right hemisphere source space projection dipole moment magnitudes, elicited by the 40-Hz probe beat during the test condition, that were statistically larger (using the F-test) than the noise at nearby frequencies. Seven of the eight participants also had statistically significant sources in the left hemisphere for the test condition. The dipole sources in the left and right hemispheres that were fit to the ECD stimulus (and then fixed in position and orientation) are shown in Table I. These dipoles were used to estimate the source space projection dipole moment magnitudes reported in Table II. The data shown in Table I are the average of the eight individuals who had significant responses in the test condition and followed the protocol of Experiment 1, and the two individuals in Experiment 2 since the same ECD stimulus was used to estimate these dipoles. Using paired t-tests, there were significant differences between some of the dipole parameters for the left and right hemispheres. For the y-axis parameter, the absolute value was used to compare the two hemispheres. The right source was significantly ($p<0.01$) more anterior than the left. Most strikingly, the response was much larger in the right hemisphere ($p<0.0001$). Every participant had a larger dipole moment magnitude in the right hemisphere compared to his or her own left hemisphere.

Table II shows the DPOAE, psychoacoustic, and MEG data from the eight participants who had significant MEG responses in the right hemisphere for test trials ($p<0.001$) in Experiment 1, and followed the same standard experimental protocol. All but one of these participants also had significant responses in the left hemisphere at $p<0.05$; the remaining individual was near significant at $p<0.09$. As shown in the third and fourth columns of Table II, there was substantial variability in both the DPOAE and the level of the probe tone chosen to generate best-beats. The probe tone level was significantly larger than the measured DPOAE level using a two-sample separate variance t-test ($p<0.001$). DPOAEs were also measured with the probe tones of 495.5 and 460 Hz present. All three DPOAE measurements are presented in Table II (no probe, 495.5-Hz probe, and 460-Hz probe). Although the means are slightly lower, paired t-tests showed no significant changes in DPOAE magnitude with a probe tone present. Figure 2 plots the ear canal spectra from a single individual for all three probe conditions. Although

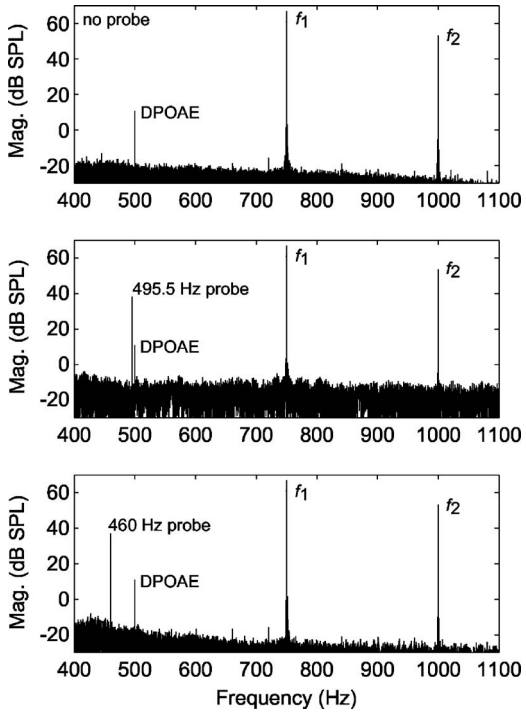


FIG. 2. Magnitude spectra of the ear canal sound pressure from a single individual in Experiment 1. The three panels show data from the three probe conditions where there was no probe tone, the probe was 4.5 Hz below the DPOAE frequency, and the probe tone was 40 Hz below the DPOAE frequency, respectively. The noise floor fluctuates between the measurements, but the DPOAE magnitude is unchanged.

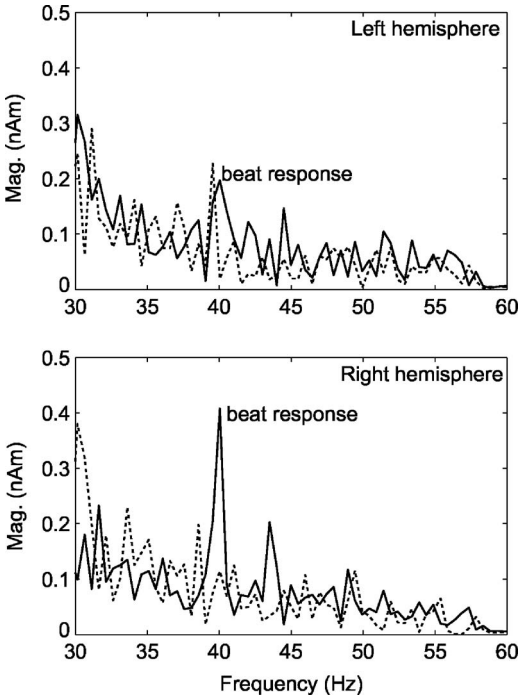


FIG. 3. Magnitude spectra of the source space projection dipole moments from each hemisphere calculated from the average of the test trials (solid line) and control trials (dash line) for the individual from Fig. 2. Note the spurious peak in the spectrum of the average control trial in the left hemisphere at 39 Hz. The left hemisphere response at 40 Hz in the average test trial only achieved a significance level of $p<0.04$.

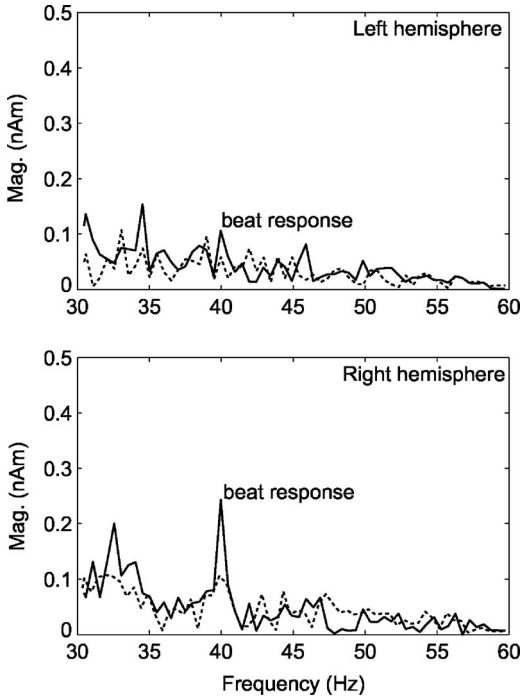


FIG. 4. Magnitude spectra of the source space projection dipole moments from each hemisphere calculated from the average of the test trials (solid line) and control trials (dashed line) for the individual from Experiment 2 with a significant response.

the noise floor fluctuates from one recording to the next, the stimuli and DPOAE do not change levels between panels.

The last four columns in Table II show the MEG source space projection dipole moment magnitudes and noise estimates for the left and right hemispheres. Using paired t-tests for both hemispheres, the source space projection dipole moment magnitude was significantly larger in the test condition than in the control (no probe) condition (left $p<0.002$; right $p<0.0001$). Although the mean source space projection dipole moment magnitude for the right hemisphere in the test condition is larger than for the left, the difference was not quite statistically significant (paired t-test $p<0.07$; sign test $p<0.07$). The right source space projection dipole moment magnitude was larger in all participants except one. The probe beat responses are shown in Fig. 3 for the same individual whose DPOAE data are plotted in Fig. 2. The 40-Hz response in the left hemisphere (ipsilateral to the left stimulus ear) was half the magnitude of the signal in the right hemisphere, although both were statistically different from the background noise ($p<0.04$ and $p<0.00001$, respectively).

Pearson's r coefficients were calculated between measurement types across the eight participants from Experiment 1. There were no significant correlations between DPOAE magnitude, best-beat probe tone level, or MEG source space projection magnitudes in either hemisphere.

B. Experiment 2: Beats between two combination tones

The 40-Hz steady-state dipoles determined from the ECD stimulus for the two participants in Experiment 2 were similar to those for the eight subjects used in Experiment 1 and are included in the average data of Table I. In the test condition, one participant had a robust 40-Hz CT beat response in the right hemisphere [$Q=0.245$ nA m, noise estimate 0.053 (s.d.=0.026) nA m, $p<0.0001$], and a small response in the left hemisphere [$Q=0.106$ nA m, noise estimate 0.043 (s.d.=0.022) nA m, $p<0.02$]. The right hemisphere response in the test condition showed a clear spike relative to noise in nearby frequencies as shown in Fig. 4. For this individual during the control condition, there was no significant response for the left hemisphere [$Q=0.058$ nA m, noise estimate 0.047 (s.d.=0.022) nA m, $p<0.3$] and a borderline response in the right hemisphere [$Q=0.105$ nA m, noise estimate 0.048 (s.d.=0.027) nA m, $p<0.04$] that was likely a type I statistical error. The second participant did not show any significant responses in either test or control conditions. Both participants had clear responses to the ECD stimulus (listed for left and right hemispheres: 0.84 and 1.89 nA m for the participant with a CT beat, and 0.92 and 1.67 nA m for the one without).

IV. DISCUSSION

All participants in Experiment 1 had measurable $2f_1$-f_2 DPOAEs that were statistically different from the background noise in the ear canal when the presented frequencies were 750 and 1000 Hz, and the corresponding DPOAE was 500 Hz. All individuals were able to perceive beating at 4.5 Hz between an externally supplied probe tone and the $2f_1$-f_2 combination tone and were able to adjust the level of the probe tone to maximize this sensation and achieve best-beats. When the frequency of the probe tone was changed so

TABLE III. Comparisons between average DPOAE magnitude and standard deviation (s.d.) obtained in the present study, and from the literature using similar stimulus parameters (N/A denotes value not explicitly given).

Study	f_2 (Hz)	f_2/f_1	L_1 (dB SPL)	L_1-L_2 (dB)	Participants (ears)	L_{dp} (s.d.) (dB SPL)
Present	1000	1.33	65	10	8 (8)	2.6 (12.1)
Harris *et al.* (1998; Fig. 2)	2000	1.33	65	0	5 (10)	−2 (6)
Abdala (1996; Fig. 5a)	1500	1.353	60	10	10 (11)	−3 (6)
Brown *et al.* (2000; Fig. 2)	1700	1.3	60	15	40 (40)	−7 (N/A)
Gorga *et al.* (1993; Fig. 3)	750	1.2	65	15	80 (80)	0 (8)
Vinck *et al.* (1996; Fig. 1, Table 4)	830	1.213	60	0	101 (101)	3.6 (6)

that the beating was at 40 Hz, all but one individual had robust 40-Hz MEG responses in the right hemisphere contralateral to the stimulus. Most participants also had statistically significant responses in the left hemisphere. Experiment 2 looked at whether 40-Hz beating responses could be obtained between two combination tones, each produced from a pair of externally presented stimuli. One of two participants showed a 40-Hz MEG response with good SNR in the right hemisphere. The 40-Hz steady-state MEG signals recorded in these two experiments objectively demonstrate activation of the human primary auditory cortex in response to $2f_1$-f_2 combination tones.

A. Psychoacoustic and DPOAE response

The absolute level of the average DPOAE recorded in the ear canal (2.6 dB SPL) is typical for human subjects at wide stimulus frequency ratios and low emission frequencies (as summarized in Table III). The s.d. reported here (12 dB) is larger than previously reported values (6–8 dB). This may be in part because we were able to include individuals with low-level emissions due to the long recording time (61 s). Such individuals may be excluded from studies with short, clinically oriented measurements due to higher noise floors.

At the stimulus frequencies used in these experiments (ratio $f_2/f_1=1.33$), it is likely that a single dominant DPOAE source is at the f_2 characteristic place where the stimuli interact most prevalently. A second source at the $2f_1$-f_2 characteristic place can dominate the DPOAE measured in the canal, but normally only when the f_2/f_1 ratio is smaller (Shera and Guinan, 1999; Knight and Kemp, 2000, 2001; Dhar *et al.*, 2005). In this experiment, the $2f_1$-f_2 combination tone is therefore presumably initiated where the stimuli interact at the f_2 characteristic place. Some energy at frequency f_{dp} then propagates basally to cross the middle ear and be recorded in the ear canal as the DPOAE. Some energy also propagates apically to the $2f_1$-f_2 characteristic place where the basilar membrane responds as it would to an externally supplied tone.

The participants in the present study could reliably estimate the level of the combination tone using best-beats. The average best-beat probe tone level (45 dB SPL) was 20 dB below the level of the f_1 stimulus tone, which is in the range of that reported previously (Goldstein, 1966; Smoorenburg, 1972b; Hall, 1975). The best-beat level of the probe tone was more than 40 dB higher than the ear canal measurement of the DPOAE at frequency $f_{dp}=2f_1$-f_2. This is consistent with the within-subject results of Zwicker and Harris (1990) who found the cancellation tone for the perceptual combination tone to be 33–60 dB higher than for the ear canal DPOAE, depending on the individual and stimulus conditions.

The best-beat probe tone is similar in frequency to f_{dp}, and would itself excite the basilar membrane near the f_{dp} characteristic place. To generate a beating sensation, the probe tone amplitude must be similar to the amplitude of the combination tone somewhere in the system in order to achieve the most prominently perceived amplitude modulation. Depth of modulation thresholds, from temporal modulation transfer functions, suggest that modulation would be impossible to detect if the probe and combination tones differed by more than 30 dB (Kohlrausch *et al.*, 2000; Strickland, 2000). Therefore, it is unlikely that the DPOAE as measured in the ear canal was causing the perceived beat.

At least two possibilities could explain the discrepancy in levels between the probe tone used to estimate the perceived combination tone and the recorded DPOAE. In the first, the reverse transmission of the f_{dp} signal on the basilar membrane and across the middle ear may have significantly attenuated the energy of the DPOAE. The DPOAE emitted into the canal could then under-represent the size of the stimulus at frequency f_{dp} on the basilar membrane at the f_{dp} characteristic place. A probe tone similar in frequency to f_{dp} but with a more favorable forward transmission pathway could have a higher level in the canal, but still be similar level to the combination tone near the f_{dp} characteristic place. Thus the probe and combination tones could possibly beat together near the f_{dp} characteristic place. Zwicker and Harris (1990) suggest cochlear origins of the gross discrepancy between perceptual and DPOAE levels that include such transmission characteristics. Reverse transmission can affect emission level in a complex fashion (Shera and Zweig, 1992a, b), but Keefe (2002) suggests that near 500 Hz the forward and reverse transfer functions are similar. Zhang and Abbas (1997) also found a similar effect of middle ear pres-

sure on forward and reverse transmission. Thus it is possible that at the f_{dp} characteristic place in the cochlea, the probe tone caused a much larger response on the basilar membrane than the $2f_1$-f_2 combination tone.

If this is true, then the perceived beating may not have originated in the cochlea. Nonlinearities also occur at all stages of neural processing above the cochlea. The probe tone would be transduced into a neural signal at its characteristic place on the basilar membrane near f_{dp}. The $2f_1$-f_2 combination tone initiated near the f_2 characteristic place is present both in neurons servicing the f_2 place, and in neurons servicing its own characteristic f_{dp} place (Kim *et al.*, 1980). If the basilar membrane response at frequency f_{dp} is small as suggested by the DPOAE, then beating between the large probe tone and small combination tone may occur in the auditory nervous system beyond the basilar membrane. The nonlinearities in these interactions may be such that a high level probe beats best with a $2f_1$-f_2 combination tone that was small in the cochlea but larger centrally. That combination tone may be larger in neural terms because of processes that enhance the detection of envelope frequencies from the firing patterns of neurons servicing the regions of the f_2 and f_1 characteristic places. Beating could then occur in the central nervous system between the $2f_1$-f_2 combination tone and the probe tone, which itself was mediated through neurons servicing the f_{dp} region of the cochlea.

A 40-Hz beat can be generated in the auditory nervous system through binaural presentation of stimuli that differ by 40 Hz (Schwarz and Taylor, 2005, Draganova *et al.*, unpublished). The binaural auditory steady-state response has different characteristics from the response evoked by two tones in the same ear. These findings clearly indicate that nonlinearities of the auditory nervous system could generate a beating response distinct from that generated by nonlinearities in the cochlea.

Correlation analysis, however, showed no linear relationship between DPOAE magnitude and either the psychoacoustic or MEG measurements. As discussed earlier, this may be due to the unique processes involved in the generation and transmission of DPOAEs, or it may be because the beat response is generated above the cochlea in the auditory nervous system. There was also no correlation between best-beat probe tone levels and the MEG source space projection dipole moment magnitudes. Although these might be expected to be more closely related (both being mediated by the brain rather than the cochlea), one involves the complexities of perception (of beating near 5 Hz), and the other the synchronous firing of large populations of neurons at 40 Hz. There may be large between-subject variations in the number of neurons that fire synchronously in response to amplitude modulation. Within-subjects, the changes in neural populations and synchrony between 5 and 40 Hz may also be large. The net result of these complexities was that no correlations were found between the psychoacoustic and MEG correlates of the $2f_1$-f_2 combination tone, or between the DPOAE measurements and either the psychoacoustic or MEG measurements.

No significant change in DPOAE level was expected (and none was found) with added probe tones since they are lower in frequency than $2f_1$-f_2. Suppression is most effective either near the stimulus tones f_2 and f_1, or about 25 Hz above the DPOAE frequency (Heitmann *et al.*, 1998; Gaskill and Brown, 1996; Talmadge *et al.*, 1999; Konrad-Martin *et al.*, 2001). In the latter case, suppression is only present if a significant $2f_1$-f_2 place source is contributing to the DPOAE measured in the ear canal. In the experiment reported here, the measured DPOAE should be from the stimulus region, and thus no suppression would be expected. Although there was no significant difference in average DPOAE level with the added probe tones, there were individuals whose response was slightly lower with the probe tones present.

B. MEG response

Dipoles were fit in each hemisphere using the 70-dB SPL ECD stimulus. The average position values in Table I can be compared to those reported previously using a similar head model (Ross *et al.*, 2000). The Ross *et al.* (2000) values are all within 1.3 s.d. of the means found here (using the present s.d. values). Since the ECD stimulus frequencies were identical to those of the probe and combination tones, it was assumed that the 40-Hz sources would be the same. Sources are known to be tonotopic with carrier frequency (Pantev *et al.*, 1996), but it is not anticipated that they would vary in location with stimulus level.

A 40-Hz beat was chosen to increase the likelihood of detecting a response in the MEG signal due to the 40-Hz peak in the transfer function of the steady-state response (Ross *et al.*, 2000). This required lowering the probe tone frequency to 460 Hz from the 495.5 Hz that was used in the psychoacoustic measurement of best-beats. The transfer function of the middle ear and cochlea should not change enough over this frequency range to affect the depth of modulation achieved between the probe and $2f_1$-f_2 combination tone.

Detection of the cortical correlate of the $2f_1$-f_2 combination tone (evoked at perceptually relevant frequencies) was only possible indirectly because the level of the MEG steady-state response drops precipitously as the modulation frequency approaches 100 Hz (Ross *et al.*, 2000). In addition, detection is influenced by the practical limitations of recording MEG signals from deep sources. Evidence suggests that the envelope following response above 100 Hz is mostly from sources in the brainstem (Herdman *et al.*, 2002; Purcell *et al.*, 2004). The indirect method used here was to cause beating between the $2f_1$-f_2 combination tone and an external probe tone in Experiment 1, or between two different $2f_1$-f_2 combination tones in Experiment 2. Beating or envelope modulation evokes the envelope following response at the difference frequency between the tones acting as stimuli. The quadratic combination tone has also been well studied psychoacoustically (e.g., Plomp, 1965; Goldstein, 1966; Humes, 1979; 1980, 1985), and has been recorded objectively along with the $2f_1$-f_2 combination tone in some physiological studies (e.g., Smoorenburg *et al.*, 1976; Kim *et al.*, 1980; Gibian and Kim, 1982; Cheatham and Dallos, 1997). The two have been employed together in psychoacoustics to facilitate judgments involving $2f_1$-f_2 (e.g.,

Goldstein, 1966). Combination tones have also been observed to evoke other combination tones (e.g., Goldstein *et al.*, 1978; Cheatham and Dallos, 1997). The purposeful use of both orders of combination tones here permitted a signal derived from the $2f_1$-f_2 combination tone to be measured in cortex that would have been otherwise unavailable.

In Table II, the average test trial source space projection dipole moment magnitude is larger (although not statistically significantly so) in the right hemisphere contralateral to the stimulus presentation to the left ear as expected from the results of Ross *et al.* (2005). The average test trial magnitudes are smaller than those given in Table I for the dipole moments elicited with the ECD stimulus. The ECD stimulus level was 70 dB SPL, whereas the average best-beat probe tone level was close to 45 dB SPL. For a 250-Hz carrier that was amplitude modulated at 39 Hz, Ross *et al.* (2000) reported that the dipole moment decreased to about 50%for a decrease in stimulus level of 25 dB. In the present study, the beating stimulus elicited a response that was less than 20% of the response to the ECD stimulus. While this could in small part be due to the presence of multiple tone pairs (John *et al.*, 1998), this increased difference than what might be expected from the probe tone level is further evidence of the disconnection between best-beat level and MEG magnitude. If the Ross *et al.* (2000) data are representative of the situation here, then the probe beating either did not have a modulation depth near 100%, or was lower in level than indicated by the probe tone level. Psychoacoustic methods involving simultaneous probe and stimulus tones may overestimate the level of combination tones (Smoorenburg, 1972b;Shannon and Houtgast, 1980). However, it is not clear where in the system the beat is generated and what the relative levels of the probe tone and combination tone may be at that generator.

In Experiment 2, a successful attempt was made to elicit a 40-Hz MEG response using two $2f_1$-f_2 combination tones. The purpose of this demonstration was to show that correlates of the $2f_1$-f_2 combination tone could be present in cortex without the administration of an additional tone. Only one of two participants in Experiment 2 showed a significant 40-Hz MEG response. This was not unexpected since a response would only be present if the two $2f_1$-f_2 combination tones were fortuitously of similar level. The two stimulus tone pair levels were the same, but f_2/f_1 was quite different for the two pairs. Stimulus frequency ratio is known to affect the level of DPOAEs (e.g., Harris *et al.*, 1989; Gorga *et al.*, 1993; Vinck *et al.*, 1996; Abdala, 1996), and psychoacoustic estimates (e.g., Goldstein, 1966; Smoorenburg, 1972a, b; Wilson, 1980). If the two $2f_1$-f_2 combination tones were sufficiently different in level where they interacted, the depth of the envelope modulation would be insufficient to elicit a recordable 40-Hz MEG response. If the absolute level of the combination tones were also too low, there would also be no response. The fact that a very definite 40-Hz response was elicited in one participant is evidence for the existence of both combination tones and their interaction in the auditory system.

In response to the ECD stimulus, the single individual with significant responses in Experiment 2 had dipoles of 0.84 and 1.89 nA m in the left and right hemispheres, respectively. The corresponding CT beat had source space projection dipole moment magnitudes that were −18.0 and −17.7 dB smaller. These values are only about −3 dB smaller than those between the ECD and probe beating stimuli discussed above for Experiment 1. Given the saturating monotonic input/output relationship reported by Ross *et al.* (2000; their Fig. 7A), the combination tones in Experiment 2 may have been slightly smaller than the average ones in Experiment 1, or the achieved depth of modulation may have been slightly less favorable.

V. CONCLUDING REMARKS

These recordings have clearly demonstrated that the human auditory cortex responds to the $2f_1$-f_2 combination tone. There were no significant correlations between the perceived level of the combination tone and the magnitude of either the cortical MEG response or the cochlear DPOAE. The perceived combination tone is therefore likely mediated by processes above those that produce the measured physiological responses. The MEG responses show that information about the combination tone is transmitted to cortex, but later complex interactions between cortical areas likely mediate its perception.

ACKNOWLEDGMENTS

Research supported by the Canadian Institutes of Health Research, and the National Institutes of Health Research.

Abdala, C. (**1996**). "Distortion product otoacoustic emission (2f1-f2) amplitude as a function of f2/f1 frequency ratio and primary tone level separation in human adults and neonates," J. Acoust. Soc. Am. **100**, 3726–3740.

Brown, D. K., Bowman, D. M., and Kimberley, B. P. (**2000**). "The effects of maturation and stimulus parameters on the optimal f2/f1 ratio of the 2f1-f2 distortion product otoacoustic emission in neonates," Hear. Res. **145**, 17–24.

Buunen, T. J., and Rhode, W. S. (**1978**). "Responses of fibers in the cat's auditory nerve to the cubic difference tone," J. Acoust. Soc. Am. **64**, 772–781.

Cheatham, M. A., and Dallos, P. (**1997**). "Intermodulation components in inner hair cell and organ of corti responses," J. Acoust. Soc. Am. **102**, 1038–1048.

Dhar, S., Long, G. R., Talmadge, C. L., and Tubis, A. (**2005**). "The effect of stimulus-frequency ratio on distortion product otoacoustic emission components," J. Acoust. Soc. Am. **117**, 3766–3776.

Dolphin, W. F. (**1997**). "The envelope following response to multiple tone pair stimuli," Hear. Res. **110**, 1–14.

Dolphin, W. F., and Mountain, D. C. (**1993**). "The envelope following response (EFR) in the mongolian gerbil to sinusoidally amplitude-modulated signals in the presence of simultaneously gated pure tones," J. Acoust. Soc. Am. **94**, 3215–3226.

Fastl, H. (**1977**). "Roughness and temporal masking patterns of sinusoidally amplitude modulated broadband noise," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 403–414.

Fastl, H. (**1990**). "The hearing sensation roughness and neuronal responses to AM-tones," Hear. Res. **46**, 293–295.

Furst, M., Rabinowitz, W. M., and Zurek, P. M. (**1988**). "Ear canal acoustic distortion at 2f1-f2 from human ears: Relation to other emissions and perceived combination tones," J. Acoust. Soc. Am. **84**, 215–221.

Gaskill, S. A., and Brown, A. M. (**1990**). "The behavior of the acoustic distortion product, 2f1-f2, from the human ear and its relation to auditory sensitivity," J. Acoust. Soc. Am. **88**, 821–839.

Gaskill, S. A., and Brown, A. M. (**1996**). "Suppression of human acoustic distortion product: Dual origin of 2f1-f2," J. Acoust. Soc. Am. **100**, 3268–3274.

Gibian, G. L., and Kim, D. O. (**1982**). "Cochlear microphonic evidence for mechanical propagation of distortion products (f2-f1) and (2f1-f2)," Hear. Res. **6**, 35–59.
Goldstein, J. L. (**1966**). "Auditory nonlinearity," J. Acoust. Soc. Am. **41**, 676–689.
Goldstein, J. L. (**1969**). "Aural combination tones," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden, The Netherlands), pp. 230–247.
Goldstein, J. L., Buchsbaum, G., and Furst, M. (**1978**). "Compatibility between psychophysical and physiological measurements of aural combination tones," J. Acoust. Soc. Am. **63**, 474–485.
Goldstein, J. L., and Kiang, N. Y. S. (**1968**). "Neural correlates of the aural combination tone 2f1-f2," Proc. IEEE **56**, 981–992.
Gorga, M. P., Neely, S. T., Bergman, B. M., Beauchaine, K. L., Kaminski, J. R., Peters, J., Schulte, L., and Jesteadt, W. (**1993**). "A comparison of transient-evoked and distortion product otoacoustic emissions in normal-hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **94**, 2639–2648.
Gutschalk, A., Mase, R., Roth, R., Ille, N., Rupp, A., Hahnel, S., Picton, T. W., and Scherg, M. (**1999**). "Deconvolution of 40 Hz steady-state fields reveals two overlapping source activities of the human auditory cortex," Clin. Neurophysiol. **110**, 856–868.
Hall, J. L. (**1975**). "Nonmonotonic behavior of distortion product 2f1-f2: Psychophysical observations," J. Acoust. Soc. Am. **58**, 1046–1050.
Harris, F. P., Lonsbury-Martin, B. L., Stagner, B. B., Coats, A. C., and Martin, G. K. (**1989**). "Acoustic distortion products in humans: Systematic changes in amplitudes as a function of f2/f1 ratio," J. Acoust. Soc. Am. **85**, 220–229.
Heitmann, J., Waldmann, B., Schnitzler, H. U., Plinkert, P. K., and Zenner, H. P. (**1998**). "Suppression of distortion product otoacoustic emissions (DPOAE) near 2f1-f2 removes DP-gram fine structure-Evidence for a secondary generator," J. Acoust. Soc. Am. **103**, 1527–1531.
Herdman, A. T., Lins, O., Van Roon, P., Stapells, D. R., Scherg, M., and Picton, T. W. (**2002**). "Intracerebral sources of human auditory steady-state responses," Brain Topogr **15**, 69–86.
Humes, L. E. (**1979**). "Perception of the simple difference tone (f2-f1)," J. Acoust. Soc. Am. **66**, 1064–1074.
Humes, L. E. (**1980**). "Growth of L(f2-f1) and L(2f1-f2) with input level: Influence of f2/f1," Hear. Res. **2**, 115–122.
Humes, L. E. (**1985**). "Cancellation level and phase of the (f2-f1) distortion product," J. Acoust. Soc. Am. **78**, 1245–1251.
John, M. S., Lins, O. G., Boucher, B. L., and Picton, T. W. (**1998**). "Multiple auditory steady state responses (MASTER): Stimulus and recording parameters," Audiology **37**, 59–82.
John, M. S., and Picton, T. W. (**2000**). "MASTER: A windows program for recording multiple auditory steady-state responses," Comput. Methods Programs Biomed. **61**, 125–150.
Jones, A. T. (**1935**). "The discovery of difference tones," Am. Phys. Teach. **3**, 49–51.
Keefe, D. H. (**2002**). "Spectral shapes of forward and reverse transfer functions between ear canal and cochlea estimated using DPOAE input/output functions," J. Acoust. Soc. Am. **111**, 249–260.
Kemp, D. T., and Brown, A. M. (**1984**). "Ear canal acoustic and round window electrical correlates of 2f1-f2 distortion generated in the cochlea," Hear. Res. **13**, 39–46.
Kim, D. O., Molnar, C. E., and Matthews, J. W. (**1980**). "Cochlear mechanics: Nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal sound pressure," J. Acoust. Soc. Am. **67**, 1704–1721.
Knight, R. D., and Kemp, D. T. (**2000**). "Indications of different distortion product otoacoustic emission mechanisms from a detailed f1, f2 area study," J. Acoust. Soc. Am. **107**, 457–473.
Knight, R. D., and Kemp, D. T. (**2001**). "Wave and place fixed DPOAE maps of the human ear," J. Acoust. Soc. Am. **109**, 1513–1525.
Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," J. Acoust. Soc. Am. **108**, 723–734.
Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**). "Sources of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast Fourier transforms in normal ears," J. Acoust. Soc. Am. **109**, 2862–2879.
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.
Mäkelä, J. P., and Hari, R. (**1987**). "Evidence for cortical origin of the 40 Hz auditory evoked response in man," Electroencephalogr. Clin. Neurophysiol. **66**, 539–546.
Nuttall, A. L., and Dolan, D. F. (**1990**). "Inner hair cell responses to the 2f1-f2 intermodulation distortion product," J. Acoust. Soc. Am. **87**, 782–790.
Pandya, P. K., and Krishnan, A. (**2004**). "Human frequency-following response correlates of the distortion product at 2f1-f2," J. Am. Acad. Audiol. **15**, 184–197.
Pantev, C., Roberts, L. E., Elbert, T., Ross, B., and Wienbruch, C. (**1996**). "Tonotopic organization of the sources of human auditory steady-state responses," Hear. Res. **101**, 62–74.
Plomp, R. (**1965**). "Detectability threshold for combination tones," J. Acoust. Soc. Am. **37**, 1110–1123.
Probst R., Loonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.
Purcell, D. W., John, S. M., Schneider, B. A., and Picton, T. W. (**2004**). "Human temporal auditory acuity as assessed by envelope following responses," J. Acoust. Soc. Am. **116**, 3581–3593.
Rickman, M. D., Chertoff, M. E., and Hecox, K. E. (**1991**). "Electrophysiological evidence of nonlinear distortion products to two-tone stimuli," J. Acoust. Soc. Am. **89**, 2818–2826.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1991**). "Two-tone distortion in the basilar membrane of the cochlea," Nature (London) **349**, 413–414.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1997**). "Two-tone distortion on the basilar membrane of the chinchilla cochlea," J. Neurophysiol. **77**, 2385–2399.
Ross, B., Borgmann, C., Draganova, R., Roberts, L. E., and Pantev, C. (**2000**). "A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones," J. Acoust. Soc. Am. **108**, 679–691.
Ross, B., Herdman, A. T., and Pantev, C. (**2005**). "Right hemispheric laterality of human 40 Hz auditory steady-state responses," Cereb. Cortex **15**, 2029–2039.
Ross, B., Picton, T. W., and Pantev, C. (**2002**). "Temporal integration in the human auditory cortex as represented by the development of the steady-state magnetic field," Hear. Res. **165**, 68–84.
Schoonhoven, R., Boden, C. J., Verbunt, J. P., and de Munck, J. C. (**2003**). "A whole head MEG study of the amplitude-modulation-following response: Phase coherence, group delay and dipole source analysis," Clin. Neurophysiol. **114**, 2096–2106.
Schwarz, D. W., and Taylor, P. (**2005**). "Human auditory steady state responses to binaural and monaural beats," Clin. Neurophysiol. **116**, 658–668.
Shannon, R. V., and Houtgast, T. (**1980**). "Psychophysical measurements relating suppression and combination tones," J. Acoust. Soc. Am. **68**, 825–829.
Shera, C. A., and Guinan, J. J., Jr. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.
Shera, C. A., and Zweig, G. (**1992a**). "Middle-ear phenomenology: The view from the three windows," J. Acoust. Soc. Am. **92**, 1356–1370.
Shera, C. A., and Zweig, G. (**1992b**). "Analyzing reverse middle-ear transmission: Noninvasive gedankenexperiments," J. Acoust. Soc. Am. **92**, 1371–1381.
Siegel, J. H. (**1994**). "Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes," J. Acoust. Soc. Am. **95**, 2589–2597.
Smoorenburg, G. F. (**1972a**). "Audibility region of combination tones," J. Acoust. Soc. Am. **52**, 603–614.
Smoorenburg, G. F. (**1972b**). "Combination tones and their origin," J. Acoust. Soc. Am. **52**, 615–632.
Smoorenburg, G. F., Gibson, M. M., Kitzes, L. M., Rose, J. E., and Hind, J. E. (**1976**). "Correlates of combination tones observed in the response of neurons in the anteroventral cochlear nucleus of the cat," J. Acoust. Soc. Am. **59**, 945–962.
Strickland, E. A. (**2000**). "The effects of frequency region and level on the temporal modulation transfer function," J. Acoust. Soc. Am. **107**, 942–952.
Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (**1999**). "Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **105**, 275–292.

Vinck, B. M., De Vel, E., Xu, Z. M., and Van Cauwenberge, P. B. (**1996**). "Distortion product otoacoustic emissions: A normative study," Audiology **35**, 231–245.

von Helmholtz, H. L. F. (**1877/1954**). *On the Sensations of Tone* [translation of *Die Lehre von den Tonempfindugen*] (Dover, New York).

Whitehead, M. L., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). "Effects of ear-canal standing waves on measurements of distortion-product otoacoustic emissions," J. Acoust. Soc. Am. **98**, 3200–3214.

Wilson, J. P. (**1980**). "The combination tone, 2f1-f2, in psychophysics and ear-canal recording," in *Psychophysical, Physiological and Behavioral Studies in Hearing*, edited by G. Van der Brink and F. A. Bilsen (Delft University, Delft, The Netherlands), pp. 43–50.

Yost, W. A. (**2000**). *Fundamentals of Hearing: An Introduction*, 4th ed. (Elsevier Academic, San Diego).

Zhang, M., and Abbas, P. J. (**1997**). "Effects of middle ear pressure on otoacoustic emission measures," J. Acoust. Soc. Am. **102**, 1032–1037.

Zwicker, E. (**1955**). "Der ungewöhnliche amplitudengang der nichtlinearen verzerrungen des ohres," Acustica **5**, 67–74.

Zwicker, E. (**1981**). "Dependence of level and phase of the (2f1-f2)-cancellation tone on frequency-range, frequency difference, level of primaries, and subject," J. Acoust. Soc. Am. **70**, 1277–1288.

Zwicker, E., and Harris, F. P. (**1990**). "Psychoacoustical and ear canal cancellation of (2f1-f2)-distortion products," J. Acoust. Soc. Am. **87**, 2583–2591.

Zwislocki, J. (**1953**). "Acoustic attenuation between the ears," J. Acoust. Soc. Am. **25**, 752–759.

Spectral modulation masking patterns reveal tuning to spectral envelope frequency

Aniket A. Saoji[a)]
Psychoacoustic Laboratory, Center for Hearing and Deafness, Department of Communicative Disorders and Sciences, State University of New York at Buffalo, 122, Cary Hall, 3435 Main Street, Buffalo, New York 14314

David A. Eddins[b)]
Department of Otolaryngology, University of Rochester, 2365 South Clinton Avenue, Suite 200, Rochester, New York 14618 and International Center for Hearing and Speech Research, Rochester Institute of Technology, Rochester, New York 14623

(Received 16 March 2006; revised 30 May 2007; accepted 30 May 2007)

Auditory processing appears to include a series of domain-specific filtering operations that include tuning in the audio-frequency domain, followed by tuning in the temporal modulation domain, and perhaps tuning in the spectral modulation domain. To explore the possibility of tuning in the spectral modulation domain, a masking experiment was designed to measure masking patterns in the spectral modulation domain. Spectral modulation transfer functions (SMTFs) were measured for modulation frequencies from 0.25 to 14 cycles/octave superimposed on noise carriers either one octave (800–1600 Hz, 6400–12 800 Hz) or six octaves wide (200–12 800 Hz). The resulting SMTFs showed maximum sensitivity to modulation between 1 and 3 cycles/octave with reduced sensitivity above and below this region. Masked spectral modulation detection thresholds were measured for masker modulation frequencies of 1, 3, and 5 cycles/octave with a fixed modulation depth of 15 dB. The masking patterns obtained for each masker frequency and carrier band revealed tuning (maximum masking) near the masker frequency, which is consistent with the theory that spectral envelope perception is governed by a series of spectral modulation channels tuned to different spectral modulation frequencies. [DOI: 10.1121/1.2751267]

PACS number(s): 43.66.Ba [JHG] Pages: 1004–1013

I. INTRODUCTION

A principal physical characteristic that often distinguishes one sound from another is the relative distribution of power across frequency, or the spectral envelope. The auditory representation of complex spectral envelopes is fundamental to certain aspects of sound localization (Musicant and Butler, 1985), speech perception (Peterson and Barney, 1952; Blandon and Lindblom, 1981), and many other auditory tasks (Bucklein, 1981). To be useful, the peaks and adjacent valleys within a spectral envelope must be maintained in the internal representation of the acoustic spectrum. In the past, auditory scientists have used various experimental procedures to study the representation of simple and complex spectral envelope patterns. However, the specific mechanism used by the auditory system to extract and encode complex spectral patterns is still a matter of speculation. One hypothesis is that the auditory system analyzes a complex spectral envelope via a modulation filter bank in the spectral domain (Shamma *et al.*, 1995; Schreiner and Calhoun, 1995). In this article, a masking experiment is reported that provides psychophysical evidence of tuning to spectral modulation frequency, supporting the modulation filter bank theory of spectral envelope perception.

Several studies of spectral envelope perception have included parametric manipulation of specific acoustic features. A common approach used to study spectral envelope discrimination is to simply measure the ability of a listener to discriminate between the overall spectral envelopes of two different acoustic spectra, the most popular technique being profile analysis (e.g., Spiegel and Green, 1982; Green *et al.*, 1983; Bernstein and Green, 1987a). In the typical profile analysis technique, a complex acoustic spectrum consists of sinusoidal components equally spaced on a logarithmic frequency axis. The task of the listener is to discriminate between a flat spectrum (the standard stimulus) and a peaked spectrum (the signal or target spectrum), where the peak is produced by adding an increment to one of the components in the standard. The profile analysis technique has led the way to differentiating between global and local processing of intensity variations across audio frequency.

A second approach used to investigate spectral shape discrimination involves the detection of sinusoidal spectral modulation. Previous studies of spectral modulation detection have shown that the lowest modulation detection thresholds correspond to spectral modulation frequencies between 2.0 and 4.0 cycles/octave, beyond which the threshold for modulation detection gradually increases (Bernstein and Green, 1987a,b; Hillier, 1991; Summers and Leek, 1994; Amagai *et al.*, 1999; Chi *et al.*, 1999; Eddins and Bero, 2007). Spectral modulation detection thresholds obtained for the various modulation frequencies are characterized by a

[a)]Electronic mail: aniket.saoji@gmail.com
[b)]Electronic mail: david_eddins@urmc.rochester.edu

spectral modulation transfer function (SMTF). The notion of SMTF is similar in concept to the temporal modulation transfer function (TMTF) that is used to describe the relation between temporal modulation detection thresholds and temporal modulation frequency (e.g., Viemeister, 1979). Interestingly, psychoacoustic studies in the past have focused their discussion of SMTFs on the failure to resolve the peaks and valleys and the separation of within-channel versus across-channel processing as revealed in specific tasks (e.g., Bernstein and Green, 1987b; Summers and Leek, 1994). Therefore, the focus has been on more traditional measures of frequency selectivity rather than on the nature of the across-(audio-frequency) channel processing underlying spectral envelope perception.

Several studies have shown tuning to spectral modulation frequency and have attributed this tuning to spectral modulation-specific channels in the auditory system. For example, physiological studies of the responses of individual cells in the auditory cortex have revealed bandpass tuning to spectral modulation frequency over the range of 0.2–3 cycles/octave (Schreiner and Mendelson, 1990; Schreiner and Sutter, 1992; Schreiner and Calhoun, 1995; Shamma *et al.*, 1993, 1995; Shamma and Versnel, 1995; Versnel *et al.*, 1995; Kowalski *et al.*, 1996a,b; Versnel and Shamma, 1998). The cortical responses to the simple sinusoidally modulated spectra have been used to predict the responses to a complex spectral envelope by using the principle of superposition (Kowalski *et al.*, 1996b). Furthermore, Versnel and Shamma (1998) have provided considerable evidence that a linear sinusoidal spectral modulation analysis can be used to predict the cortical representation of complex spectral envelopes of vowel stimuli. Thus, physiological evidence indicates that cells in the auditory cortex function as a spectral modulation filter bank and, in effect, are analogous to a crude Fourier analysis of the complex spectral envelope. To date, however, behavioral evidence of tuning to sinusoidal spectral modulation has not been reported.

Tuning to spectral modulation and the associated domain-specific filtering operations demonstrated physiologically are similar in concept to the analysis of complex wave forms by the auditory periphery and the analysis of complex temporal envelopes by the central auditory system. In the audio-frequency domain, psychophysical (Fletcher, 1940; Hamilton, 1957; Greenwood, 1961; Spiegel, 1981; Schooneveldt and Moore, 1989; Bernstein and Raab, 1990; Moore and Ohgushi, 1993) and physiological (Rhode *et al.*, 1978; Evans, 1975; Costalupes *et al.*, 1984; Young and Barta, 1986; Evans, 1992) evidence indicates that the auditory system analyzes complex sounds by selective filtering that occurs at the cochlear level; specifically the basilar membrane. Likewise, tuning to temporal modulation has been demonstrated via psychophysical adaptation and masking experiments (Bacon and Grantham, 1989; Houtgast, 1989; Tansley and Suffield, 1983; Yost *et al.*, 1989) performed in humans with sinusoidally amplitude modulated noises and tones. Furthermore, physiological recordings from cells in the cochlear nucleus (e.g., Frisina *et al.*, 1990; Møller, 1976), the inferior colliculus (e.g., Rees and Møller, 1987; Rees and Palmer, 1989), and the auditory cortex (e.g., Eggermont, 1994; Schreiner and Urbas, 1986, 1988) of various mammals indicate that the neurons at various levels in the auditory system act as a cascade of temporal modulation filters that are tuned to temporal modulation frequency. Thus, the notion of spectral modulation-frequency specific channels is consistent with a series of domain-specific filtering operations that analyze complex acoustic patterns into different auditory channels.

The primary purpose of the present study is to determine whether or not the auditory system exhibits tuning to spectral modulation frequency. Therefore, a masking approach using simple sinusoidal spectral modulation was adopted because of its successful use in the characterization of channel-specific tuning in the audio-frequency and temporal modulation domains (e.g., Fletcher, 1940; Houtgast, 1989). Following the logic of previous investigators, if spectral modulation channels exist, then one sinusoidal spectral modulation should maximally interfere with the perception of another sinusoidal spectral modulation when the two are similar in modulation frequency. Likewise, when the masker and signal modulation frequencies differ substantially, little masking is predicted because the two spectral modulation frequencies would be processed by separate modulation channels. Such tuning to modulation frequency would support the theory of spectral modulation channels in the auditory system. The lack of spectral modulation-frequency specific tuning would fail to support such a theory.

II. METHODS

A. Subjects

Four subjects ranging in age from 22 to 26 years participated in these experiments. Each subject had normal hearing based upon pure-tone thresholds ($<$20 dB HL from 250 to 8000 Hz, ANSI, 1996) and screening tympanograms (Y, 226 Hz). Pure-tone thresholds from 8000 to 13000 Hz were assessed by Bekesy audiometry and all listeners fell within laboratory norms. With the exception of the first author, the subjects received an hourly wage and were given a 20% bonus upon completion of the project. One of the four subjects had previous experience in psychoacoustic listening tasks.

B. Conditions

Both unmasked and masked spectral modulation detection thresholds were measured. Unmasked spectral modulation detection thresholds were obtained by determining the modulation depth (dB) necessary to discriminate between a signal having a sinusoidally modulated spectrum and a standard having a flat spectrum at each of the following modulation frequencies: 0.25, 0.5, 1, 2, 3, 4, 5, 7, 10, or 14 cycles/octave.

Masked modulation detection thresholds were obtained by varying the modulation depth of the signal modulation in the presence of a sinusoidal masker modulation with a fixed spectral modulation depth of 15 dB. A depth of 15 dB was chosen both to ensure that the masker modulation at each frequency was salient and, as a first approximation to the spectral contrast, typical of American English vowels (based

TABLE I. Stimulus parameters. Carrier bandwidths, masker, and signal frequencies for which the masking functions were obtained.

Carrier bandwidth	Masker (cycles/octave)	Signal (cycles/octave)
200–12 800 Hz (6 oct)	1, 3, 5	0.25–14.14
800–1600 Hz (1 oct)	3	0.75–8.48
6400–12 800 Hz (1 oct)	1, 3, 5	Various

on a female talker) as well as head-related transfer functions (taken from the CIPIC database; Algazi *et al.*, 2001) computed prior to this experiment. The masker modulation frequency was either 1, 3, or 5 cycles/octave. For each masker frequency, signal frequencies at, above, and below the masker modulation frequency were chosen to best evaluate potential tuning to spectral modulation frequency. The various conditions for which the masked thresholds were obtained are summarized in Table I. To investigate the role of carrier frequency range, unmasked and masked spectral modulation detection thresholds were obtained for carrier bandwidths of 200–12 800, 800–1600, or 6400–12 800 Hz.

C. Stimuli

All stimuli were generated using a digital array processor (TDT AP2) and a 16 bit D/A converter with a sampling period of 24.4 μs (40,983 Hz). The first step was to compute the complex spectrum of the desired signal. First, two 8192-point buffers, X (real part of the spectrum) and Y (imaginary part of the spectrum) were filled with the same sinusoid computed on a logarithmic frequency scale. Next, X and Y were multiplied by independent 8192-point samples from a Gaussian distribution. Then X and Y were multiplied by an 8192-point buffer filled with values corresponding to the magnitude response of a second-order Butterworth filter with appropriate passband (200–12 800, 800–1600, or 6400–12 800 Hz). An inverse Fast Fourier Transfor (FFT) was performed on the complex spectrum (X,Y) to produce the desired 400 ms spectrally shaped noise wave form. Finally, the wave forms were shaped with a 10 ms $\cos^2$ window and scaled to the desired presentation level. For the masker plus signal stimuli, the same technique was used, however, the buffers corresponding to the masker and signal modulators were summed prior to multiplication with the noise buffer.

Independent noise stimuli were presented on each observation interval. For conditions where the masker and signal spectral modulation frequencies differ, the starting phase of each modulator was chosen randomly from a uniform distribution (0 to 2π) on each presentation. When the masker and the signal spectral modulation were same, the starting phase for the signal modulation was chosen randomly and the masker modulation was added in quadrature phase in relation to the signal modulation. Similar phase combinations have been used to generate complex temporal modulation combinations used in the temporal modulation masking experiments reported in the literature, e.g., Wakefield and Viemeister (1990), Grantham and Bacon (1988), Bacon and Grantham (1989). While a previous study demonstrated that spectral modulation detection thresholds were not influenced by interval-to-interval randomization of overall level (rove) or modulator phase (Eddins and Bero, 2007), stimulus generation in the present study included randomization of the starting phase to encourage listeners to focus on the overall spectral pattern and to minimize the probability of using local (in the audio-frequency domain) changes in level as the basis for spectral modulation detection.

The spectrum level of the flat-spectrum standard stimuli was 35 dB SPL. The overall level of the spectrally modulated stimuli was adjusted to be equal to that of the flat-spectrum standard. As a result, the levels of the peaks in the modulated spectrum exceeded 35 dB SPL slightly, in a manner dependent on the required spectral modulation depth.

D. Procedure

A three-interval, single-cue, two alternative, forced choice procedure was used to measure spectral modulation detection thresholds. Prior to data collection, subjects received a few sample trials for any new condition to gain familiarity with the stimuli. On each trial, the three observation intervals were separated by 400 ms silent intervals. The standard stimulus was presented in the first interval as an anchor or reminder. A second standard stimulus and the signal stimulus were randomly assigned to the remaining two presentation intervals. The threshold was estimated using an adaptive psychophysical procedure employing 60 trials. The spectral modulation depth was reduced after three consecutive correct responses and increased after a single incorrect response. The step size was initially 2 dB and was reduced to 0.4 dB after three reversals in the adaptive track. The equilibrium point of such a procedure is 79.4% correct. Stimuli were presented monaurally to the left ear via ER-2 insert phones and subjects were seated in a sound treated booth. Unmasked spectral modulation detection thresholds were based on the average of three successive 60-trial runs. When required to compute masking patterns (see the following), the unmasked thresholds for intermediate signal frequencies were determined by interpolation. Masked spectral modulation detection thresholds were based on the average of six successive 60-trial runs.

III. RESULTS

For clarity of exposition, the results for the wideband and narrowband noise carriers will be considered separately. In each case, unmasked thresholds and masking patterns will be presented as a function of the signal modulation frequency for individual listeners as well as the average across listeners.

A. Wideband noise carrier (200–12 800 Hz)

Individual unmasked thresholds (S1–S4) and the SMTF reported by Eddins and Bero (EB) are shown in Fig. 1 with spectral modulation detection thresholds (dB) as a function of signal modulation frequency (cycles/octave). The error bar refers to the standard error of the mean averaged across all subjects and conditions. These SMTFs indicate that the auditory system is most sensitive in the midspectral modulation

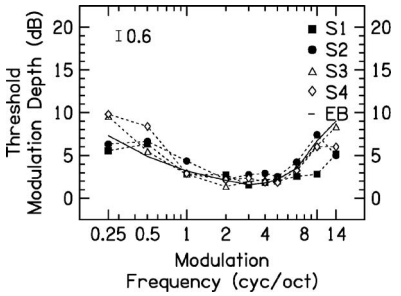


FIG. 1. Spectral modulation transfer functions (SMTFs) for four subjects (S1–S4) over the modulation frequency range of 0.25–14 cycles/octave using a 6 octave wide band of noise (200–12 800 Hz) as a carrier. The solid line labeled EB shows data from Eddins and Bero (2007). The standard error, averaged across all subjects and conditions, is shown by the bar on the upper left.

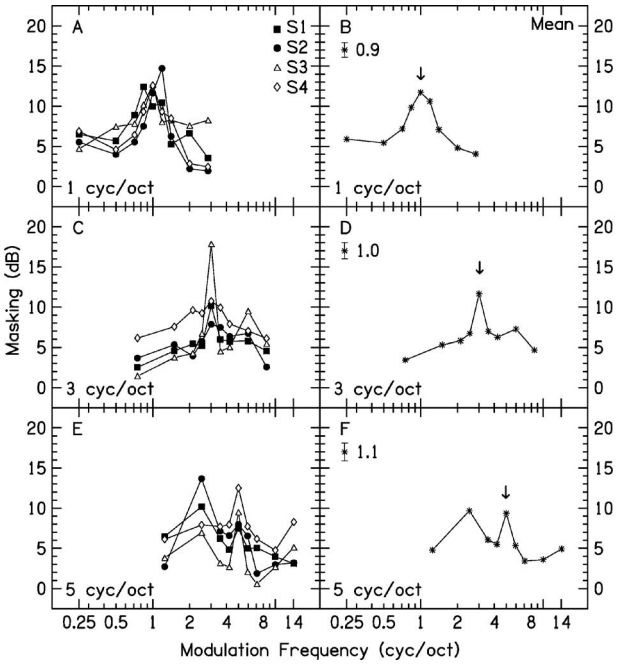


FIG. 2. (A), (C), and (E) (Left column) Individual masking patterns for four subjects (S1–S4) obtained in the presence of masker modulations of 1, 3, and 5 cycles/octave, respectively. (B), (D), and (F) (Right column) The standard error bar averaged across all subjects and conditions and the mean masking patterns for the four subjects (S1–S4) for the masker frequencies of 1, 3, and 5 cycles/octave, respectively. The carrier bandwidth was 200–12 800 Hz.

frequency range, around 1–4 cycles/octave. The sensitivity to modulation is reduced at low- and high-modulation frequencies, confirmed by a one-way repeated measures ANOVA, indicating a significant effect of modulation frequency ($F_{3,9}=14.951$, $p<0.001$). The SMTFs obtained in this study are similar across subjects and follow the same pattern as the transfer functions reported by others (Bernstein and Green, 1987b; Summers and Leek, 1994; Chi *et al.*, 1999; Eddins and Bero, 2007).

Masked SMTFs were obtained for masker modulation frequencies of 1, 3, and 5 cycles/octave superimposed on a broadband carrier 6 octaves in width. The masker modulation depth was 15 dB. For the masker modulation of 1 cycle/octave, masked thresholds were obtained for signal frequencies ranging from 0.25 to 2.82 cycles/octave. To determine the amount of spectral modulation masking for each subject and condition, the unmasked SMTFs were subtracted from the masked SMTFs, revealing spectral modulation masking patterns as shown in Fig. 2. Individual spectral modulation masking patterns for a masker modulation of 1 cycle/octave show some variability across subjects [Fig. 2(A)] in terms of peak frequency, degree of tuning (width), and the amount of masking (height). Closer examination reveals a peak in the masked SMTF at 1 cycle/octave for S3 and S4, whereas the peaks occur at 0.84 and 1.19 cycles/octave for S1 and S2, respectively. The reason for this slight mistuning remains unclear. The mean data obtained for the masker modulation of 1 cycle/octave indicates a peak at 1 cycle/octave [Fig. 2(B)]. The error bars in Fig. 2 represent the standard error of the mean averaged across all subjects and conditions. For the masker modulation of 3 cycles/octave, spectral modulation masking patterns were obtained over a signal frequency range of 0.75–8.48 cycles/octave. The individual [Fig. 2(C)] and the mean [Fig. 2(D)] data show a distinct peak at masker frequency of 3 cycles/octave for all four subjects. The individual masking patterns reveal a small secondary peak at 6 cycles/octave for three of the four subjects (S1, S2, and S3). For a masker modulation of 5 cycles/octave, spectral modulation masking patterns were obtained over a signal frequency range of 1.25–14.14 cycles/octave. The individual [Fig. 2(E)] and the mean [Fig. 2(F)] masking patterns show a double peaked masking function. Masking functions for two subjects (S1 and S2) reveal greater masking at 2.5 cycles/octave than at 5 cycles/octave. The masking patterns of the other two subjects (S3 and S4) show more masking at 5 cycles/octave than 2.5 cycles/octave. The occurrence of secondary peaks 1 octave above or below the masker frequency will be addressed in the general discussion to follow.

The masking patterns can be characterized in terms of the amount of masking in decibels, the dynamic range in decibels, and the bandwidth in octaves (Table II). The amount of masking refers to the peak value in the masking function. The dynamic range was calculated by subtracting the lowest value from the peak value in the masking function. The bandwidth was computed using the following two methods. In the "half DR" method, two points were first determined. They were the half-way point between the peak and the lowest point in the masking pattern *below* the peak and the half-way point between the peak and the lowest point in the masking pattern *above* the peak. The bandwidth in octaves was computed between these two points. The "half DR Q" factor was computed by dividing the peak frequency by the difference between the two points determined in the half DR method, yielding a "quality" factor proportional to the degree of tuning. In the "3 dB" method, the difference between the 3 dB down point relative to the peak was calculated on the upper and lower slope of the tuning curve. The bandwidth in octaves was computed between these two points. The "3 dB Q" factor was computed by dividing the peak frequency by the difference between the two points obtained in the 3 dB method. The average data for the four subjects is depicted in Table II. Using the above stated criteria, it was not possible to estimate a bandwidth for all the

TABLE II. Summary of tuning characteristics based on masking patterns. Modulation frequency=masker modulation frequency (cycles/octave); Masking=the amount of masking in decibels; DR=dynamic range in decibels; Half DR BW=estimated bandwidth of the spectral modulation frequency filter in octaves using the half DR method; Half DR Q=value obtained from the half DR method; 3 dB BW=estimated bandwidth of the spectral modulation frequency filter in octaves using the 3 dB method; 3 dB Q=value obtained from the 3 dB method. See the text for descriptions of the four methods of bandwidth computation. The standard deviation for each is shown in parentheses. The number of asterisks (*) indicates the number of subjects for whom the bandwidth could not be calculated. Closed triangle (▲) indicates that the values calculated are based on the limited set of data collected on those experimental conditions.

Modulation frequency	Masking (dB)	DR (dB)	Half DR BW (Oct)	Half DR Q	3 dB BW (Oct)	3 dB Q
Carrier bandwidth: 200–12 800 Hz						
1	13.08 (1.08)	9.92 (2.08)	0.79 (0.22)	1.94 (0.71)	0.52 (0.17)	2.98 (0.99)
3	11.64 (4.31)	8.48 (5.43)	0.88 (0.63)	2.54 (1.87)	0.93 (0.80)	3.61 (3.62)
5	11.46 (1.95)	8.89 (2.08)	0.82 (0.92)	3.31 (1.62)	0.91 (1.11)*	11.43 (11.60)*
Carrier bandwidth: 800–1600 Hz						
3	11.1 (2.24)	6.6 (2.33)	0.62 (0.55)	3.55 (2.12)	0.69 (0.73)	3.97 (2.87)
Carrier bandwidth: 6400–12 800 Hz						
1▲	14.06 (2.59)▲	5.30 (2.09)▲	…	…	…	…
3	13.17 (5.40)	8.15 (4.43)	0.36 (0.06)**	4.09 (0.55)**	0.27 (0.12)**	6.0 (2.68)**
5▲	13.29 (3.31)▲	7.77 (3.03)▲	…	…	…	…

subjects for the 5 cycles/octave masker due to the nonmonotonic lower or upper slope of the masking pattern.

The mean amount of masking and the dynamic range was similar across the spectral modulation frequencies of 1, 3, and 5 cycles/octave. The large standard deviations, however, indicate that there were substantial differences across subjects. The bandwidth estimates were similar across the three masker modulation frequencies. For the spectral modulation frequency of 5 cycles/octave, the bandwidth was determined for the peak obtained at the signal frequency of 5 cycles/octave rather than 2.5 cycles/octave. The Q factor indicated greater tuning at 5 cycles/octave and progressively less tuning at 3 and 1 cycles/octave, suggesting an inverse relation between putative filter bandwidth and spectral modulation frequency as determined using the $\log_2$ frequency scale. However, there is substantial variability in the Q factor across subjects and masker modulation frequencies. It should be noted that quantification of these tuning characteristics is limited in some cases by the choice of signal modulation frequencies. As such, these measures should be considered as best estimates given the precision of measurement used.

B. Narrowband noise carriers

Masking patterns were obtained for two narrowband noise carriers to allow a comparison of tuning to spectral modulation across widely separate audio-frequency regions. The unmasked SMTFs obtained for the two narrowband noise carriers (800–1600 Hz, closed circles and 6400–12 800 Hz, open triangles) are shown in Fig. 3 and are very similar to the unmasked SMTF obtained for the broadband (200–12 800 Hz, open squares) noise carrier. The error bar represents the standard error of the mean averaged across all subjects and conditions. The similarities in SMTFs across carrier frequency regions are consistent with the data of Eddins and Bero (2007). A two-way repeated measures ANOVA indicated a significant effect of carrier band ($F_{2,95}=13.901$, $p=0.006$), modulation frequency ($F_{6,95}=7.583$, $p<0.001$), and a significant interaction between the two ($F_{14,95}=5.491$, $p<0.001$). Post-hoc (Tukey) tests indicated that the thresholds for the two 1 octave carriers were not different from each other but were higher than for the broadband carrier. For the broadband and 800–1600 Hz carriers, thresholds for the 8.48 cycles/octave condition were higher than other modulation frequencies except 1.5 cycles/octave, whereas for the 6400–12 800 Hz carrier, threshold at 8.48 cycles/octave was significantly higher than threshold for 6.0 and 4.24 cycles/octave and threshold at 1.5 cycles/octave was significantly higher than 4.24 cycles/octave.

1. Octave band (800–1600 Hz)

Masked SMTFs were obtained for a masker frequency of 3 cycles/octave and signal frequencies ranging from 1.5 to 8.48 cycles/octave. The unmasked SMTFs were subtracted from the masked SMTFs to determine the amount of masking, as shown in Fig. 4. Both the individual [Fig. 4(A)] and mean [Fig. 4(B)] masking patterns show a peak at the signal frequency of 3 cycles/octave. The bar in Fig. 4(B) represents

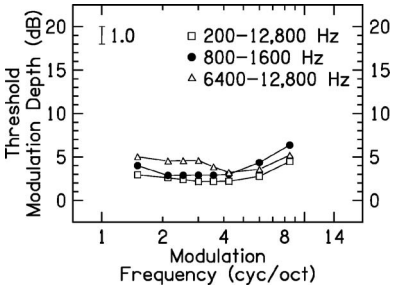


FIG. 3. Mean unmasked SMTFs (1.5–8.48 cycles/octave) and the standard error bar averaged across all subjects and conditions obtained for the three carrier bandwidths (200–12 800, 800–1600, and 6400–12 800 Hz).

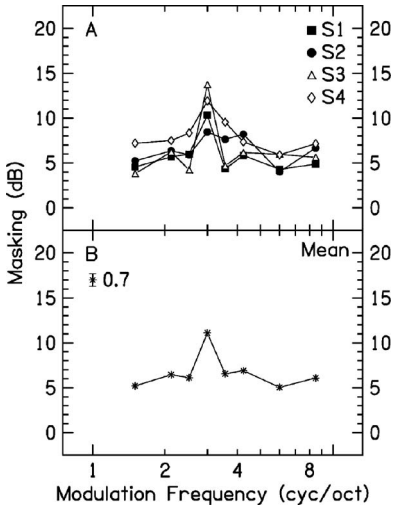


FIG. 4. (A) The individual masking patterns for the four subjects (S1–S4) for a masker modulation of 3 cycles/octave on a 1 octave band of noise (800–1600 Hz). (B) The standard error bar averaged across all subjects and conditions and the mean masking function obtained for the four subjects.

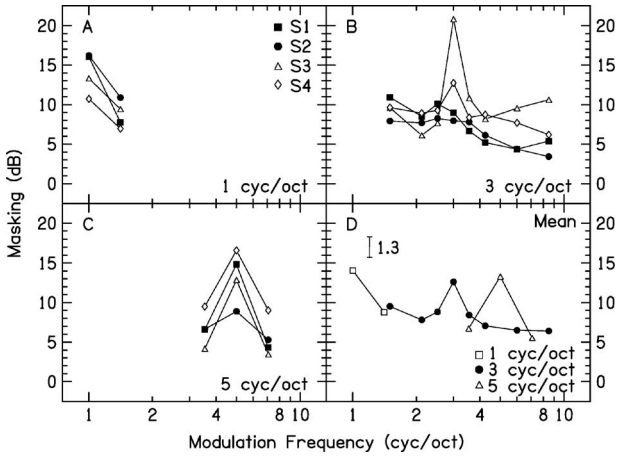


FIG. 5. (A); (B), and (C) The masking pattern for the four subjects (S1–S4) for 1, 3, and 5 cycles/octave, respectively. (D) The standard error bar averaged across all subjects and conditions and the mean masking function for the masker frequencies of 1, 3, and 5 cycles/octave. The carrier bandwidth is a 1 octave narrow band of noise (6400–12 800 Hz).

the standard error of the mean averaged across all subjects and conditions. Careful inspection of the individual data reveals small secondary peaks at various spectral modulation frequencies. Similar to the masking patterns for the broadband masker (200–12 800 Hz), the amount of masking for this 1 octave (800–1600 Hz) noise carrier decreased with increasing separation between the signal and masker modulation frequencies. The amount of masking (dB), dynamic range (dB), bandwidth (oct), and the Q factor characterizing the tuning curves shown in Fig. 4 are summarized in Table II.

2. Octave band (6400–12 800 Hz)

Initially masked SMTFs were obtained for a masker frequency of 3 cycles/octave and signal frequencies ranging from 1.5 to 8.48 cycles/octave superimposed on a noise carrier from 6400 to 12 800 Hz. The unmasked SMTFs were subtracted from the masked SMTFs to determine the amount of masking, as shown in Fig. 5(B).

Inspection of the mean data for the 3 cycles/octave masker indicated a peak in the masking function at a signal modulation of 3 cycles/octave [Fig. 5(D), closed circles], however, the peak at 3 cycles/octave was not consistent across subjects [Fig. 5(B)]. For two subjects (S3 and S4), a peak occurred at a signal frequency of 3 cycles/octave. For the remaining two subjects (S1 and S2), a low-pass masking function was obtained. Because low-pass functions were not seen for the other two carrier conditions (200–12 800 Hz and 800–1600 Hz), an additional experiment was conducted to explore the nature of tuning for this carrier at other spectral modulation frequencies. Abbreviated masking functions were obtained for masker frequencies of 1 and 5 cycles/octave, using only adjacent signal frequencies to determine if tuning existed. For the 1 cycle/octave masker, signal frequencies below the masker frequency (i.e., less than 1 cycle/octave) superimposed upon a 1 octave carrier will have less than 1 cycle of spectral modulation. In this case, the equivalent of "splatter" in the spectral modulation domain would limit interpretation of the data. Thus, masked thresholds were obtained at and above the masker frequency (signal frequencies of 1.0 and 1.1 cycles/octave), but not below the masker frequency. The individual data for the masking functions [Fig. 5(A)] show a distinct elevation at the signal frequency of 1.0 cycle/octave relative to 1.1 cycles/octave, consistent with tuning at that frequency region, but not distinguishing between low-pass or bandpass filter characteristics. However, the amount of change on the high frequency side of the masking function is more similar to the bandpass tuning characteristics than the low-pass tuning characteristics in Fig. 5(B). For the masker frequency of 5 cycles/octave, masked thresholds were obtained at the signal frequencies of 3.54, 5, and 7.07 cycles/octave. Individual [Fig. 5(C)] as well as the mean data [Fig. 5(D), open triangles] show a peak in the masking function obtained for the masker frequency of 5 cycles/octave. The error bar in Fig. 5(D) represents the standard error of the mean averaged across all subjects and conditions. Although a low-pass function was seen for the two subjects S1 and S2 for 3 cycles/octave masker modulation, that pattern did not generalize across masker modulation frequencies for the 6400–12 800 carrier and did not generalize across all carrier bandwidths for those subjects. The reason for this anomalous result is unknown. The amount of masking (dB), dynamic range (dB), bandwidth (oct), and the Q factor for the spectral modulation frequencies of 1, 3, and 5 cycles/octave are summarized in Table II. For subjects S1 and S2 the bandwidth could not be computed for the low-pass masking patterns obtained for the maskermodulation of 3 cycles/octave. For the masker modulations of 1 and 5 cycles/octave the bandwidth could not be calculated for the limited set of data collected.

IV. DISCUSSION

A. Tuning to spectral modulation frequency

The results indicated that a peak occurs in the modulation masking patterns when the signal and the masker are similar in spectral modulation frequency. The amount of masking decreases as the signal and the masker modulation

frequencies diverge. There are several possible explanations for the tuning to spectral modulation frequency demonstrated here. At the physiological level, similar tuning to spectral modulation frequency has been reported in the cells of the ferret auditory cortex (e.g., Shamma *et al.*, 1995). These behavioral results, combined with physiological studies by Shamma and colleagues, support the hypothesis that families of cells in the central auditory system tuned to different spectral modulations function as channels tuned to spectral modulation frequency.

Similar to the notion of channels in the spectral modulation domain, the auditory system is known to analyze complex temporal envelopes via channels tuned to temporal modulation frequencies. The evidence for frequency selectivity in the temporal modulation domain was initially established in the masking experiments performed by Bacon and Grantham (1989) and Houtgast (1989). Bacon and Grantham (1989) reported the detection of temporal modulation for frequencies between 2 and 512 Hz in the presence of masker modulation at either 4, 16, or 64 Hz superimposed upon a broadband noise carrier. Their results show low-pass tuning for the masker modulation of 4 Hz and bandpass tuning for the masker modulation frequencies of 16 and 64 Hz. Furthermore, an increase in masking was obtained with an increase in the masker modulation depth. Similar bandpass tuning to various masker modulation frequencies was documented by Houtgast (1989). Another experimental paradigm used to demonstrate modulation-frequency-specific channels in the auditory system is selective adaptation to temporal modulation frequencies. Tansley and Suffield (1983) reported a reduction in the sensitivity to temporal modulation following prolonged exposure (20–30 min) to temporal and frequency modulation. Similar results have been reported in the adaptation studies performed by other investigators (Green and Kay, 1973; Regan and Tansley, 1979; Tansley and Regan, 1979; Tansley *et al.*, 1982). The prolonged exposure to a particular temporal modulation frequency results in an increase in the temporal modulation detection threshold at or near the adaptation frequency. A comparison of the TMTFs measured before and after adaptation to a particular temporal modulation frequency reveals a depression in the TMTF near the frequency of the adapting stimulus.

An alternative explanation for the apparent tuning to spectral modulation frequency shown in Fig. 2 is that listeners simply rely on the overall change in the spectral modulation depth to detect signal modulation in the presence of masker modulation. Assuming that the excitation pattern provides a reasonable estimate of the internal representation of spectral shape, the change in spectral modulation depth preserved by the auditory system that results from adding signal to masker modulation may be gauged by computing the spectral modulation depth in the excitation pattern for the masker alone (standard) and for the masker-plus-signal (signal) stimuli. To determine whether or not a simple change in the overall excitation pattern might explain the pattern of thresholds shown in Fig. 2, excitation patterns were computed for each condition in the present experiment based on the model first proposed by Moore and Glasberg (1987) and later modified by Moore and Glasberg (2004) in the context of their loudness pattern model. This model assumes that the auditory periphery can be represented by a bank of overlapping bandpass filters, the characteristics of which were determined on the basis of the results of several investigations of auditory filter shape using the notched noise masking paradigm (e.g., Patterson, 1976).

With the same stimulus generation software used in the main experiments, average stimulus spectra were computed based on 100 signal and masker samples for each of the 27 signal and masker modulation frequency combinations used with the 200–12 800 Hz carrier as well as for the standard condition with only masker modulation. The spectral modulation depth of the signal frequency was equal to the masked signal threshold averaged across the four subjects. In the actual experiment, the signal and masker modulation starting phases were random with respect to each other and from interval to interval. However, acoustical analyses based on 100 random signal and masker phase combinations revealed minimal variation in overall spectral modulation depth (standard deviation=0.16 dB). Therefore, to reduce computation time, average spectra for a given condition were computed only for five different randomly selected signal and masker phase combinations. Excitation patterns were estimated over the range of 3–40 ERBs or roughly 90–16 800 Hz. The computations included a correction for the middle-ear transfer function, but did not include correction for the earphone transfer function. The internal representation of spectral modulation was indexed by computing the maximum modulation depth (local peak-to-valley difference in decibels) over the range of audio frequencies between the nominal lower and upper cutoff frequencies of the carrier condition. The maximum modulation depth for a given signal/masker modulation frequency combination was averaged across the five random phase conditions.

To assess the change in modulation depth associated with the addition of signal modulation depth at threshold, first the modulation depth preserved in the excitation patterns associated with the masker alone (standard) was computed for each masker modulation frequency (1, 3, and 5 cycles/octave). Next, the modulation depth preserved in the excitation patterns associated with the signal-plus-masker (signal) conditions was computed. Finally, the difference or change in modulation depth associated with the addition of the signal to the masker was computed by subtracting the modulation depth associated with the masker alone from the modulation depth associated with the signal-plus-masker stimuli. The results of these computations are summarized in Fig. 6, where the change in modulation depth computed from the excitation patterns (in dB) is shown on the ordinate as a function of the signal modulation frequency relative to the masker modulation frequency (signal divided by masker frequency) on the abscissa. Each curve represents a separate masker modulation frequency.

If masked spectral modulation detection thresholds (e.g., Fig. 2) were determined by a simple change in the overall modulation depth preserved in the excitation pattern, then one would expect the functions to be flat, corresponding to a constant change in the excitation pattern as a function of spectral modulation frequency. These functions demonstrate

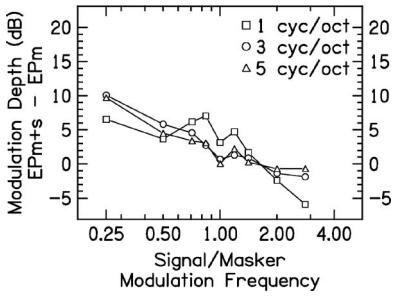


FIG. 6. Change in modulation depth (dB) as a function of the signal modulation frequency relative to the masker modulation frequency (signal divided by masker frequency). The modulation depth associated with the excitation pattern of masker alone (EP_m) was subtracted from the modulation depth associated with the excitation pattern of signal-plus-masker (EP_{m+s}) stimuli to obtain the change in modulation depth (dB).

that the addition of signal modulation at masked modulation detection threshold to masker modulation with a modulation depth of 15 dB does not result in a constant change in the modulation depth preserved by the excitation pattern. Indeed, the amount of change in the modulation depth in the excitation pattern is a complex function of both the signal and the masker modulation frequencies. Thus, one can conclude that the apparent tuning to spectral modulation revealed by the spectral modulation masking paradigm, as shown in Fig. 2, does not simply reflect a change in the excitation pattern associated with the addition of signal to masker modulation.

B. Variations in tuning characteristics

For most carrier conditions, spectral modulation frequencies, and subjects, a single bandpass tuning characteristic was obtained. However, variations in the masking patterns were obtained for some subjects and conditions. These deviations from the typical masking patterns obtained for some of the experimental conditions are as follows.

Occasionally, a low-pass masking pattern was obtained. These masking patterns showed a greater elevation for lower modulation frequencies than for the higher modulations and had no distinct peak in the masking function. Similar low-pass masking patterns have been reported for low modulation frequencies in the temporal domain (e.g., Bacon and Grantham, 1989). Although it is difficult to explain the exact underlying mechanism for this phenomenon, some speculations can be made based on the physiological findings in cells of the ferret auditory cortex. Shamma *et al.* (1993) suggest that, based on the inhibitory and excitatory responses of the cells, a variety of response patterns (i.e., high pass or low pass) can be obtained. This may be attributed to different cortical cell types and morphologies (Shamma *et al.*, 1993). Behavioral manifestations of these patterns would imply a population of cells dominated by the responses of such cell types.

In addition to the above-mentioned alteration in the masking patterns, bandpass masking functions with relatively broad passbands were obtained on some experimental conditions for some of our subjects. In this case, the masker modulation interfered with the detection of signal modulation when the two were similar, but there was a spread of masking to the neighboring frequencies. The center frequency of the passband was similar or near to the masker modulation frequency. Thus, there appears to be a wide range of characteristic bandwidths associated with tuning to spectral modulation. It is possible that these variations in tuning reflect variations in complex inhibitory and excitatory sidebands of cells sensitive to spectral modulation (Shamma *et al.*, 1995).

The results obtained in this study indicate a local maximum in the masking patterns when the signal and the masker were similar in their spectral modulation frequency. Along with this primary peak, the masking patterns show secondary peaks at various modulation frequencies. For example, masking patterns obtained with a 5 cycles/octave masker and a 200–12 800 Hz carrier showed a primary peak for the signal frequency of 5 cycles/octave and a secondary peak at 2.5 cycles/octave. Similarly, when the masker modulation frequency was 3 cycles/octave, a secondary peak occurred at 6 cycles/octave. Neither the cause nor the significance of the occurrence of secondary peaks is understood at this time, although similar secondary peaks have been reported in other modulation masking experiments (e.g., DeValois and Tootell, 1983; Bacon and Grantham, 1989).

C. Effect of carrier bands

Masking patterns were obtained for the three different carrier bands (200–12 800, 800–1600, and 6400–12 800 Hz). The broadband condition was designed to investigate the processing of spectral features across most of the functional hearing range of an individual. The two 1 octave carrier bands were used to estimate spectral shape perception near the two ends of the functional hearing range. The unmasked SMTFs obtained for the two 1 octave carrier bands (800–1600 and 6400–12 800 Hz) were similar to the unmasked SMTFs obtained for the broadband condition (200–12 800 Hz). This indicates that spectral modulation detection is similar across the functional audio-frequency range, consistent with the results of Eddins and Bero (2007). In contrast, Moore *et al.* (1989) reported that the detection of a single spectral notch in a broadband noise (i.e., spectral decrement detection) was worse for notches centered at 8000 than 1000 Hz, while the detection of a spectral peak did not vary markedly with center frequency. The three peaks or notch widths (0.125, 0.25, and 0.5 fc) at center frequencies (fc) of either 1000 or 8000 Hz corresponded roughly to a half cycle of square wave modulation with a fundamental frequency of 1.36, 2.76, or 5.54 cycles/octave. Although comparisons with the present data are not simple, the general lack of a strong frequency effect in the present data is indicative of a fundamental difference between the detection of a single, rather sharp spectral notch, and the detection of spectral modulation.

In general, the masking patterns obtained across the three carrier bands were similar to each other for all subjects. It is interesting to note, however, that the masking patterns obtained for subjects S1 and S2 deviated in several ways from the other two subjects. First, they both demonstrated mistuning, as revealed by the masking patterns obtained for the masker modulation frequency of 1 cycle/octave. For

these same subjects more masking was evident at 2.5 cycles/octave than at 5 cycles/octave for the masker modulation of 5 cycles/octave (carrier band: 200–12 800 Hz). Furthermore, the masking patterns obtained for the 3 cycles/octave (carrier band: 6400–12 800 Hz) masker for these two subjects showed a low-pass function, whereas the other two subjects showed a distinct peak in the masking function. On the other hand, all the subjects, including S1 and S2, showed a peak at 3 cycles/octave for the masker modulation of 3 cycles/octave and a carrier band of 200–12 800 Hz or 800 to 1600 Hz. Thus, subjects S1 and S2 show variability in the masking functions across different modulation frequencies and carrier conditions. These differences indicate variation in the processing of the spectral envelopes across individuals.

V. CONCLUSIONS

To better understand the nature of spectral envelope perception, a masking paradigm was used to test the hypothesis that auditory processing of spectral shape reveals tuning to spectral modulation frequency. Unmasked SMTFs revealed the greatest sensitivity to spectral modulation in the range of 1–3 cycles/octave, consistent with several previous investigations. Using a spectral modulation masking technique, the resulting spectral modulation masking patterns reveal distinct tuning to spectral modulation frequency. This tuning is consistent with the hypothesis that spectral envelope patterns are processed by channels tuned to spectral modulation frequency. The results obtained in the different experiments reported here indicate that the perception of spectral modulation is similar in many respects to frequency selectivity in the audio-frequency domain and the auditory temporal domain. The study of the internal representation of simple sinusoidal spectral patterns may help to predict the internal representation of arbitrary spectral envelopes on the basis of a few simple measurements and may provide insight into the rules governing, and the mechanisms underlying, spectral envelope perception.

ACKNOWLEDGMENTS

This work was supported by NIH NIDCD R01 DC04403 and NIA P01 AG09524. The authors wish to express their sincere appreciation for the helpful comments of Dr. Frederic Wightman, an anonymous reviewer, and the associate editor.

Algazi, V. R., Duda, R. O., Morrison, R. P., and Thompson, D. M. (**2001**). "The CIPIC HRTF database," in *WASPAA'01, Proceedings of the 2001 IEEE AASP Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, pp. 99–102.

Amagai, S., Dooling, R. J., Shamma, S., Kidd, T. L., and Lohr, B. (**1999**). "Detection of modulation in spectral envelopes and linear-rippled noises by budgerigars (Melopsittacus undulatus)," J. Acoust. Soc. Am. **105**, 2029–2035.

American National Standards Institute (ANSI). ANSI S3.6–1996, (**1996**). "Specification for audiometers," (American National Standards Institute, New York).

Bacon, S. P., and Grantham, D. W. (**1989**). "Modulation masking: Effects of modulation frequency, depth, and phase," J. Acoust. Soc. Am. **85**, 2575–2580.

Bernstein, L. R., and Green, D. M. (**1987a**). "The profile-analysis bandwidth," J. Acoust. Soc. Am. **81**, 1888–1895.

Bernstein, L. R., and Green, D. M. (**1987b**). "Detection of simple and complex changes of spectral shape," J. Acoust. Soc. Am. **82**, 1587–1592.

Bernstein, R. S., and Raab, D. H. (**1990**). "The effects of bandwidth on the detectability of narrow- and wide-band signals," J. Acoust. Soc. Am. **88**, 2115–2125.

Blandon, R. A. W., and Lindblom, B. (**1981**). "Modeling the judgment of vowel quality differences," J. Acoust. Soc. Am. **69**, 1414–1422.

Bucklein, R. (**1981**). "The audibility of frequency response irregularities," J. Audio Eng. Soc. **29**, 126–131.

Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (**1999**). "Spectro-temporal modulation transfer functions and speech intelligibility," J. Acoust. Soc. Am. **106**, 2719–2732.

Costalupes, J. A., Young, E. D., and Gibson, D. J. (**1984**). "Effects of continuous noise backgrounds on rate response of auditory nerve fibers in cat," J. Neurophysiol. **51**, 1326–1344.

DeValois, K. K., and Tootell, R. B. H. (**1983**). "Spatial frequency specific inhibition in cat striate cortex cells," J. Physiol. (London) **336**, 359–376.

Eddins, D. A., and Bero, E. (**2007**). "Spectral modulation detection as a function of modulation frequency, carrier bandwidth, and carrier frequency region," J. Acoust. Soc. Am. **121**, 363–372.

Eggermont, J. J. (**1994**). "Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. Effects of carrier type, modulating waveform and intensity," Hear. Res. **74**, 51–66.

Evans, E. F. (**1975**). "The sharpening of cochlear frequency selectivity in the normal and abnormal cochlea," Audiology **14**, 419–442.

Evans, E. F. (**1992**). "Auditory processing of complex sounds: An overview," Philos. Trans. R. Soc. London, Ser. B **336**, 295–306.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–61.

Frisina, R. D., Smith, R. L., and Chamberlain, S. C. (**1990**). "Encoding of amplitude modulation in the gerbil cochlear nucleus. I. A hierarchy of enhancement," Hear. Res. **44**, 99–122.

Grantham, D. W., and Bacon, S. P. (**1988**). "Detection of increments and decrements in modulation depth of SAM noise," J. Acoust. Soc. Am. **84**, S140.

Green, G. G., and Kay, R. H. (**1973**). "The adequate stimuli for channels in the human auditory pathways concerned with the modulation present in frequency-modulated tones," J. Physiol. (London) **234**, 50P–52P.

Green, D. M., Kidd, G., Jr., and Picardi, M. C. (**1983**). "Successive versus simultaneous comparison in auditory intensity discrimination," J. Acoust. Soc. Am. **73**, 639–643.

Greenwood, D. D. (**1961**). "Auditory masking and the critical band," J. Acoust. Soc. Am. **33**, 484–501.

Hamilton, P. M. (**1957**). "Noise masked thresholds as a function of tonal duration and masking noise bandwidths," J. Acoust. Soc. Am. **29**, 506–511.

Hillier, D. A. (1991). "Auditory processing of sinusoidal spectral envelopes," dissertation, Sever Institute of Technology, Washington University, St. Louis, MO.

Houtgast, T. (**1989**). "Frequency selectivity in amplitude-modulation detection," J. Acoust. Soc. Am. **85**, 1676–1680.

Kowalski, N., Depireux, D. A., and Shamma, S. A. (**1996a**). "Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra," J. Neurophysiol. **76**, 3503–3523.

Kowalski, N., Depireux, D. A., and Shamma, S. A. (**1996b**). "Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra," J. Neurophysiol. **76**, 3524–3534.

Møller, A. R. (**1976**). "Dynamic properties of primary auditory fibers compared with cells in the cochlear nucleus," Acta Physiol. Scand. **98**, 157–167.

Moore, B. C. J., and Glasberg, B. R. (**1987**). "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns," Hear. Res. **28**, 209–225.

Moore, B. C. J., and Glasberg, B. R. (**2004**). "A revised model of loudness perception applied to cochlear hearing loss," Hear. Res. **188**, 70–88.

Moore, B. C. J., and Ohgushi, K. (**1993**). "Audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **93**, 452–461.

Moore, B. C. J., Oldfield, S. R., and Dooley, G. J. (**1989**). "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," J. Acoust. Soc. Am. **85**, 820–836.

Musicant, A. D., and Butler, R. A. (**1985**). "Influence of monaural spectral cues on binaural localization," J. Acoust. Soc. Am. **77**, 202–208.

Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640–654.

Peterson, G. E., and Barney, H. L. (**1952**). "Control method used in a study

of the vowels," J. Acoust. Soc. Am. **24**, 175–184.
Regan, D., and Tansley, B. W. (**1979**). "Selective adaptation to frequency-modulated tones: Evidence for an information-processing channel selectively sensitive to frequency changes," J. Acoust. Soc. Am. **65**, 1249–1257.
Rees, A., and Møller, A. R. (**1987**). "Stimulus properties influencing the response of inferior colliculus neurons to amplitude-modulated sounds," Hear. Res. **27**, 129–143.
Rees, A., and Palmer, A. R. (**1989**). "Neuronal responses to amplitude-modulated and pure-tone stimuli in the guinea pig inferior colliculus, and their modification by broadband noise," J. Acoust. Soc. Am. **85**, 1978–1994.
Rhode, W. S., Geisler, C. D., and Kennedy, D. T. (**1978**). "Auditory nerve fiber response to wide-band noise and tone combinations," J. Neurophysiol. **41**, 692–704.
Schooneveldt, G. P., and Moore, B. C. (**1989**). "Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth, and signal duration," J. Acoust. Soc. Am. **85**, 273–281.
Schreiner, C. E., and Mendelson, J. R. (**1990**). "Functional topography of cat primary auditory cortex: Distribution of integrated excitation," J. Neurophysiol. **64**, 1442–1459.
Schreiner, C. E., and Sutter, M. L. (**1992**). "Topography of excitatory bandwidth in cat primary auditory cortex: Single-neuron versus multiple-neuron recordings," J. Neurophysiol. **68**, 1487–1502.
Schreiner, C. E., and Calhoun, B. M. (**1995**). "Spectral envelope coding in cat primary auditory cortex: Properties of ripple transfer functions," Aud. Neurosci. **1**, 39–61.
Schreiner, C. E., and Urbas, J. V. (**1986**). "Representation of amplitude modulation in the auditory cortex of the cat. I. Anterior auditory field," Hear. Res. **21**, 227–241.
Schreiner, C. E., and Urbas, J. V. (**1988**). "Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields," Hear. Res. **32**, 49–64.
Shamma, S. A., Fleshman, J. W., Wieser, P. R., and Versnel, H. (**1993**). "Organization of response areas in ferret auditory cortex," J. Neurophysiol. **69**, 367–383.
Shamma, S. A., and Versnel, H. (**1995**). "Ripple analysis in ferret primary auditory cortex. II. Prediction of unit responses arbitrary spectral profiles," Aud. Neurosci. **1**, 255–270.
Shamma, S. A., Versnel, H., and Kowalski, N. (**1995**). "Ripple analysis in ferret primary auditory cortex. I. Response characteristics of single units to sinusoidally rippled spectra," Aud. Neurosci. **1**, 233–254.
Spiegel, M. F. (**1981**). "Thresholds for tones in maskers of various bandwidths and for signals of various bandwidths as a function of signal frequency," J. Acoust. Soc. Am. **69**, 791–795.
Spiegel, M. F., and Green, D. M. (**1982**). "Signal and masker uncertainty with noise maskers of varying duration, bandwidth and center frequency," J. Acoust. Soc. Am. **71**, 1204–1210.
Summers, V., and Leek, M. R. (**1994**). "The internal representation of spectral contrast in hearing-impaired listeners," J. Acoust. Soc. Am. **95**, 3518–3528.
Tansley, B. W., and Regan, D. (**1979**). "Separate auditory channels for unidirectional frequency modulation and unidirectional amplitude modulation," Sens Processes **3**, 132–140.
Tansley, B. W., Regan, D., and Suffield, J. B. (**1982**). "Measurement of the sensitivities of information-processing channels for frequency change and for amplitude change by a titration method," Can. J. Psychol. **36**, 723–730.
Tansley, B. W., and Suffield, J. B. (**1983**). "Time course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation," J. Acoust. Soc. Am. **74**, 765–775.
Versnel, H., Kowalski, N., and Shamma, S. A. (**1995**). "Ripple analysis in ferret primary auditory cortex. III. Topographic distribution of ripple response parameters," Aud. Neurosci. **1**, 271–285.
Versnel, H., and Shamma, S. A. (**1998**). "Spectral-ripple representation of steady-state vowels in primary auditory cortex," J. Acoust. Soc. Am. **103**, 2502–2514.
Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.
Wakefield, G. H., and Viemeister, N. F. (**1990**). "Discrimination of modulation depth of sinusoidal amplitude modulation (SAM) noise," J. Acoust. Soc. Am. **88**, 1367–1373.
Young, E. D., and Barta, P. E. (**1986**). "Rate responses of auditory nerve fibers to tones in noise near masked threshold," J. Acoust. Soc. Am. **79**, 426–442.
Yost, W. A., Sheft, S., and Opie, J. (**1989**). "Modulation interference in detection and discrimination of amplitude modulation," J. Acoust. Soc. Am. **86**, 2138–2147.

Theory construction in auditory perception: Need for development of teaching materials

Nathaniel I. Durlach
Hearing Research Center, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

Frederick J. Gallun
VA RR&D National Center for Rehabilitative Auditory Research, 3710 SW U.S., Veterans Hospital Road, Portland, Oregon 97239

(Received 22 May 2007; revised 31 May 2007; accepted 11 June 2007)

Beginning investigators in the field of auditory perception receive essentially no instruction in how to go about constructing new and useful theories. This article considers some of the characteristics of good theory in this field and is intended to serve as a call for further discussion of the processes by which such theory is created and for the development of appropriate theory-construction teaching materials. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2754060]

PACS number(s): 43.66.Ba [RAL] Pages: 1014–1016

I. INTRODUCTION

Almost all courses on auditory perception include material concerned with the design of experiments (often packaged under the heading "Psychophysical Methods"). In contrast, to the best of the writers' knowledge, very few courses include material on the design of theories. In this article, an attempt is made to stimulate discussion that might lead to the development of such material. Ideally, such material would not only improve the teaching of theory construction, but also theory construction itself.[1]

As one might expect, there is a considerable amount of material discussing scientific theory and theory construction in writings concerned with the philosophy and/or history of science (e.g., Bechtel, 1988; Kuhn, 1970; Langley *et al.*, 1987; Popper, 1968; Reynolds, 1971; Wertheimer, 1959). However, much of this material, as well as the material in the *Journal of Theory Construction and Testing*, appears to be of relatively little value for our purposes either because it sheds little light on how one should go about attempting to create useful new theories or because the kinds of theories focused upon are poorly matched to the kinds that are (or would be) most useful in the domain of auditory perception. In our opinion, the strongest exceptions to this statement can be found in the book *A Primer in Theory Construction* by Reynolds (1971). Not only does Reynolds provide a general overview of different kinds of theories, but he addresses head on (and in a highly readable manner) issues related to the *creation of new theories* and to the acquisition of scientific *understanding* (in his terms, "understanding" is most likely to occur when the theory in question combines the "axiomatic form" and the "causal process form").

In this article, we make some general comments on the nature of theory and on dimensions of theory goodness, then consider some paths to theory construction that we believe are particularly relevant to the field of auditory perception, and then end with a few concluding remarks concerned with the development of theory-construction teaching materials. Consistent with the lack of any need in this article to distinguish between the terms "theory" and "model," we use the term "theory" throughout.

II. ON THE NATURE OF THEORY AND THE DIMENSIONS OF THEORY GOODNESS

As indicated in the above-mentioned references, there are a variety of conceptual structures that can be referred to as theories. In this article, unless indicated otherwise, the word "theory" refers to a conceptual structure that is created by identifying the elemental real-world entities of interest with the primitives (i.e., undefined terms) of some mathematical system. In general, a mathematical system includes, in addition to the primitives, axioms relating these primitives and theorems that can be derived from these axioms by deductive reasoning. Once the identification between primitives and real-world entities has been made, the axioms and theorems of the mathematical system automatically become statements about the real world, the mathematical system has been *applied*, and a *real-world theory* has been created.[2,3] Examples of such structures are given by the application of Euclidean geometry (or one of the non-Euclidean geometries) to the physical space surrounding us, the application of mathematical group theory to sets of physical transformations (where the binary operator postulated in the definition of a group is realized by the composition of transformations), and, closer to home, Helmholtz's application of the mathematics associated with vibratory oscillations to the behavior of the basilar membrane and to our auditory sensations (Helmhotz, 1954).

In the pure-mathematics domain, questions about truth are meaningless. Assuming the axioms are consistent (so that no internal contradictions can arise), one can only consider issues of parsimony, power, elegance, beauty, etc. In contrast, in the applied mathematics domain, questions about truth, specifically, questions concerning the extent to which the statements about the real-world entities are consistent with empirical evidence, are central.

Finally, it is worth noting that even though much of the work in pure mathematics has been motivated by purely aes-

 0001-4966/2007/122(2)/1014/3/$23.00

thetic considerations and without any application in mind, it is an historical fact that most mathematical systems (including, for example, non-Euclidean geometry, imaginary numbers, and transfinite arithmetic) have, subsequent to their creation, found important application in real-world theory.[4]

Apart from the extent to which a real-world theory is formalized in the sense indicated above, there are many dimensions along which it can be rated as good or bad. The most important class of dimensions presumably relates to the extent to which the theory is true, i.e., describes the real world. One factor here concerns the degree of correctness or precision of match between theory and empirical data (observations and/or experimental results); a given match can be approximate or arbitrarily precise. A second factor concerns the range of predictions; the set of predictions that can be derived from the theory can be very extensive or very limited (with the accuracy of the predictions varying more or less widely across the set). A third factor concerns the ease with which the predictions can be derived; such derivations can be simple and straightforward or complex and time consuming. A fourth factor concerns the extent to which the theory leads the empirical findings or merely follows them. In other words, to what extent does the theory predict phenomena that have never previously been measured, or observed, or even contemplated? A second major class of dimensions concerns the extent to which the theory provides understanding of the domain to which the theory is addressed. Although there exists significant correlation between predictive power (accuracy and range) on the one hand, and basic scientific understanding on the other hand, they are by no means the same thing. In particular, it is sometimes possible to build an entity with substantial predictive power about a specific set of empirical phenomena without understanding these phenomena or even without understanding the entity one has constructed. In contrast, it is impossible to have complete understanding and not have predictive power. In other words, understanding implies predictive power, but predictive power does not necessarily imply understanding. Further factors within this second class of dimensions concern the benefits of understanding beyond those related to prediction. Although one cannot substitute "I understand" for "I think" in Descartes' "I think, therefore I am," understanding is a core characteristic of human existence. At the very least, understanding, like listening to a beautiful piece of music or looking at a beautiful painting, provides great pleasure. A third class of dimensions concerns the relations of the given theory to other theories. Is the given theory relatively unique in the sense that there are no other theories that can explain the phenomena in question? Similarly, to what extent does the theory in question serve as a stimulant not only to further exploration in the domain of the theory (both theoretical and experimental), but also to exploration and changed thinking in other domains?

III. THEORY CONSTRUCTION IN THE DOMAIN OF AUDITORY PERCEPTION

All factors relevant to theory construction in general apply to theory construction in the case of auditory perception. Two important special factors that shape theory construction in this case, however, concern (1) the role of the auditory system as a sensor and processor of acoustic environmental signals and (2) the make-up of the auditory system as a very complex hierarchical biological system with many feedback and feedforward subsystems.

Included among the features of theory construction associated with these two special factors are the following. First, in the attempt to model the detailed components of the system, there is often heavy involvement in the biophysics of these components. Second, because the system is so complex, true theory construction is sometimes replaced by simulation activities. The aim in these activities is to develop a simulator (perhaps a neural net) and a simulator training technique that results in behavior that matches that of the target. Frequently, the match is achieved with only limited understanding of how either the target or the trained simulator actually work. Taken to the limit, claiming that such a procedure leads to a scientific theory is equivalent to claiming that one had constructed a theory of a particular component of human behavior by creating a baby and training that baby as it matured to exhibit that behavior. Although theory construction by simulation is still not very common, its use is likely to grow in the future as a result of the increasing complexity of the material considered and the increased power of computers.

In most cases of interest, theory construction in the area of auditory perception makes use of five main inputs. One is the wealth of information continuously evolving about the behavior of various types of physiological elements in various animals in various states in response to various types of stimuli. The second is the ever-growing collection of psychoacoustic data resulting from objective tests of human auditory performance (that may involve substantial sensory, perceptual, and cognitive elements). The third is the knowledge available on other sensory systems, particularly vision. The fourth consists of the scientist's own subjective experience as an experimental subject in tests of auditory performance. Finally, the fifth input is derived by determining to the extent possible what processing by the listener's auditory system would be most useful (optimal performance plus robustness) in performing the given task in order to achieve the specified goal under various types of constraints.

Variations in overall approach to theory construction result not only from variations in the inputs considered, but also from the way in which the inputs are weighted and integrated. At one extreme, for example, attention is focused almost exclusively on the first input (on the physiology) with the other inputs being seriously downplayed. At another extreme, attention is focused strongly on the fifth input (on what the system "should do" to achieve effective and robust processing) and the model is of the "black-box" type.[5] Ideally, theories should take serious account of all inputs.

Another special feature of theory construction in auditory perception arises because of the substantial variations in the subjects studied. This variation can become evident in a single subject as a function of time (arising from fatigue or learning) or in the testing of different subjects. This problem has become increasingly prominent as the issues being probed become more cognitive and higher levels of the sys-

tem are involved. In some cases, intersubject variation is avoided by throwing the poor performing subjects out of the study (i.e., the scientist wants to focus on the upper limits of performance). In other cases, the theory must either have enough flexibility (i.e., enough fitting parameters) to encompass the intersubject variability, or there must be a separate theory created for each subject. Clearly, both intrasubject variation and intersubject variation greatly complicate the problem of theory construction in the domain of auditory perception.

IV. DEVELOPMENT OF TEACHING MATERIALS FOR COURSES IN AUDITORY PERCEPTION

Independent of precisely how the issues related to theory construction and theory goodness are identified and addressed, there remains the question of how best to develop teaching materials that can be usefully incorporated in courses on auditory perception. At present, we believe that such materials could best be developed by means of a cooperative effort among auditory scientists who have themselves developed useful theories in which each such scientist

(a) attempts to articulate and describe the processes gone through in the creation of the theory in question,
(b) communicates the resulting information about these processes via a note to JASA and/or a special session arranged for some future ASA meeting, and
(c) participates in the preparation of a report on theory construction in the domain of auditory perception based on the results of the above work.

Ideally, the resulting report would not only contain material directly relevant to the task of theory construction, but also information related to how this material could best be integrated into various types of auditory-science courses.

We would greatly appreciate hearing from anyone who might be interested in the above-described effort. Also, of course, if we are incorrect in our assessment of the extent to which theory construction is currently being taught in courses on auditory perception, we would greatly appreciate hearing about it.

ACKNOWLEDGMENTS

We are deeply indebted to our associate Chris Mason, to the Associate Editor of JASA, and to the reviewers of this article for many useful comments on drafts of this paper. This work was supported by AFOSR Grant No. FA9950-05-1-2005 and NIH/NIDCD Grants Nos. DC04545 and F32 DC006526. Frederick Gallun was also supported by the National Center for Rehabilitative Auditory Research through VA RR&D Award No. C4855H.

[1]It is assumed here that theory construction, in the classical sense, will survive. Based on the general changes currently taking place in society (e.g., the increasing severity of the time famine, the increased reliance on computation as opposed to thought, etc.), it is not at all obvious that this assumption is correct.

[2]We have used the term "*mathematical system*" rather than "*mathematical theory*" so as to avoid confusion over the word "*theory*." As indicated, our attention in this article is focused on *real-world theories*, not on the mathematical systems (i.e., mathematical theories) used in the construction of the real-world theories.

[3]Given the meaning assigned to the word "theory" in this article, it is interesting to consider what meaning should be assigned to the term "*quantitative theory*." On the one hand, to the extent that all the real-world theories considered are derived (by definition) from mathematical systems, one could refer to all of them as quantitative theories. On the other hand, it might make more sense to restrict use of the term "quantitative theory" to cases in which the mathematical system in question is concerned in some way with numbers.

[4]This very remarkable historical fact implies either (1) that our thinking is so strongly dominated by our experiences in the outside objective world that we are totally incapable of creating anything really new or (2) that our perception of the world is so dominated by our characteristics as perceivers (as opposed to the characteristics of the world) that we are totally incapable of perceiving the real world.

[5]Consideration of what the system *should* do clearly makes sense in a scientific environment like that of audition (or other sensory-perceptual systems) where notions of ideal processing are well established and one can determine not only the operations performed by the ideal processor but also the results of these operations [see, for example, the work of Siebert (1968) and Colburn (1973)]. To what extent such considerations can be applied to fields nominally not focused on signal processing is an interesting open question. For example, is there any way of looking at the physical universe that would enable such questions as "What should the laws of gravity be" to make sense?

Bechtel, W. (**1988**). *Philosophy of Science: An Overview for Cognitive Science* (Erlbaum, Hillsdale, NJ).
Colburn, H. S. (**1973**). "Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination," J. Acoust. Soc. Am. **54**, 1458–1470.
Helmholtz, H. von (**1954**). *On the Sensations of Tone* (translation by A. J. Ellis) (Dover, New York).
Kuhn, T. S. (**1970**). *The Structure of Scientific Revolutions* (Chicago U. P., Chicago, IL).
Langley, P., Simon, H. A., Bradshaw, G. L., and Zytkow, J. M. (**1987**). *Scientific Discovery: Computational Explorations of the Creative Process* (MIT, Cambridge, MA).
Popper, K. R. (**1968**). *The Logic of Scientific Discovery* (Hutchinson, London).
Reynolds, P. D. (**1971**). *A Primer in Theory Construction* (Prentice Hall, Englewood Cliffs, NJ).
Siebert, W. M. (**1968**). "Stimulus transformations in the auditory system," in *Recognizing Patterns*, edited by P. Kolers and M. Eden (MIT, Cambridge, MA).
Wertheimer, M. (**1989**). *Productive Thinking* (Harper, New York).

Individual differences in source identification from synthesized impact sounds

Robert A. Lutfi[a] and Ching-Ju Liu
Department of Communicative Disorders and Waisman Center, University of Wisconsin, Madison, Wisconsin 53706

(Received 6 July 2006; revised 31 May 2007; accepted 31 May 2007)

Impact sounds were synthesized according to standard textbook equations given for the motion of freely vibrating membranes, bars, and plates. In a two-interval, forced-choice procedure, highly practiced listeners identified from these sounds predefined target sources based on their material and size, the hardness of the striking mallet, and the presence or absence of light damping applied to the center of the source. Listener decision strategy in each case was determined from a discriminant analysis of trial-by-trial responses resulting in a vector of regression weights given to different acoustic parameters. The analysis revealed significant differences in decision strategy across listeners within identification task, but similarity in decision strategy within listeners across variations in task. Only when the acoustic information for identification was highly constrained (identification of damping) did listeners adopt similar decision strategies approaching that of an ideal observer. Despite the large individual differences in decision strategy, identification accuracy was, in most cases, similar across listeners. Where there were differences in identification accuracy the differences appeared largely related to differences in internal noise and not decision strategy. The results are generally comparable to those obtained for the discrimination of arbitrary tone patterns. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2751269]

PACS number(s): 43.66.Ba, 43.66.Fe, 43.66.Jh [AJO] Pages: 1017–1028

I. INTRODUCTION

Impact sounds are ubiquitous in nature and are a vital source of information about objects and events present in our everyday environment. It is no surprise then that studies of sound source identification have largely focused on the ability of listeners to identify simple attributes of sources from the sound of impact (see Lutfi, 2007 for a review). Conclusions of these studies have, for the most part, been based on the analysis of group-mean data. Individual differences in performance when reported have rarely been given much consideration beyond mention and have at times been purposely excluded from analysis by setting arbitrary performance criteria. Yet, from the earliest studies it has been clear that listeners do often perform very differently on source identification tasks, even in those tasks involving the identification of the most rudimentary source attributes.

In one of the earliest studies, Gaver (1988) had listeners judge the size of wood and metal bars from recordings of the bars struck with a soft mallet as they rested on a carpeted surface. The 11 listeners were roughly split as to whether they perceived the wood or metal bars to be longer. Notably, the split made it appear as though the size judgments were independent of material when ratings were averaged across listeners. In a related study, Carello *et al.* (1998) had listeners judge the length of wooden rods from the sounds they produced when dropped to the floor. Though judgments, on average, were positively correlated with actual length, the strength of the relation was reported to vary considerably across listeners. Lakatos *et al.* (1997) examined the perception of the cross-sectional height/width ratios of metal and wooden bars from recordings of impact sounds. They report that several of their listeners consistently reversed labeled shapes, resulting in significantly less than chance performance. Similar incidences of reverse labeling have been reported for the recognition of gender from the sound of hands clapping (Repp, 1987) and for judgments regarding the larger or faster of two wooden balls from the sounds of the balls rolling across a wooden plate (Houben *et al.*, 2005).

In all of these studies little or no information was given beforehand to listeners about the sounds and/or their relation to the sound-producing object or event. Lack of prior information likely increases the chances of observing individual differences, but it is clearly not a necessary condition for obtaining individual differences. Other studies have reported individual differences even when listeners are highly trained in the task and are given correct feedback from trial to trial. This was the case, for example, in a study by Lutfi (2001) where seven highly trained listeners judged from synthetic impact sounds whether the sound source (a bar) was hollow or solid. While identification accuracy was similar for all listeners, a discriminant analysis of the trial-by-trial responses showed listeners to be split in how they approached the task. Roughly half attended to a specific relation between frequency and decay (corresponding to the analytic solution for hollowness), while the remaining half relied on frequency alone. Similar individual differences in listening strategy were obtained in a companion study involving the discrimination of source material (Lutfi and Oh, 1997), and in a separate study involving the discrimination of mallet hardness from impact sounds (Giordano and Petrini, 2003).

[a]Electronic mail: ralutfi@wisc.edu

The finding in these studies that substantially different approaches to a task can yield similar levels of identification accuracy is particularly noteworthy because, until recently, the focus of studies has been largely on identification accuracy. The presumption of studies has been that accurate identification implies the use of a particular decision strategy on the part of the listeners, which is then subsequently analyzed (cf. Cabe and Pittenger, 2000; Carello *et al.*, 1998; Kunkler-Peck and Turvey, 2000; Li *et al.*, 1991; Repp, 1987; Tucker and Brown, 2003; Warren and Verbrugge, 1984). Though the assumption may, in many cases, be correct, the fact that different strategies can yield the same level of performance leaves open the possibility that individual differences in decision strategy may have been overlooked in these studies. Indeed, even when an implicated decision strategy is shown to correlate significantly with listener judgments, as is often the case in these studies, one cannot rule out the possibility that an even stronger correlation might be observed with some other decision strategy not considered. The problem is compounded by the fact that performance levels in these studies are often quite high, which can increase the likelihood there would be more than one viable decision strategy for the task (see Lutfi, 2001 for a detailed discussion of this problem).

Notwithstanding these practical considerations, a compelling reason to investigate individual differences in sound source identification pertains to their theoretical implications. Since Helmholtz (1877), auditory theory has attached great significance to how one's individual experience in the world can shape his/her perception of sounding objects. The theoretical interest has served to motivate a large body of research on the perception of random tone patterns, sounds that do not normally occur in nature and that are unfamiliar to the listener (Watson and Kelly, 1981; Neff and Dethlefs, 1995; Lutfi, 1993; Lutfi *et al.*, 2003). One of the most outstanding features of these studies is the large individual differences observed, both in overall performance and in listening strategy. Such differences are believed to reflect the listener's lack of familiarity with these sounds and the absence of predictable structure that is found in most naturally occurring sounds. Though the idea is now generally accepted, it has not been widely tested. If correct, one should expect greater uniformity in performance and listening strategy across listeners in tasks involving the identification of natural sounds.

The present study was undertaken to gain a better understanding of individual differences in source identification from impact sounds. We pursued three specific goals that were intended to identify the possible causes of individual differences observed in past studies and to permit a closer comparison to studies involving random tone patterns. First, we wished to determine whether there are particular identification tasks involving impact sounds for which individual differences are more likely to be observed. We tested the hypothesis that individual differences are more likely in tasks for which there are multiple potential cues for identification compared to tasks in which the information for identification is highly constrained. Second, we wished to determine whether there are specific ways in which individuals differ consistently from one another across identification tasks. We considered the possibility that listeners have characteristic *listening styles* that generalize across tasks, and that can be used to predict patterns of results across experiments involving the same listeners. Finally, we considered the possible cause of individual differences in performance accuracy. In particular, we evaluated whether such differences result from differences in listener decision strategy or whether they are largely due to other factors (e.g., differences in sensitivity, attentiveness, or memory) not directly tied to the stimulus.

II. GENERAL METHODS

A. Stimuli

In all experiments the stimuli were approximations of the airborne sounds of struck bars, plates, and membranes synthesized according to theoretical equations for the motion of these sources from standard acoustics texts (Fletcher and Rossing, 1991; Morse and Ingard, 1968). Specific details pertaining to the synthesis are provided by Lutfi (2001) and Lutfi and Oh (1997). We chose to use synthetic sounds rather than "live" or recorded sounds because the equations of motion used in the synthesis provide a basis for analyzing *all* relevant sources of information in the sounds. This, as described in Sec. II C, allows listener decision strategy to be analyzed as vector regression weights on individual acoustic parameters and compared to that of a maximum-likelihood detector for each identification task.

For each source, the sound-pressure wave form resulting from the synthesis was a sum of exponentially damped sinusoids whose individual frequencies, ν, amplitudes, A, and decay moduli, τ, were uniquely determined by the specific material and geometric properties of the source, as well as the manner in which the source was struck. Material and size in this synthesis affected the frequency and decay of partials, while mallet hardness and external damping affected the relative amplitude of partials (specific details are given in the description of results from each experiment). The resonant source in different conditions was a circular bar clamped at one end, a loosely hinged circular membrane or a loosely suspended circular plate. For the bar, the ratios of modal frequencies $k=\nu_n/\nu_1$ were 1.00, 6.26, and 17.54; for the membrane they were 1.00, 1.594, 2.136, 2.296, 2.653, and 2.918; and for the plate they were 1.00, 2.80, 5.15, 5.98, 9.75, and 14.09. In Experiments 1 and 3 the amplitudes and the decay moduli of modes varied in inverse proportion to modal frequency, $A_n=A_1/k_n$ and $\tau_n=\tau_1/k_n$. In Experiment 2 the decay moduli were constant across frequency. The values of all other parameters of the stimuli are given for each experiment in Table I. These values were chosen to be typical of sources that might be encountered in everyday listening, but otherwise their selection was arbitrary.

To prevent listeners from simply discriminating a fixed difference between wave forms, a random perturbation in the frequency, amplitude, and/or decay of each partial (depending on the experiment) was introduced on each presentation. This practice was also necessary to obtain regression weights on individual acoustic parameters as described in Sec. II C.

TABLE I. Values of stimulus parameters used in Experiments 1–3. Note that a negative value for Δ indicates that the mean value for the target was smaller than that for the nontarget. See the text for details regarding the values of Δ and σ expressed in JND units.

Experiment	Source	ν_1(Hz)	τ_1(s)	Δ(JNDs)	σ(JNDs)	d'_{ML}
1. Material and size	Bar	250	2.0	5 (ν_i, τ_i)	5 (ν_i, τ_i)	2.4
	Plate	250	0.4	4 (ν_i, τ_i)	5 (ν_i, τ_i)	2.8
	Membrane	500	0.2	4 (ν_i, τ_i)	5 (ν_i, τ_i)	2.8
2. Mallet hardness	Membrane	250	0.4	$-2k_i$ (A_i)	$2k_i$ (A_i)	2.5
3. Point of Contact	Membrane	125	0.4	-6 $(A_{1,4,9})$	3 (A_i)	3.5

The perturbations were added independently to each partial and were the same for all listeners. The distribution of perturbations was normal in log units (i.e., $\log \nu$, $\log A$, and $\log \tau$) with standard deviations expressed in JND units. One JND in this case corresponded to the value $\log(1+\Delta z/z)$, where $\Delta z/z=0.002$, 0.12, and 0.10 is respectively, the Weber fraction for frequency, amplitude, and decay (cf. Wier *et al.*, 1977; Jesteadt *et al.*, 1977; Schlauch *et al.*, 2001). The standard deviations of the perturbations for each experiment are given in Table I. Note that the variation in the wave form resulting from these perturbations is not unlike that which occurs for real impact sounds where inhomogeneities in the material of objects and small variations in how they are struck affect the resulting sound.

Finally, as practical matter, a 5-ms cosine-squared ramp was used to truncate signals after 1 s. This kept trials at a reasonable length, while allowing adequate time for most partials of the sounds to decay to inaudibility. All sounds were played at a 44,100-Hz sampling rate and were delivered over headphones (Beyerdynamic DT 990) without subsequent filtering to the right ear of individual listeners seated in a double-walled, IAC sound-attenuation chamber. Average total sound power at the headphone was calibrated to be approximately 65 dB SPL using a binaural loudness balancing procedure and a TDH-50 earphone with known transfer function.

B. Procedure

A standard two-interval, forced-choice procedure was used. On each trial the listener heard two impact sounds separated by 400 ms. The listener was instructed to select the sound corresponding to a predefined target source, which was the first or second sound with equal probability. In different conditions the target source was identified by its material and size (Experiment 1), by the hardness of the striking mallet (Experiment 2), or by the presence or absence of external damping applied to the surface of the object (Experiment 3). The relevant acoustic information in these conditions was, respectively, a difference in the frequency and decay of partials, the presence or absence of spectral tilt, and the attenuation of specific partials (1, 4, and 9). Listeners were given simple task instructions (e.g., "chose the sound made with the hard mallet") but were not in anyway coached as to what sound quality or cues to listen for. Correct feedback was given after each trial.

Seven normal-hearing female adults (ANSI, 1996), ages 21–30 years, were paid at an hourly rate for their participation. The listeners were students in the Department of Communicative Disorders at the University of Wisconsin - Madison. All had extensive previous experience with the two-interval, forced-choice task and all received at least 400 trials of practice prior to data collection for each condition in which they participated. The data were collected in 1-h sessions, including breaks, conducted on different days. A total of at least 400 experimental trials was run for each listener for each condition. There was no special reason for collecting more than 400 trials except that listener schedules permitted, and this increased the reliability of the data.

C. Discriminant analysis

The decision strategy of each listener was estimated as a vector of regression weights, b_i, using the generalized linear model:

$$\text{logit}[P(R=2)] = b_0 + \sum_{i=1}^{n} b_i(\log x_{i2} - \log x_{i1}) + e, \qquad (1)$$

where $\text{logit}[P(R=2)]$ is the log-likelihood of a second interval response, x_{i2} is the frequency, amplitude, or decay of the ith partial in the second interval, and e is a residual term (cf. Berg, 1990; Anderson, 1971). In Experiment 1 the x_i corresponded to values of frequency and decay, in Experiments 2 and 3 they corresponded to values of amplitude. In all cases, the perturbations in the x_i within and across presentations varied independently of one another. Two independent estimates of the b_i for each listener were obtained by analyzing separately the data for the target in the first and second interval of the trial. The estimates and their associated error were computed using the GLMFIT routine of the software application MATLAB v.6.5. They were then compared to those of a theoretical detector that maximizes the likelihood of a correct response on each trial based on the equations of motion for the resonant source and the trial-by-trial perturbations in acoustic parameters for each condition. Except for Experiment 3 the maximum-likelihood (ML) detector weighted each x_i equally. Overall performance levels of the ML detector for each experiment, d'_{ML}, are given in Table I.

III. RESULTS

A. Individual differences in the identification of size and material

In Experiment 1 the listener identified a prespecified target within a given class (membrane, bar, or plate) based on its particular material and size. The actual material and size

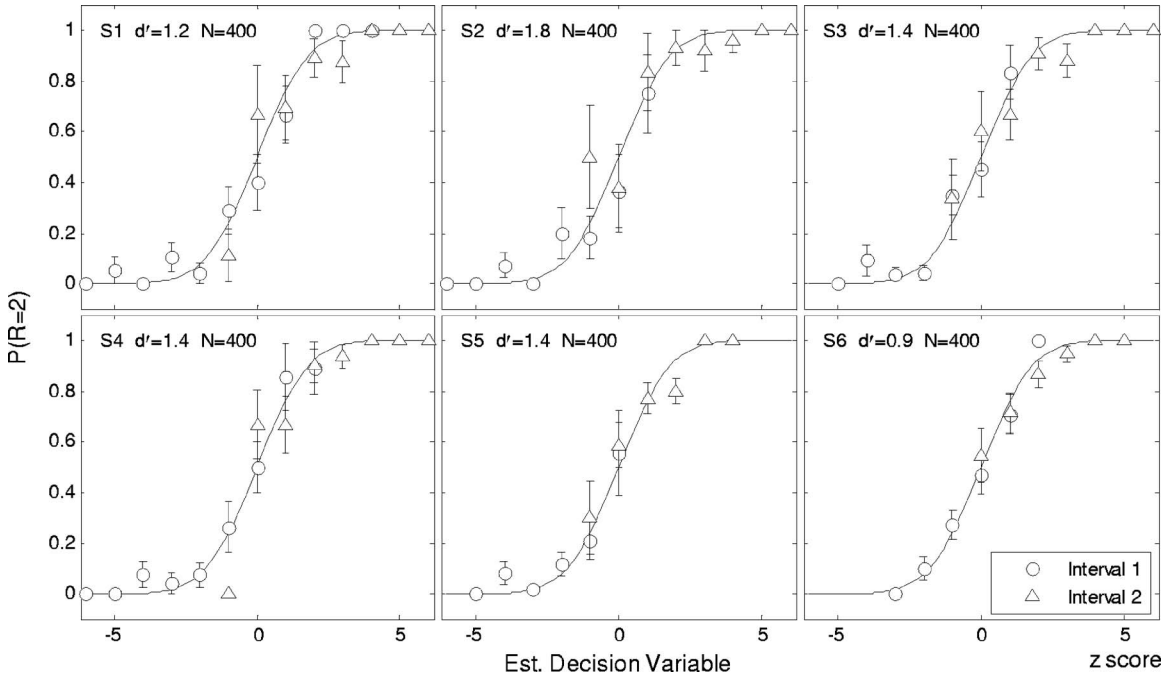


FIG. 1. For each listener (panels) the probability of a second-interval response is plotted as a function of the decision variable estimated from the generalized linear model given by Eq. (1). The data are from Experiment 1 for the case in which the resonant source (target and nontarget) was the plate. Circles and triangles indicate, respectively, trials in which the target was in the first and second interval. Error bars give the standard error of estimate. Predictions of the model, given by the sigmoidal curves, are representative of those obtained for all conditions of the study.

was somewhat arbitrary inasmuch as the simple equations of motion yield identical impact sounds for different combinations of these parameters. Notwithstanding, the target differed acoustically from the nontarget in that the partials were, on average, higher in frequency and decayed more slowly over time (see Table I). In physical terms, this would correspond to the case in which the target was smaller than the nontarget but made of a more dense material, all other factors constant (cf. Morse and Ingard; pp. 175–191).

Figure 1 shows, for each of the six listeners participating in this experiment, the predictions (sigmoidal curves) of the generalized linear model for the case in which the resonant source (target and nontarget) was the plate. The abscissa gives the decision variable [right-hand side of Eq. (1)] corresponding to the best estimates of b_i for each listener. The open symbols are the data with error bars. The fit to the data is quite good and it is representative of that obtained for the other conditions of this study. Figures 2–4 next show the obtained estimates of b_i (symbols) for each listener separately for the membrane, bar, and plate and separately for the change in frequency and change in decay (panel columns). To facilitate comparisons across listeners (panel rows) we have adopted the common practice of normalizing the b_i so that their unsigned values sum to unity (cf. Berg, 1990). Later, in Sec. III D, we consider as a separate factor the influence of the raw regression weights prior to normalization. The two estimates of b_i agree well and, with only one exception (S6, bar), the error of estimate is quite small.[1]

Three features of these data are worth noting. First, there are clear individual differences in the decision strategy of listeners for each object class. The differences have largely to do with whether listeners' judgments are more influenced by the change in frequency or the change in decay of partials. To facilitate this comparison the panel rows corresponding to each listener's data in these figures have been roughly ordered with respect to the relative influence of frequency and decay. The differences are most evident in the case of the plate and the bar. Note, in particular, that for these two classes S1 exhibits a near exclusive reliance on the change in frequency while S6 shows a near exclusive reliance on the change in decay, both involving the lowest-frequency partials. There are also individual differences in which partials tend to dominate judgments, though here the similarities among listeners are perhaps more striking. In the case of the membrane, for example, the change in the decay of the high-frequency partials tends to dominate the judgments of S3, while the reverse appears true for S6.

The second point to note is that, while listeners differ in their approach to any one identification task, for any one listener the approach is similar across tasks. Note, for example, that S5 and S6's reliance on the change in decay is evident across all object classes, as is S1 and S3's reliance on the change in frequency, at least in the case of the bar and plate. We performed an F-ratio test to determine whether the differences across listeners were in fact greater than those across task. We only included the weights for the highest and lowest frequency partials since most of the other weights were at or near zero. The test was not significant at the $p=0.05$ level ($F_{5,2}=5.33$, $p=0.16$), however, the p value was small enough that the pooling of variance across listeners to evaluate a main effect of task would be ill-advised (cf. Bancroft and Han, 1984). There also appears in these data a tendency across object class for listeners to place greatest reliance on the changes occurring in the lowest frequency partials, though notable exceptions occur again in the case of the membrane. Taken together these observations hint at the

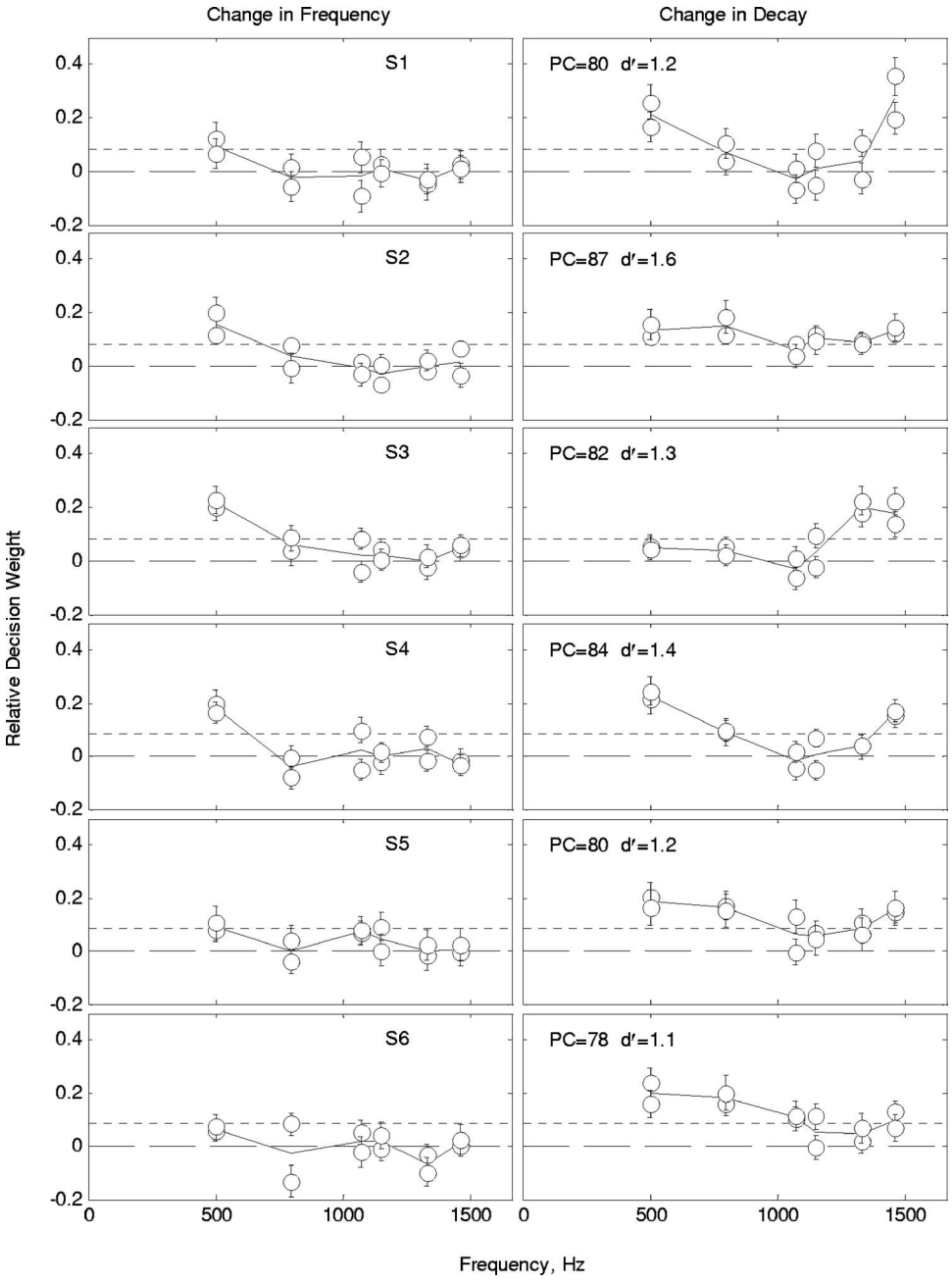


FIG. 2. Obtained regression weights b_i for the identification of size and material (Experiment 1) are given for each listener (panel rows) for the case in which the resonant source was the hinged membrane. Error bars give the standard error of estimate in each case. Short-dashed lines give the ML weights. The two panel columns give the regression weights separately for the change in frequency and change in decay of each partial. Performance levels (PC and d') are also indicated for each listener.

possibility that individual differences in decision strategy may be generally characterized by a few basic "listening styles." The idea has been considered in the literature on multitone pattern discrimination (e.g., Doherty and Lutfi, 1996) and is readdressed in Experiment 2.

The final point to note is that, despite the clear individual differences in decision strategy, performance accuracy (indicated in the upper left-hand corner of panels) is roughly similar across listeners. Excluding listener S6, who performed well below the other listeners, individual performance accuracy ranges from $d'=1.2$ to 1.9 across the different object classes. This performance is considerably less than that for the ML detector, for which $d'_{\mathrm{ML}}=2.4-2.8$ across tasks. The similar shortfall from optimal performance suggests that there may be a common upper limit on the amount of information listeners can process in these sounds. The limit might, for example, be related to inefficiencies in the decision strategy of listeners evidenced in Figs. 2–4. The optimal decision strategy in each figure (ML regression weights) is given by dashed curves. Note that in every case the regression weights for listeners deviate significantly from these curves. Whereas the ML detector gives equal weight to changes occurring in the different partials, in most cases, listener judgments appear to be dominated by the changes occurring in only two or three partials. Similar observations implying an upper limit on information capacity have been reported for the discrimination of arbitrary multitone patterns (Lutfi, 1990; Lenhart and Lutfi, 2006). We will return to this topic in Sec. III D.

B. Individual differences in the identification of mallet hardness

Next to material and size, the perception of mallet hardness has received the most attention in the literature on

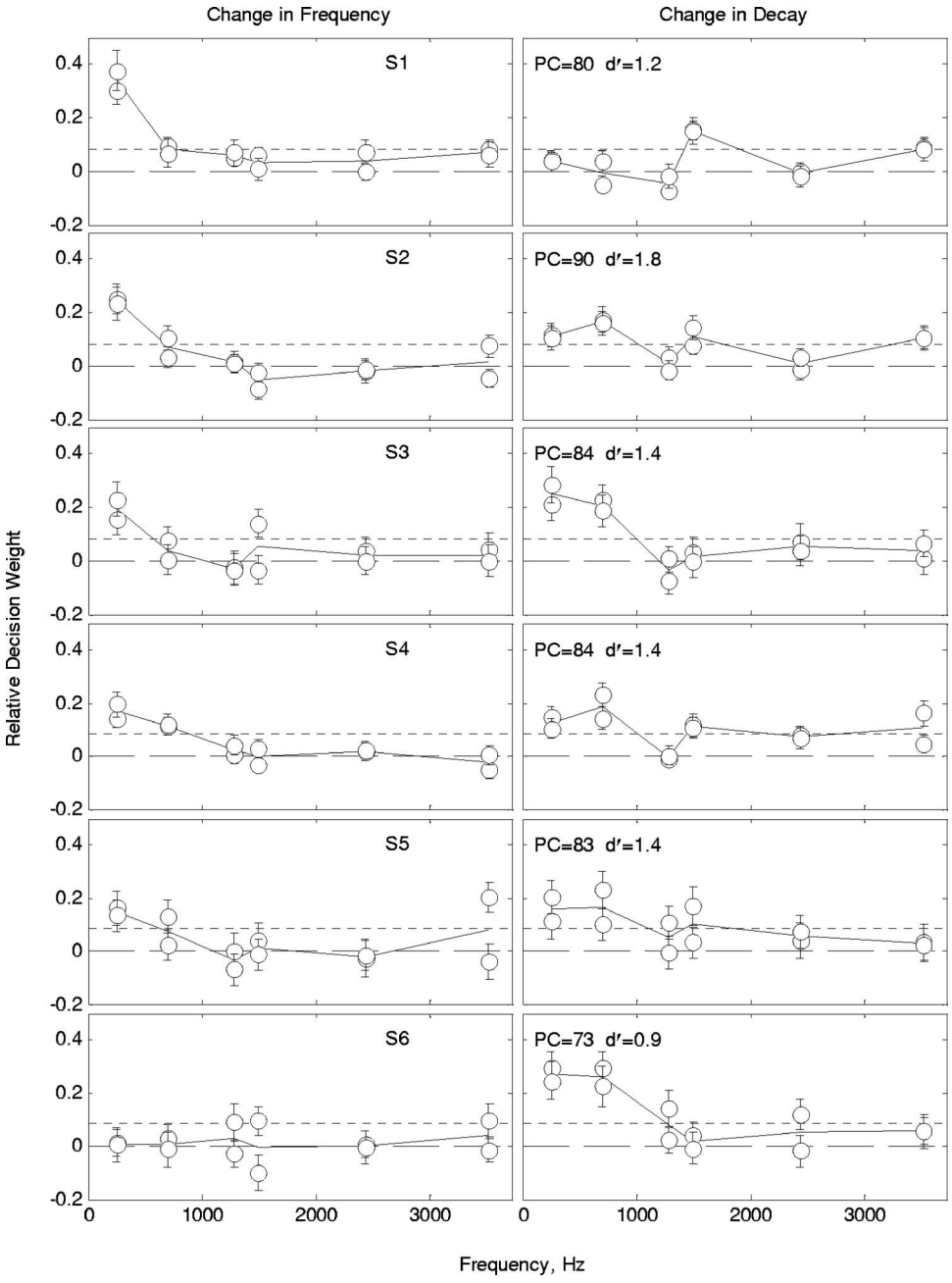


FIG. 3. The same as Fig. 2, except that the resonant source was the loosely suspended plate.

source identification from impact sounds (e.g. Freed, 1990; Giordano and Petrini, 2003). Mallet hardness was chosen as the focus of the next experiment partly for this reason, but also because it is a simple property that relates to the impact imparted to the source. The properties of the impact can be equally as important as the properties of the resonantor for source identification. Generally, softer mallets spend more time in contact with the resonator, so that vibrational modes with period shorter than the time of contact are only weakly excited. The emitted sound, therefore, has low-pass characteristic; perceptually it is less bright or more muted in quality (Benade, 1979; Fletcher and Rossing 1991, pp. 547–548). In Experiment 2 we simulated this property for the circular membrane by imposing a –2 dB/octave tilt on the spectrum of the target sound. The perturbation σ in amplitude was increased proportionally to maintain a constant Δ/σ for the individual partials (see Table I). The listener's task was to distinguish this target from a foil having, on average, a flat spectrum. All other conditions were identical to those of Experiment 1 with the exception that the frequency and decay of partials were fixed with $\nu_1=250$ Hz and $\tau_1=0.4$ s for both target and nontarget. Figure 5 shows, for four of the listeners participating in Experiment 1 and one newly recruited listener (S7), the regression weights obtained for this experiment with replication (right panels) obtained on the following day. The figures show clear idiosyncratic differences in the pattern of regression weights across listeners. Moreover, the patterns are reliably replicated on different days indicating that the differences are robust.

C. Individual differences in the identification of external damping

We next consider the question as to whether there are conditions in which listeners *do not* differ in their approach

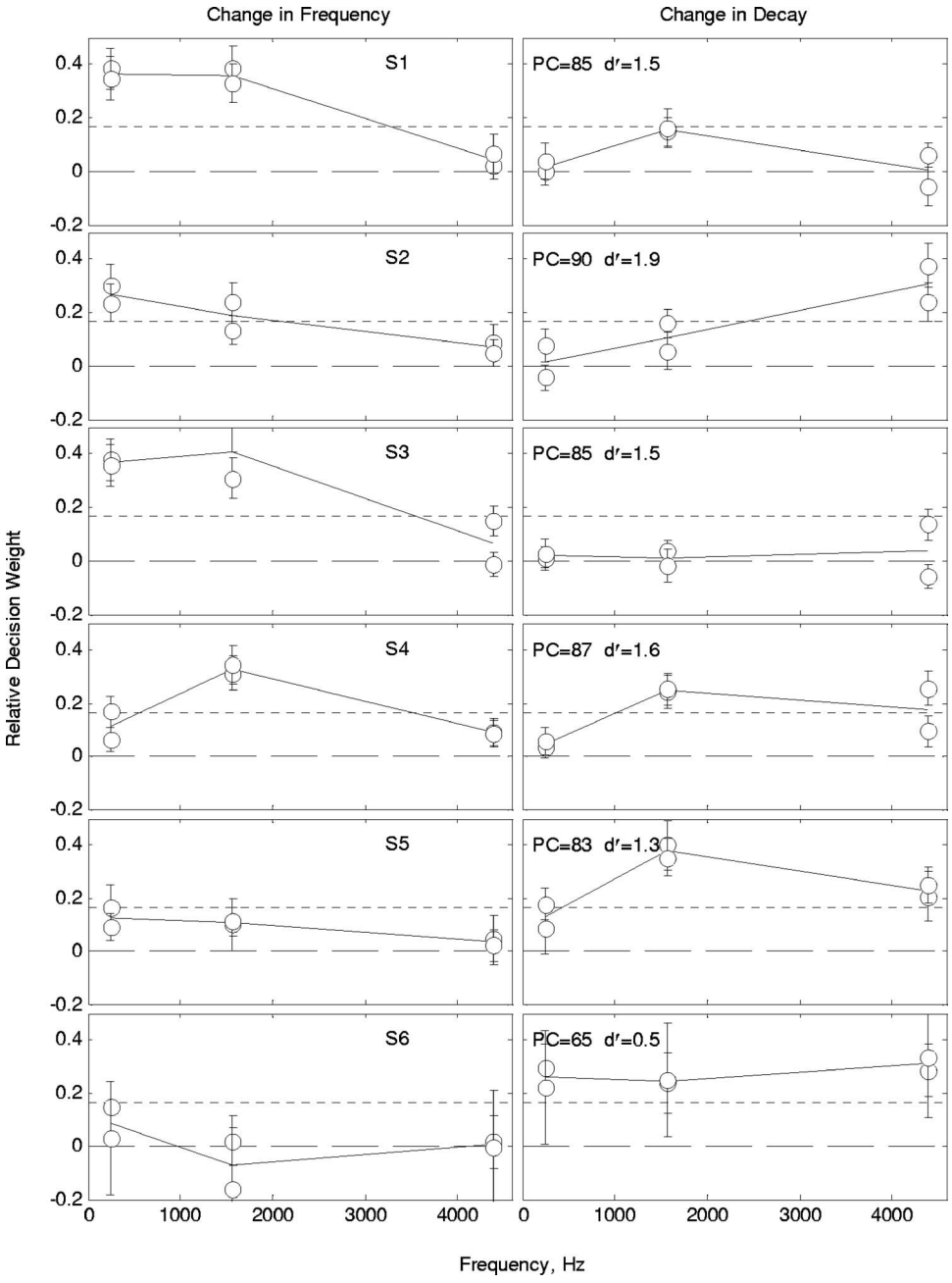


FIG. 4. The same as Fig. 2, except that resonant source was the clamped bar.

to sound source identification. The stimuli of Experiments 1 and 2 offered many opportunities to observe differences in decision strategy as the information for identification was highly redundant. The changes occurring in any one partial alone contained sufficient information to perform significantly above chance. In Experiment 3 we considered whether listeners would agree more nearly in their approach if the information for identification was instead constrained to a small subset of the partials making up the sound. The expectation is that it would, since there are fewer opportunities to differ in strategy. However, such a result is not guaranteed, particularly if any one of the subsets of partials continues to contain sufficient information to perform well above chance, as was the case in this next experiment.

The task was simulated to be that of identifying light external damping applied to the center of the resonator, in this case the circular membrane. Such damping, as might result from the light touch of a finger, has the effect of partially nulling modes 1, 4, and 9 (Hall, 1991, p.171). The task of the listener, then, was to detect a reduction in amplitude of these modes with the amplitudes of all other modes, on average, unchanged. The frequency and decay of partials were fixed and were identical for target and nontarget. Additional partials were also introduced at $k=14.91$, 20.66, and 26.99 with $\nu_1=125$ Hz. Otherwise conditions were identical to those of Experiment 1.

Figure 6 shows the regression weights on the changes in amplitude for each partial for five of the listeners who had previously participated in Experiment 1. Performance metrics are given in the upper-left-hand corner of each panel and the ML decision weights are given by the dashed curves as before. In contrast to the results of Experiments 1 and 2, Fig. 6 shows greater similarity in the decision strategy of listeners. Here each listener clearly places greatest reliance on the

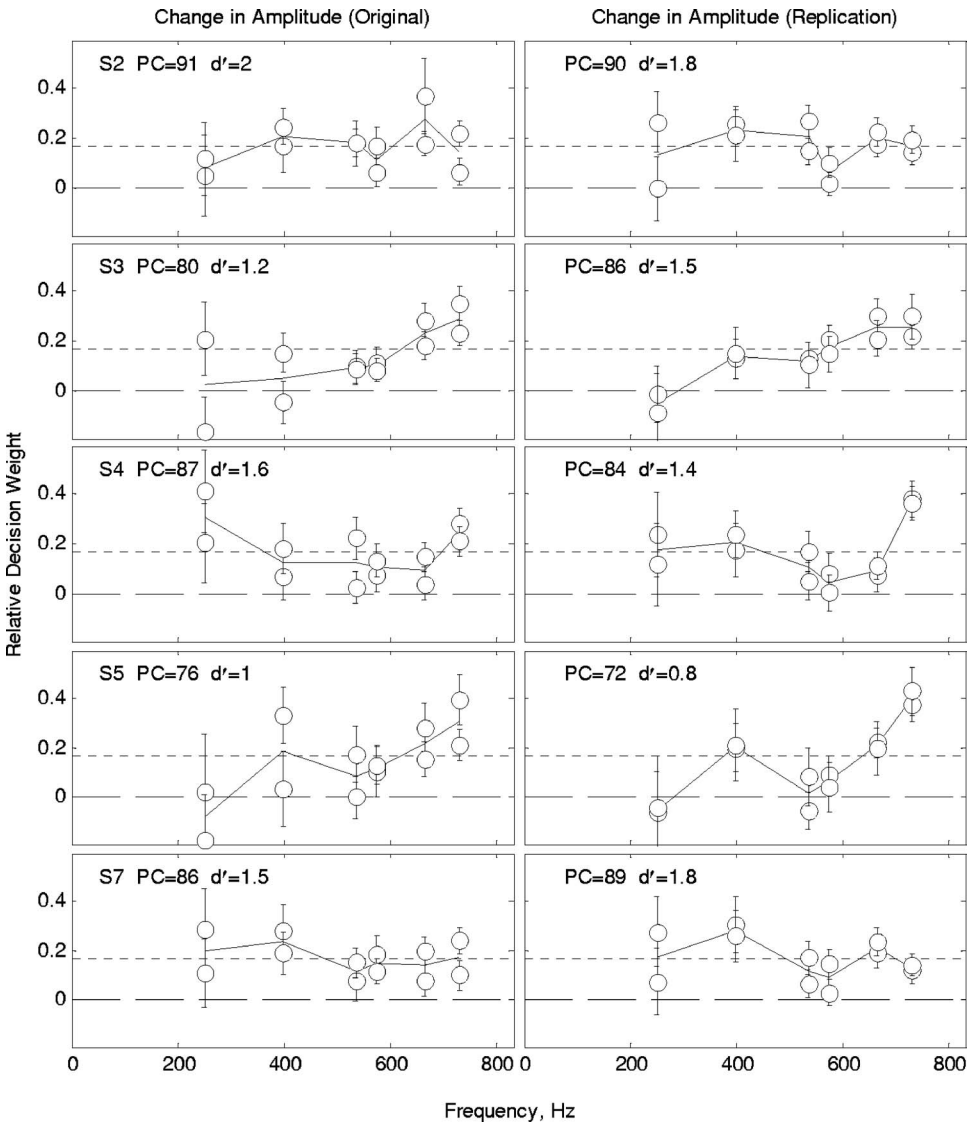


FIG. 5. Obtained regression weights b_i for identification of mallet hardness (Experiment 2) are given for each listener (panel rows) with standard error of estimate. Performance levels (PC and d') and ML weights (short-dashed lines) are given as before. Right panels show a replication of the Experiment 2 after one day.

highest frequency partial. The values of d' are also similar, ranging from 1.1 to 1.5. The results, in general, are consistent with the expectation that listeners will adopt similar decision strategies when the information for identification is constrained to a small number of partials, even when any one of those partials contains adequate information to perform significantly above chance.

D. Individual differences in weighting efficiency and internal noise

In this section we address the final goal of the study; determining the source of individual differences in identification accuracy. Broadly speaking, two factors can be expected to contribute to individual differences in performance. The first relates to how closely the listener's decision strategy approaches that of the ML observer, what we refer to as weighting efficiency (cf. Berg, 1990). This is the component of performance that can be predicted from the stimulus and is related to the normalized values of the regression coefficients. The second is internal noise, which relates to factors not directly tied to the stimulus or its variation from trial to trial. These are factors which affect the magnitude of the raw regression weights, which might include limited sensitivity, lapses in memory, or general inattentiveness.

We used a *weighting efficiency* metric following Berg (1990) to evaluate the relative contribution of the stimulus-dependent and stimulus-independent components of performance. Let d'_{wgt} represent the performance of each listener predicted exclusively from their normalized regression weights; that is, assuming no other limits imposed by internal noise (for a general discussion of the d' metric of performance see Swets, 1996). Weighting efficiency is then defined as

$$\eta_{\mathrm{wgt}} = (d'_{\mathrm{wgt}}/d'_{\mathrm{ML}})^2, \tag{2}$$

where d'_{ML} is the performance of the ML detector as given in Table I. In practice, the d'_{wgt} for each listener is obtained from the trial-by-trial responses of a hypothetical listener that adopts the following decision rule:

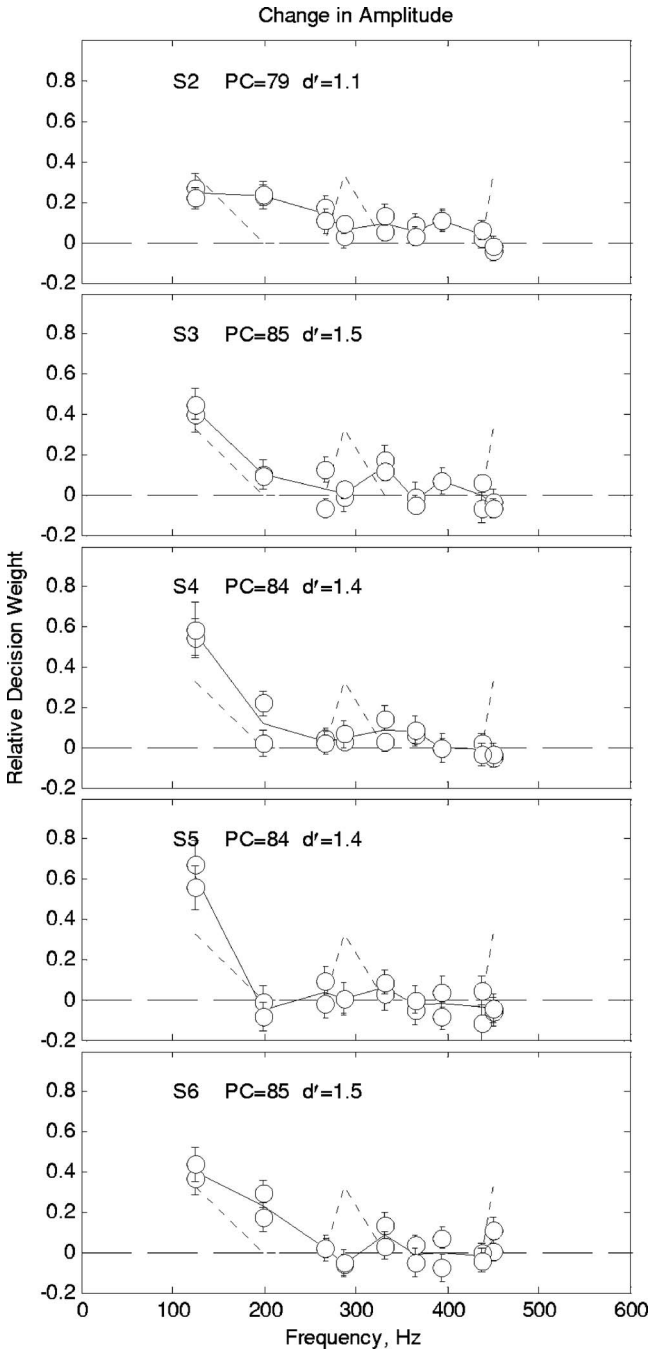


FIG. 6. The same as Fig. 5, except the data are for the identification of damping (Experiment 3).

$$\text{Respond interval 2 iff } \sum_{i=1}^{n} b_i(\log x_{i2} - \log x_{i1}) > 0 \text{ else respond interval 1}, \quad (3)$$

where the b_i are the normalized regression weights for each listener and the responses are determined for the same sequence of stimuli used in the different conditions of experiments. We next define an overall performance efficiency η_{obt} representing the combined influence of the weighting efficiency and internal noise,

$$\eta_{\text{obt}} = (d'_{\text{obt}}/d'_{\text{ML}})^2 = \eta_{\text{wgt}} \times \eta_{\text{nos}}, \quad (4)$$

where d'_{obt} is obtained performance and $\eta_{\text{nos}}=(d'_{\text{obt}}/d'_{\text{wgt}})^2$ represents the influence of internal noise. Note that η_{nos} is simply that component of performance not accounted for by the weights. We can now address the question as to the source of the individual differences by determining the relative contribution of η_{wgt} and η_{nos} to η_{obt}. Figures 7–11 show the values of η_{obt}, η_{wgt}, and η_{nos} obtained from the two independent estimates of the regression weights for each listener in each experiment of the study. The dashed curves give the prediction assuming that internal noise is responsible for all individual variation in η_{obt} within each task, i.e., η_{wgt} is constant within task. The constant value in each case is given by the mean of the computed η_{wgt} values across listeners. The mean values of η_{wgt} are neither consistently greater nor consistently less than 0.5, indicating that the dominant factor influencing performance (decision weights or internal noise) depends on the task. More important, what the figures do consistently show is greater variability in the individual values of η_{nos} compared to η_{wgt}. An F-ratio test revealed the differences to be significant at the $p=0.05$ level in every case except Fig. 2, which fell just slightly short of significance ($p=0.06$). For Figs. 7–11 the F-ratio values were, respectively, $F_{11,11}=5.49$, $F_{11,11}=2.58$, $F_{13,13}=3.30$, $F_{25,25}=2.19$, and $F_{9,9}=4.91$. Taking into account experimentwise error rate, at best only one of the F values would be expected to be significant by chance at the $p=0.05$ level. By this analysis, then, it appears that most of the individual variation in identification performance is due to factors unrelated to the stimulus; that differences in the decision strategies of listeners play only a secondary role.

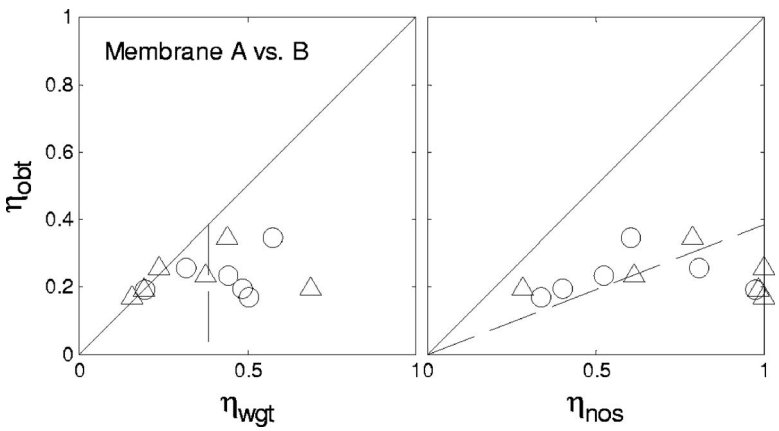


FIG. 7. For all listeners, the relative influence of internal noise η_{nois} and weighting efficiency η_{wgt} on overall performance efficiency η_{obt} for the identification of membrane size and material (Experiment 1). See the text for details.

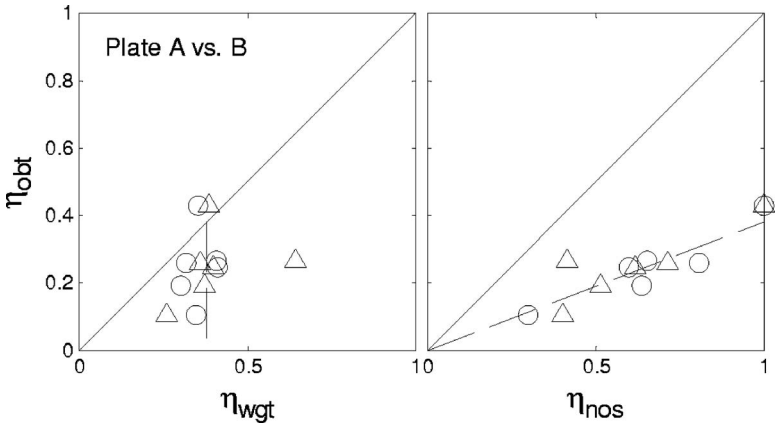


FIG. 8. The same as Fig. 7, except the data are for the identification of plate size and material (Experiment 1).

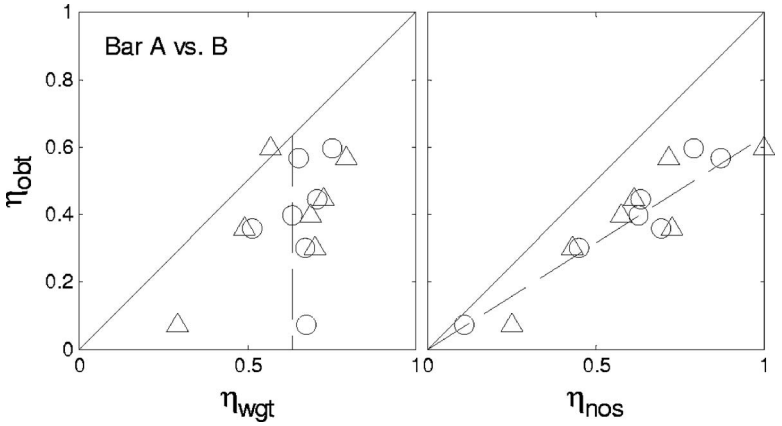


FIG. 9. The same as Fig. 7, except the data are for the identification of bar size and material (Experiment 1).

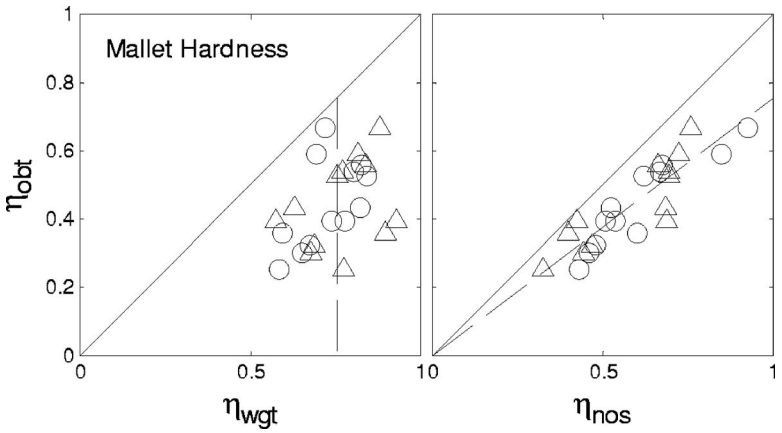


FIG. 10. The same as Fig. 7, except the data are for the identification of mallet hardness (Experiment 2).

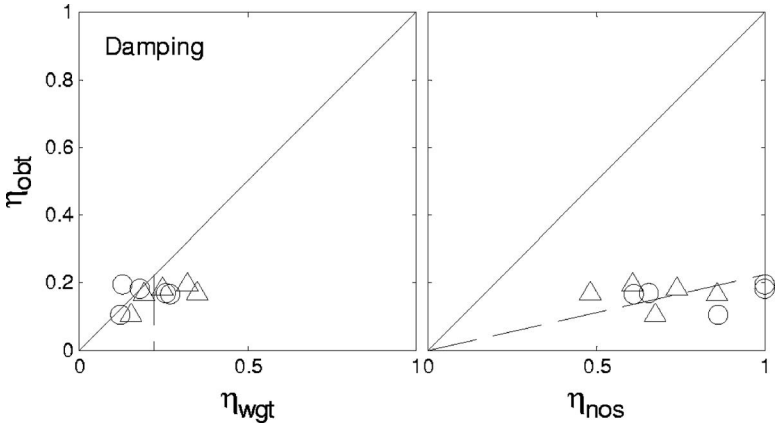


FIG. 11. The same as Fig. 7, except the data are for the identification of damping (Experiment 3).

IV. DISCUSSION

The results of the present study are consistent with past studies showing that listeners can differ substantially in their approach to the identification of simple resonant sources from impact sounds. The study reports several new findings regarding these differences: First, the differences in decision strategy that are observed within a given identification task do not always lead to observable differences in identification accuracy. Experiments 1 and 2 showed that when identifying differences in size, material, and manner of impact, listeners largely based their judgments on the changes occurring in a small number of partials. It mattered little to overall performance which changes the listener chose to emphasize, as all changes in these experiments conveyed equally reliable information for identification. Second, listeners appear to have individualized listening styles. Experiments 1 and 2 further showed that the particular changes listeners chose to emphasize in a given task carry over across variations of that task and replications on different days. Third, individual differences in decision strategy are less likely to be observed when the information for identification is restricted to a subset of the partials making up the sound. In Experiment 3 the information for the identification of damping was simulated as a change in the amplitude of three of nine partials. Though identification accuracy varied across listeners, regression weights were similar for all listeners and close to those of the ML detector. Finally, despite the considerable individual variation in decision strategy across listeners within a task, most of the variation in identification accuracy appears due to factors other than decision strategy. This was the conclusion indicated by an efficiency analysis of the data from all three experiments.

The results suggest a need to reexamine reported findings of past studies where conclusions have been based on group-averaged data. Such practice has been widespread in the literature, both as applies to measures of performance accuracy and estimates of the strength of relation between stimulus parameters and the listeners' response. Moreover, while studies have typically exercised care in reporting measures of dispersion associated with group averages, such measures do not distinguish whether the observed differences are real or are simply the result of measurement error, the latter being assumed in most cases. This can have one of two negative consequences: It can cause real relations between stimulus parameters and the listener's response to be missed, or worse, it can lead to false conclusions about the existence of such relations. An example of the former case was given by Gaver (1988), as described in Sec. I. An example of the latter case is provided by the data of Fig. 5, where averaging across listeners might lead one to conclude that listener decision strategy was much closer to that of ML detector than, in fact, it was for any one listener.

The present results also suggest care in presuming too strong of a relation between decision strategy and performance accuracy when evaluating decision strategy. This is the risk, for example, in studies where accurate identification is taken to imply the perception of invariant acoustic relations associated with a particular object class (e.g., Carello *et al.*, 1998; Kunkler-Peck and Turvey, 2000), or where a listener's reliance on a particular acoustic cue is inferred from the effect on performance of eliminating that cue (e.g., Warren and Verbrugge, 1984). We have seen in the present study cases in which different decision strategies yield similar levels of performance (Experiments 1 and 2), and cases where different levels of performance are obtained for equally efficient decision strategies (as indicated by the efficiency analysis of the data from these experiments). Indeed, the problem of inferring decision strategy from performance has been recognized for some time and has provided incentive to exploit alternative methods of evaluating decision strategy that place less significance on performance accuracy (e.g., Giordano and McAdams, 2006; Lutfi, 2000, 2001; Lutfi and Oh, 1997; McAdams *et al.*, 2004).

Finally, we should note certain similarities of our results to those obtained in studies involving the discrimination of unfamiliar, multitone patterns. Lutfi (1990), for example, conducted an experiment, somewhat comparable to our Experiment 1, in which highly trained listeners distinguished target from nontarget tone sequences based on an average difference in the frequency, amplitude, and duration of tones. As in Experiment 1, random perturbations in the tone parameters were added on each presentation. The finding was that performance grew at a less than optimal rate with the number of tones in the sequence, corresponding to an information limit per stimulus of about 2 bits. Lenhart and Lutfi (2006) later showed the limit to result from the tendency of listeners to rely predominantly on two to three tones in the sequence, a result consistent with that of Experiment 1. Doherty and Lutfi (1996) provide data comparable to those of Experiment 2. The task was to detect a level increment in a multitone complex with level-perturbation added independently to each tone. Individual regression weights on the change in level for each tone were obtained from 11 highly trained listeners. Like our Experiment 2, performance levels were similar across listeners, while regression weights revealed clear differences in the reliance listeners placed on the different tones in the complex. Remarkably, the idiosyncratic pattern of regression weights obtained for each listener was replicated after a period of a week, raising the specter that individual listening styles might be used to predict broad patterns of behavior across different studies. The agreement among these studies suggests that much of what has been learned from multitone-pattern discrimination studies might apply as well to the identification of rudimentary sources from impact sounds. Indeed, those who have invested much effort in studying the discrimination of multitone patterns have done so with the idea that the knowledge gained would generalize to other listening tasks. If the results of the present study do not greatly advance this view, they at least suggest that lessons learned regarding individual differences similarly apply.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Andrew Oxenham and two anonymous reviewers for helpful comments on an earlier version of the manuscript. This research was supported by NIDCD Grant No. R01 DC006875-01.

[1]For this reason, we have decided not to undertake extensive ad hoc statistical tests of significance that are likely, in any case, to show the observed differences across listeners to be real.

Anderson, T. W. (**1971**). *An Introduction to Multivariate Statistical Analysis* (Wiley, New York), pp. 205–217.

ANSI S3.6-1996 (**1996**). *American National Standards Specification for Audiometers* (American National Standards Institute, New York).

Bancroft, T. A. and Han, C. P. (**1984**). "A note on pooling variances," J. Am. Stat. Assoc. **78**, 981–983.

Benade, A. H. (**1979**). *Fundamentals of Musical Acoustics* (Oxford University Press, London).

Berg, B. G. (**1990**). "Observer efficiency and weights in a multiple observation task," J. Acoust. Soc. Am. **88**, 149–158.

Cabe, P. A., and Pittenger, J. B. (**2000**). "Human sensitivity to acoustic information from vessel filing," J. Exp. Psychol. Hum. Percept. Perform. **26**, 313–324.

Carello, C., Anderson, K. A., and Kunkler-Peck, A. J. (**1998**). "Perception of object length by sound," Psychol. Sci. **9**, 211–214.

Doherty, K. A., and Lutfi, R. A. (**1996**). "Spectral weights for overall level discrimination in listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **99**, 1053–1058.

Fletcher, N. H., and Rossing, T. D. (**1991**). *The Physics of Musical Instruments* (Springer, New York).

Freed, D. J. (**1990**). "Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events," J. Acoust. Soc. Am. **87**, 311–322.

Gaver, W. W. (1988). "Everyday listening and auditory icons," Ph.D. dissertation, University of California, San Diego, CA.

Giordano, B. L., and McAdams, S. (**2006**). "Material identification of real impact sounds: Effects of size variation in steel, glass, wood and plexiglass plates," J. Acoust. Soc. Am. **119**, 1171–1181.

Giordano, B. L., and Petrini, K. (**2003**). "Hardness recognition in synthetic sounds." *Proceedings of the Stockholm Music Acoustics Conference*, Stockholm, Sweden.

Hall, D. E. (**1991**). *Musical Acoustics*. (Cole, Pacific Grove, CA).

Helmholtz, H. (**1877**). *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, 4th ed., translated by A. J. Ellis (Dover, New York, **1954**).

Houben, M., Kohlrausch, A., and Hermes, D. J. (**2005**). "The contribution of spectral and temporal information to the auditory perception of the size and speed of rolling balls," Acta Acust. **6**, 1007–1015.

Jesteadt, W., Wier, C. C., and Green, D. M. (**1977**). "Intensity discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 169–177.

Kunkler-Peck, A. J., and Turvey, M. T. (**2000**). "Hearing shape," J. Exp. Psychol. **26**, 279–294.

Lakatos, S., McAdams, S., and Causse, R. (**1997**). "The representation of auditory source characteristics: Simple geometric form," Percept. Psychophys. **59**, 1180–1190.

Lenhart, E., and Lutfi, R. A. (**2006**). "Effect of decision weights and internal noise on the growth of d' with N," J. Acoust. Soc. Am. **119**, 3236–3237.

Li, X., Logan, R. J., and Pastore, R. E. (**1991**). "Perception of acoustic source characteristics: Walking sounds," J. Acoust. Soc. Am. **90**, 3036–3049.

Lutfi, R. A. (**1990**). "Informational processing of complex sound. II. Cross-dimensional analysis," J. Acoust. Soc. Am. **87**, 2141–2148.

Lutfi, R. A. (**1993**). "A model of auditory pattern analysis based on component-relative-entropy," J. Acoust. Soc. Am. **94**, 748–758.

Lutfi, R. A. (**2000**). "Source uncertainty, decision weights, and internal noise as factors in auditory identification of a simple resonant source," J. Assoc. Res. Otolaryngol. **23**, 171.

Lutfi, R. A. (**2001**). "Auditory detection of hollowness," J. Acoust. Soc. Am. **110**, 1010–1019.

Lutfi, R. A. (**2007**). "Human Sound Source Identification," in *Sound Source Perception*, edited by W. A. Yost (Springer, New York).

Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (**2003**). "One factor underlies individual differences in auditory informational masking within and across age groups," Percept. Psychophys. **65**, 396–406.

Lutfi, R. A., and Oh, E. (**1997**). "Auditory discrimination of material changes in a struck-clamped bar," J. Acoust. Soc. Am. **102**, 3647–3656.

McAdams, S., Chaigne, A., and Roussarie, V. (**2004**). "The psychomechanics of simulated sound sources: Material properties of impacted bars," J. Acoust. Soc. Am. **115**, 1306–1320.

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics*. (Princeton University Press, Princeton, NJ), pp. 175–191.

Neff, D. L., and Dethlefs, T. M. (**1995**). "Individual differences in simultaneous masking with random-frequency, multicomponent maskers," J. Acoust. Soc. Am. **98**, 125–134.

Repp, B. H. (**1987**). "The sound of two hands clapping: An exploratory study," J. Acoust. Soc. Am. **81**, 1100–1109.

Schlauch, R. S., Ries, D. T., and DiGiovanni, J. J. (**2001**). "Duration discrimination and subjective duration for ramped and damped sounds," J. Acoust. Soc. Am. **109**, 2880–2887.

Swets, J. A. (**1996**). *Signal Detection Theory and ROC Analysis in Psychology and Diagnosis*. (Erlbaum, Mahwah, NJ).

Tucker, S., and Brown, G. J. (**2003**). "Modelling the auditory perception of size, shape and material: Applications to the classification of transient sonar sounds," presented at the 114th Audio Engineering Society Convention, Amsterdam, The Netherlands.

Warren, W. H., and Verbrugge, R. R. (**1984**). "Auditory perception of breaking and bouncing events: A case study in ecological acoustics," J. Exp. Psychol. **10**, 704–712.

Watson, C. S., and Kelly, W. J. (**1981**). "The role of stimulus uncertainty in the discrimination of auditory patterns," in *Auditory and Visual Pattern Recognition*, edited by D. J. Getty and J. H. Howard, Jr. (Erlbaum, Mahwah, NJ), pp. 37–59.

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.

Interaural fluctuations and the detection of interaural incoherence. III. Narrowband experiments and binaural models

Matthew J. Goupell[a] and William M. Hartmann
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824

(Received 17 November 2005; revised 26 March 2007; accepted 30 March 2007)

In the first two articles of this series, reproducible noises with a fixed value of interaural coherence (0.992) were used to study the human ability to detect interaural incoherence. It was found that incoherence detection is strongly correlated with fluctuations in interaural differences, especially for narrow noise bandwidths, but it remained unclear what function of the fluctuations best agrees with detection data. In the present article, ten different binaural models were tested against detection data for 14- and 108-Hz bandwidths. These models included different types of binaural processing: independent-interaural-phase-difference/interaural-level-difference, lateral-position, and short-term cross-correlation. Several preprocessing transformations of the interaural differences were incorporated: compression of binaural cues, temporal averaging, and envelope weighting. For the 14-Hz bandwidth data, the most successful model postulated that incoherence is detected via fluctuations of interaural phase and interaural level processed by independent centers. That model correlated with detectability at $r=0.87$. That model proved to be more successful than short-term cross-correlation models incorporating standard physiologically-based model features ($r=0.78$). For the 108-Hz bandwidth data, detection performance varied much less among different waveforms, and the data were less able to distinguish between models. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2734489]

PACS number(s): 43.66.Ba, 43.66.Pn, 43.66.Qp [AK] Pages: 1029–1045

I. INTRODUCTION

In a previous article, to be called "Article I," Goupell and Hartmann (2006) studied the ability of listeners to detect a small amount of interaural incoherence. The experiments employed selected noises in which the interaural coherence was fixed at a value of 0.992, where the interaural coherence was defined as the maximum of the cross-correlation function, as computed over the entire duration (500 ms) of the stimulus. Physical analysis of the noises studied the fluctuations in interaural phase difference (IPD) and interaural level difference (ILD). It was found that these fluctuations were increasingly variable across different noises for decreasing bandwidth.

In the psychoacoustical experiments of Article I, the listener's task was to distinguish between the incoherent noises (coherence=0.992) and diotic noises with a coherence of 1.0. In spite of the fact that the incoherent noises all had the same coherence, the experiments showed that for narrow bandwidths the incoherence was much more readily detectable in some noises than in others. Listeners found it significantly easier to detect incoherence when the fluctuations in IPD or ILD were larger. As the bandwidth increased, the incoherence became equally detectable in all the different noises, consistent with a model in which detection is predictable from interaural coherence alone.

The stimuli for the experiments of Article I were selected based on large or small fluctuations in interaural phase or level. For any given noise, the fluctuations were measured by the standard deviations over time in IPD or ILD. The corresponding variations in detectability, especially when the bandwidth was as narrow as 14 Hz, indicated that these fluctuation measures have considerable perceptual validity. However, it is possible, even likely, that some other measure of stimulus fluctuation would correlate better with human perception of incoherence.

It is also not clear that IPD and ILD fluctuations, as used in Article I, should be considered as comparably important. The experiments showed that variability in IPD and variability in ILD led to similar variability in detectabilty, but phase and level fluctuations are so strongly correlated within an ensemble of noises that comparable data do not clearly demonstrate comparable importance. For instance, it is possible that listeners only responded to interaural phase fluctuations. Experiments using controlled level fluctuations would then lead to significant effects only because the level fluctuations are so strongly correlated with phase fluctuations.

The purpose of the present article is to address the uncertainties left by Article I by testing a variety of different binaural detection models against incoherence detection data. The binaural detection models were derived from models previously used to study the masking-level difference (MLD). This was a sensible approach because the MLD is closely related to incoherence detection (Durlach *et al.*, 1986; Bernstein and Trahiotis, 1992). Although the set of models tested affords a notable variety, it must be acknowledged in advance that the set is not exhaustive.

[a]Electronic mail: goupell@kfs.oeaw.ac.at

Domnitz and Colburn (1976) summarized the two major types of binaural models historically presented to explain the MLD phenomenon. The first type uses interaural parameter differences, IPD and ILD. For example, the vector model (Jeffress *et al.*, 1956) predicts the largest release from masking for a signal phase difference of 180° (NoSπ). Another example is the lateralization model (Hafter, 1971), in which a signal is detected by a shifted lateral image that is formed by combining IPD and ILD. The second type of binaural model includes energy and cross-correlation models. When an out-of-phase tone is added to homophasic noise, the interaural correlation of the entire stimulus is reduced. Models such as the equalization-cancellation (EC) model by Durlach (1963) and the correlation model of Osman (1971) fall into this category.

There is still debate as to which type of model best describes binaural detection. Gilkey *et al.* (1985) found that wideband reproducible-noise masking data were incompatible with several interaural parameter models. On the other hand, Colburn *et al.* (1997) showed that the EC model was incompatible with reproducible noise data from Isabelle and Colburn (1991). Several recent articles have favored EC-like models to describe binaural detection data (Breebaart *et al.*, 1999; Breebaart and Kohlrausch, 2001; Breebaart *et al.*, 2001a, b, c). However, not all of the data can be described by EC-like models. Breebaart and Kohlrausch (2001) found that correlation and energy models cannot entirely describe thresholds for stimuli with a fixed-interaural correlation. Using narrowband multiplied noise, Breebaart *et al.* (1999) found that neither interaural difference parameters nor the EC model could account for the results of experiments that included static level differences.

Article I showed that incoherence detection cannot be understood in terms of coherence derived from the cross-correlation of the stimulus as computed over a long duration. A second article, Article II (Goupell and Hartmann, 2007), found the same negative result for coherence computed over short-duration stimuli. Therefore, the present article focuses on interaural parameter models and on cross-correlation models that include a physiologically-based preprocessor.

II. EXPERIMENT 1: 14-Hz BANDWIDTH

The purpose of Experiment 1 was to obtain incoherence detection data from a large set of narrowband noises that were randomly generated and unselected so as to be a fair representation of all noises with a given bandwidth, duration, and interaural coherence. The detection data were collected in order to test the models presented in this article.

A. Stimuli

A collection of 100 dual-channel noises with 14-Hz bandwidth was created for Experiment 1. It was the same collection from which particular noises were selected in Article I. In the present experiment all 100 noises were used to avoid any bias.

Each noise was constructed from equal-amplitude random-phase components that spanned a frequency range of 490–510 Hz with a frequency spacing of 2 Hz. Components between 495 and 505 Hz had equal amplitudes of unity. Frequencies below 495 and above 505 Hz were attenuated with a raised-cosine spectral window. The 3-dB bandwidth was 14 Hz. An orthogonalization procedure guaranteed that the interaural coherence of each noise was precisely 0.992.

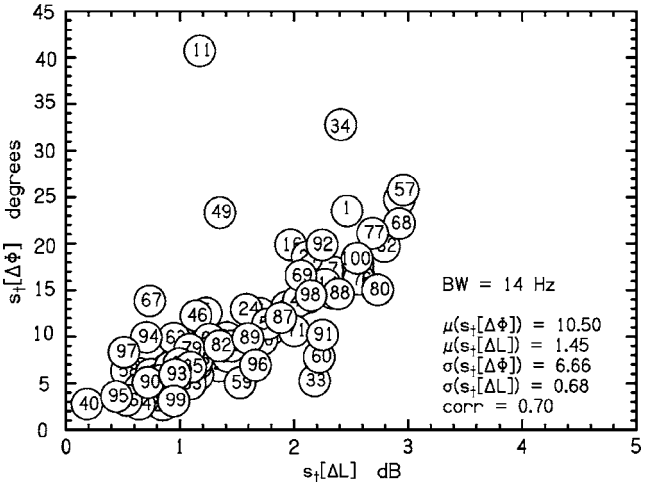


FIG. 1. Fluctuations of IPD vs fluctuations of ILD for the collection of 100 reproducible noises having a 14-Hz bandwidth used in Experiment 1. Each noise is labeled by a serial number indicating only the order of creation. The means, standard deviations, and IPD-ILD correlation of the distributions are reported.

As in Articles I and II, our attention focused on the interaural phase difference,

$$\Delta\Phi(t) = \phi_R(t) - \phi_L(t), \tag{1}$$

and the interaural level difference,

$$\Delta L(t) = 20 \log_{10}\left[\frac{E_R(t)}{E_L(t)}\right], \tag{2}$$

where ϕ is the phase and E is the envelope calculated from the analytic signal. Fluctuations in these interaural differences were initially defined in terms of their standard deviations over time, computed over the duration of the stimulus T, and indicated by the functions

$$s_t[\Delta\Phi] = \sqrt{\frac{1}{T}\int_0^T [\Delta\Phi(t) - \overline{\Delta\Phi}]^2 dt} \tag{3}$$

and

$$s_t[\Delta L] = \sqrt{\frac{1}{T}\int_0^T [\Delta L(t) - \overline{\Delta L}]^2 dt}. \tag{4}$$

The fluctuations $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ were calculated for each noise, and average quantities, indicated by an overbar, refer to a time-averaged interaural difference for the noise—normally very close to zero. These fluctuations are shown in Fig. 1 for the 100 noises, labeled by serial number (order of creation). Figure 1 also indicates the mean, standard deviation, and correlation of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$, computed over the ensemble of 100 noises.

As in Article I, noise stimuli were presented to the listeners in three observation intervals, each with a total duration of 500 ms and with 30-ms Hanning windows for attack

and decay. Noises were computed by a Tucker-Davis AP2 array processor (System II) and converted to analog form by 16-bit DACs (DD1). The buffer size was 4000 samples per channel and the sample rate was 8000 samples per second (sps). The noise was low-pass filtered with a corner frequency of 4 kHz and a −115-dB/octave rolloff. The noises were presented at 70±3 dB SPL with levels determined by programmable attenuators (PA4) operating in parallel on the two channels prior to the low-pass filtering. The level was randomly chosen in 1-dB steps for each of the three intervals within a trial to discourage the listener from trying to use level cues to perform the task.

B. Procedure

Listeners were seated in a double-wall sound-attenuating room and used Sennheiser HD414 headphones. The 100 noises were presented in sets of ten as ordered by serial number. Thus, the first set had noises 1–10, the second set had noises 11–20, and so on. Six runs were devoted to listening to a set of ten reproducible noises. Listeners completed each set before moving on to the next set.

The structure of runs, trials within a run, and the data collection procedure were the same as in Articles I and II. It is briefly described as follows: A noise could be presented either incoherently (the dichotic presentation of x_L and x_R) or it could be presented coherently (the diotic presentation of x_L). A run consisted of 60 trials, where each of the ten reproducible noises in a set was presented incoherently a total of six times. Thus, a listener heard an individual noise incoherently a total of 36 times (six runs times six presentations per run).

On each trial the listener heard a three-interval sequence. The first interval was the standard interval, which was always a coherent noise. The second interval was randomly chosen to be either incoherent or coherent. The third interval was the opposite of the second (e.g., if the second interval was coherent, the third interval was incoherent). The two coherent presentations were randomly selected from the remaining nine reproducible noises in the set except that they were required to be different from the x_L and x_R in the incoherent "odd" interval and to be different from one another. The interinterval duration was 150 ms. The listener was required to decide which of the two latter intervals was the incoherent interval. As described in Articles I and II, listeners were allowed to indicate that they were confident about a response, leading to a confidence adjusted score (CAS) on a scale of 0 to 72 for a run of 36 trials. Although the data collection procedure also kept track of the percentage of correct responses (P_c), it was found that the CAS provided greater "dynamic range" by preventing most of the ceiling effects for the most successful listeners. Confidence ratings and multipoint decision scales have been shown by signal detection theory to be valid psychophysical techniques (Egan *et al.*, 1959; Schulman and Mitchell, 1966).

C. Listeners

Experiments in this article employed three male listeners from Article I—D, M, and W. Listeners D and M were between the ages of 20 and 30 and had normal hearing according to standard audiometric tests and histories. Listener W was 65 and had a mild bilateral hearing loss, but only at frequencies four octaves above those used in the experiment. Listeners M and W were the authors.

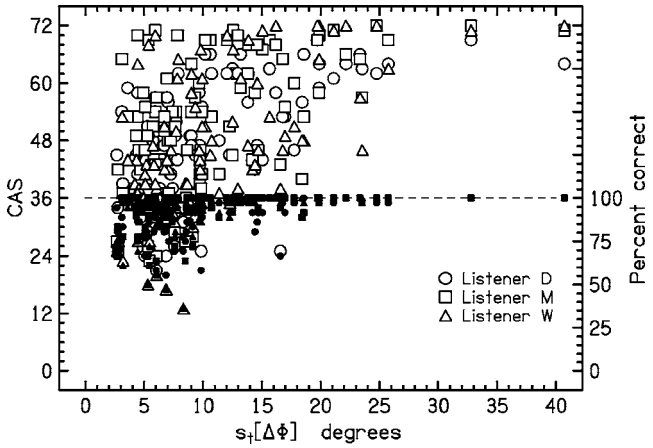


FIG. 2. All the detection data for the 100 noises from Experiment 1 for three listeners, D, M, and W, are plotted twice, once as the number correct—on a scale from 0 to 36 (0 to 100%), and once as CAS—on a scale from 0 to 72. The data are plotted as a function of the standard deviation of the interaural phase in an attempt to give some order to the plot.

D. Results

The results from experiments using all 100 noises can be seen in Fig. 2. The open symbols show the CAS while the closed symbols show the number of correct responses, essentially equivalent to the P_c. These values are plotted as a function of $s_t[\Delta\Phi]$ only to give some order to the plot, not because $s_t[\Delta\Phi]$ is thought to be the best model for detection. Figure 2 illustrates the advantage of using the CAS over P_c because the number of correct responses reaches a ceiling, especially for listener M. The CAS increases the dynamic range of the experiment, though it has not completely removed ceiling effects.

Agreement between the listeners for individual noises is difficult to see in Fig. 2, but agreement is actually good. The interlistener Pearson correlation was 0.73 for D and M, 0.71 for D and W, and 0.80 for M and W. These interlistener correlations are smaller than those reported for the ten noises in Article I—approximately 0.9 on average. The reason for the difference is probably that members of the entire collection of 100 noises are less distinctive than are the five largest and five smallest fluctuation noises used in Article I.

III. MODELS FOR INCOHERENCE DETECTION

In order to discover the stimulus features that best predict human perception of incoherence, models of perception were constructed using transformed interaural parameters, as described below, and the models were tested against the large set of perceptual data from Experiment 1.

A. Model preprocessing assumptions

Several assumptions, common to all models, were made to reflect auditory preprocessing of the complex incoherent

stimuli. Two free parameters are introduced in the following as well as a scale of lateralization for the $\Delta\Phi(t)$ and $\Delta L(t)$.

1. Temporal averaging

The fluctuation measures used to construct stimulus sets for Article I were based on instantaneous values of interaural differences as they appeared with our 8 ksps sample rate. But it is not evident that, for example, a large interaural difference with a duration of only 0.125 ms would receive much respect from the binaural system. Therefore, the present models include a parametric temporal averaging operation, following other models, e.g., Viemeister (1979), in using an exponential averaging window to represent temporal modulation transfer functions of the form

$$\Delta\Phi'(t) = \hat{e}[\Delta\Phi(t)] = \frac{\int_0^{T_D} \Delta\Phi(t-t')e^{-t'/\tau}dt'}{\int_0^{T_D} e^{-t''/\tau}dt''} \quad (0 < t < T), \tag{5}$$

and

$$\Delta L'(t) = \hat{e}[\Delta L(t)] = \frac{\int_0^{T_D} \Delta L(t-t')e^{-t'/\tau}dt'}{\int_0^{T_D} e^{-t''/\tau}dt''} \quad (0 < t < T). \tag{6}$$

Parameter T is the duration of the stimulus, and the time constant τ was a free parameter. The averaging window, with running variable t', was terminated when t' became greater than t or when the weight of the exponential function dropped to 0.1, which determined the upper limit of the integration T_D.

2. Compression of binaural cues

A small static interaural difference leads to a small displacement in the lateral position of the auditory image from a centered position. A greater interaural difference leads to a greater displacement, but increasing interaural differences produce diminishing returns because the laterality is a compressive function of interaural differences. A perceptual model for fluctuations can easily adopt this effect from static experiments. The compression functions, to be called "laterality compression," used in the present analysis were exponential fits to the data from Yost's 1981 experiments. They are of the form

$$\Psi'_{\Delta\Phi}(t) = 10\ \mathrm{sgn}[\Delta\Phi'(t)](1 - e^{-|\Delta\Phi'(t)|/40}), \tag{7}$$

and

$$\Psi'_{\Delta L}(t) = 10\ \mathrm{sgn}[\Delta L'(t)](1 - e^{-|\Delta L'(t)|/8}), \tag{8}$$

where $\Psi'_{\Delta\Phi}(t)$ and $\Psi'_{\Delta L}(t)$ are on a scale of lateral position that ranges from −10 to 10. In Eq. (7), the weighting constant of the exponential is 40°. In Eq. (8), the weighting constant of the exponential is 8 dB. These functions correspond to the experimentally determined lateral position of a sine tone at a frequency of 500 Hz, the center frequency of our noise bands. A further benefit of the compressive laterality transformation is that IPD and ILD are put on the same scale so that they can be easily combined in mathematical models.

3. Critical envelope value weighting

Maxima can occur in the IPD, $\Delta\Phi(t)$, at times when the envelope in one ear is very small. But if the envelope is near zero, the listener may not be able to detect this fluctuation in $\Delta\Phi(t)$. Therefore, it would be wrong for a model to give much weight to this phase fluctuation. We sought to reduce the problem by discounting phase fluctuations that coincided with very small envelope values[1] by employing a weighting function,

$$w_g(t) = \begin{cases} 1 & \text{if } E_L(t) \text{ and } E_R(t) \geqslant gE_{\mathrm{rms}} \\ 0 & \text{if } E_L(t) \text{ or } E_R(t) < gE_{\mathrm{rms}}, \end{cases} \tag{9}$$

where E_L and E_R are Hilbert envelopes for left and right channels, and E_{rms} actually indicates a comparison with corresponding left- or right-channel overall rms values. Parameter g is the critical envelope fraction, a free parameter. If the envelope in either channel is less than g times the rms envelope, then the weight is set to zero. Otherwise, the weight is set to one.

After all the modeling assumptions, the *transformed* IPD and ILD are described by the notation

$$\Psi_{\Delta\Phi}(t) = \Psi'_{\Delta\Phi}(t)w_g(t) \tag{10}$$

and

$$\Psi_{\Delta L}(t) = \Psi'_{\Delta L}(t). \tag{11}$$

Because the allowed values of the preprocessing parameters τ (exponential averaging) and g (envelope weighting) include the entire physical range, the transformed interaural differences admit the possibility of no transformation. The exception is in the laterality compression, which was always applied to models 1–7.

B. Models for binaural combination

Ten different binaural combination models with adjustable parameters were studied. Each model produced a decision statistic intended to predict the detectability of incoherence. The models and their parameters were then independent variables in regressions comparing predictions with listener detection performance.

The models tested three different hypotheses concerning binaural combination: (1) the independent-interaural-difference or independent-centers model, (2) the lateral-position or lateralization model, and (3) the short-term cross-correlation model. In models of the independent-difference type, averaged fluctuations in IPD and in ILD are combined with a relative weighting parameter a. In models of the lateral-position type, an image location is calculated based on IPD and ILD values that are combined with a time/intensity trading parameter b. The decision statistic is then based on fluctuations in that location. In the short-term-cross-correlation models, only the IPD is used, as will be shown later in this section.

The models are based on transformed (i.e., preprocessed) values of IPD and ILD combined in different ways. It should be noticed that there is no important distinction between the IPD and the interaural time difference (ITD) in this work. With bandwidths as narrow as ours, the ITD can

be determined from the IPD by dividing by the band center frequency of 500 Hz. Consequently, although the models are expressed in terms of IPD, they could equally well be expressed in terms of ITD with no important changes.

Model 1: Sum of interaural differences. A simple model of the independent-interaural-difference type hypothesizes that incoherence is detected on the basis of a linear combination of the standard deviation in transformed IPD and the standard deviation in transformed ILD. The standard deviation of a transformed interaural difference is

$$s_t[\Psi] = \sqrt{\frac{1}{T}\int_0^T [\Psi(t) - \overline{\Psi}]^2 dt}, \tag{12}$$

where $\Psi(t)$ is either the transformed IPD or ILD, and the integral spans the entire stimulus of duration T. Therefore, the sum of transformed standard deviations of IPD and ILD is

$$d_1 = a s_t[\Psi_{\Delta\Phi}] + (1-a) s_t[\Psi_{\Delta L}]. \tag{13}$$

This model has three free parameters: a, τ, and g. The *nontransformed* fluctuations in IPD and ILD were, in fact, the basis for choosing stimuli in Articles I and II. There, it was found that larger values of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ correlated with a greater detectability of incoherence in noises for a given value of coherence.

Model 2: Sum of mean square variations. As a close relative to the decision statistic d_1, an independent-differences model could use the square of the fluctuation, as introduced by Isabelle and Colburn (1987) in connection with a masking level difference experiment with reproducible stimuli,

$$d_2 = a s_t^2[\Psi_{\Delta\Phi}] + (1-a) s_t^2[\Psi_{\Delta L}]. \tag{14}$$

This model has the same free parameters as model 1. This model "…intended to capture the subjective increase in image width caused by the addition of a target tone to the narrowband masker…" (Isabelle and Colburn, 2004).

Model 3: Sum of integrations. An alternative decision statistic is based on an integration of the absolute value of the IPD and ILD over the duration of the stimulus. In this model the contributions of the IPD and ILD are computed separately,

$$d_3 = a\frac{1}{T}\int_0^T |\Psi_{\Delta\Phi}(t)|dt + (1-a)\frac{1}{T}\int_0^T |\Psi_{\Delta L}(t)|dt. \tag{15}$$

This model has the same free parameters as model 1. We do not know of any precedent for such an independent-integration model in the literature.

Model 4: Sum of threshold deviations. A fourth kind of decision statistic measures the fraction of the time that interaural differences are far from zero (the center position). This thresholded statistic is defined as

$$d_4 = a\frac{1}{T}\int_0^T W[h,\Psi_{\Delta\Phi}(t)]dt + (1-a)\frac{1}{T}\int_0^T W[h,\Psi_{\Delta L}(t)]dt, \tag{16}$$

where

$$W[h,\Psi(t)] = \begin{cases} 1 & \text{if } \Psi(t) \geqslant h \\ 0 & \text{if } \Psi(t) < h. \end{cases} \tag{17}$$

In addition to the same three free parameters of other models, model 4 has a fourth free parameter, h, to set the level of threshold. Since both transformed interaural parameters are on the same scale of lateral position, it was assumed that the threshold is the same for both interaural differences.

Webster (1951) proposed a similar model that used only deviations of IPD to determine the influence of interaural phase on masking thresholds. Our model permits large deviations in either IPD or ILD to be the basis for incoherence detection. Model 4 reduces to Webster's model for $a=1$.

Model 5: Standard deviation of the lateral position. Model 5 comes from a suggestion by Hafter (1971) that a signal might be detected by a shift in the lateral position of an image formed by combining IPD and ILD with a time-intensity trading ratio. Model 5 is the first model of three in this article based on a time-varying lateral position, and it hypothesizes that the standard deviation of fluctuations in the lateral position describes incoherence detection. The key distinction is that in a lateral-position model, a fluctuation in phase can cancel a fluctuation in level, but such cancellation is not possible in an independent-centers model such as models 1–4. The lateral position itself can be defined as

$$\Psi_z(t) = b\Psi_{\Delta\Phi}(t) + (1-b)\Psi_{\Delta L}(t), \tag{18}$$

where b is a dimensionless time-intensity trading parameter for transformed interaural differences. The overall time-intensity trading ratio is a combination of b and the laterality-compression factors in Eqs. (7) and (8). Then the standard deviation of the lateral position becomes

$$d_5 = s_t[\Psi_z(t)] = s_t[b\Psi_{\Delta\Phi}(t) + (1-b)\Psi_{\Delta L}(t)], \tag{19}$$

where there are three free parameters: b, τ, and g.

Model 6: Integration of the lateral position. The particular model that Hafter proposed in 1971 was actually a model based on the integrated absolute value of lateral-position incorporating time-intensity trading. Converted to use transformed variables, the model gives

$$d_6 = \frac{1}{T}\int_0^T |\Psi_z(t)|dt = \frac{1}{T}\int_0^T |b\Psi_{\Delta\Phi}(t) + (1-b)\Psi_{\Delta L}(t)|dt. \tag{20}$$

Here, the instantaneous lateral position corresponds to the fluctuation because it is assumed that the undisplaced position corresponds to $z=0$. This model has the same free parameters as model 5.

Model 7: Threshold deviation of the lateral position. Deviations that exceed a threshold value constitute events, and the durations of these events are summed in a decision statistic given by

$$d_7 = \frac{1}{T}\int_0^T W[h, \Psi_z(t)]dt. \tag{21}$$

As for model 4, W has the value 1 if Ψ_z is greater than h and is zero otherwise. In addition to the three free parameters of the other lateral-position models (models 5 and 6), model 7 has a fourth free parameter, h, to set the level of threshold.

Model 8: rms deviation of the short-term cross-correlation function. In connection with the MLD, Osman (1971) proposed a model based on the interaural cross-correlation computed over the entire observation interval. An alternative computes the cross-correlation as a function of running time t,

$$\gamma(t) = \frac{\int_{t-\Delta t}^{t} x_L(t')x_R(t')dt'}{\sqrt{\int_{t-\Delta t}^{t} x_L^2(t_1)dt_1 \int_{t-\Delta t}^{t} x_R^2(t_2)dt_2}}, \tag{22}$$

where $x_L(t')$ is the left-channel waveform and $x_R(t')$ is the right-channel waveform. This cross-correlation function is evaluated at zero lag because an incoherence detection experiment includes no offset ITD. The integration window Δt is brief. For instance, Isabelle and Colburn (2004) take it to be the inverse of the center frequency of the noise band.

Rewritten in terms of the Hilbert envelope and phase, the running cross-correlation is

$$\gamma(t) = \frac{\int_{t-\Delta t}^{t} E_L(t')E_R(t')\cos[\omega t' + \Phi_L(t')]\cos[\omega t' + \Phi_R(t')]dt'}{\sqrt{\int_{t-\Delta t}^{t} |E_L(t_1)|^2\cos^2[\omega t_1 + \Phi_L(t_1)]dt_1 \int_{t-\Delta t}^{t} |E_R(t_2)|^2\cos^2[\omega t_2 + \Phi_R(t_2)]dt_2}}. \tag{23}$$

Isabelle and Colburn (2004) showed that $\gamma(t)$ is approximately given by the cosine of the instantaneous interaural phase difference when the bandwidth is small, as for Experiment 1. For the bandwidth of 14 Hz, the Hilbert envelope and phase should vary on the time scale of $1/14 = 74$ ms. For the center frequency of $f_c = 500$ Hz, this time scale is much slower than the period, $\Delta t = 2$ ms. It then can be assumed that E_L, E_R, Φ_L, and Φ_R are approximately constant over the integration intervals in Eq. (23). Over one period of the stimulus the denominator reduces to the product $E_L E_R$, which then cancels the envelope factors in the numerator. Therefore the short-term cross-correlation (STCC) function can be approximated as

$$\gamma(t) \approx \cos \Delta\Phi(t). \tag{24}$$

The deviation from the diotic value is $1 - \gamma(t)$ and the transformed deviation is

$$\Psi_{CC}(t) = \hat{e}\{1 - \cos[\Delta\Phi(t)]\}w_g(t). \tag{25}$$

The transformed deviation in Eq. (25) includes exponential temporal averaging, which potentially reduces the effectiveness of brief lateral excursions, and it incorporates critical envelope weighting wherein a deviation from perfect correlation is not noticed if an envelope becomes too small.

The root-mean square of the transformed deviation then forms a decision statistic

$$d_8 = \left[\frac{1}{T}\int_0^T \Psi_{CC}^2(t)\, dt\right]^{1/2}. \tag{26}$$

Like the other models, the short-term cross-correlation incorporates temporal averaging and envelope weighting. Unlike the other models, it does not include laterality compression so that the interaural phase remains in units of radians. Because d_8 does not include any form of ILD, it does not include IPD-ILD weighting, and it has only two free parameters, τ and g. Unlike models 1 and 5, which compute a standard deviation, the decision statistic d_8 was computed as a deviation of Ψ_{CC} from zero to represent the deviation from a diotic noise.

Model 9: Integration of the short-term cross-correlation function. Just as models 3 and 6 integrated the absolute value of model percepts, model 9 integrates the absolute deviation,

$$d_9 = \frac{1}{T}\int_0^T \Psi_{CC}(t)\, dt. \tag{27}$$

By definition, $\Psi_{CC}(t) \geqslant 0$. Model 9 has the same free parameters as model 8, τ and g. Also, like model 8, laterality compression was not included in the transformed variable so that $\Delta\Phi(t)$ is in radians.

Model 10: Threshold of the short-term cross-correlation function. Just as models 4 and 7 were based on thresholded values of model percepts, model 10 integrates a thresholded short-term cross-correlation,

$$d_{10} = \frac{1}{T}\int_0^T W[h, \Psi_{CC}(t)]dt. \tag{28}$$

Model 10 has three parameters, τ, g, and h, where h sets the *magnitude* of the threshold deviation. Per Eq. (25) the magnitude of the deviation, $\Psi_{CC}(t)$, can be as large as 2—the difference between $\cos(0)$ and $\cos(\pi)$. As in the other STCC models, the laterality compression was omitted.

C. Models compared with Experiment 1

The ten models presented above were tested against the data from Experiment 1. A linear regression of the form $y = mx + b$ was used to evaluate the effectiveness of a model to describe incoherence detection. The y variable was the CAS for the individual listeners or for an average over listeners. The x variable was d_n from one of the ten models. Figure 3 shows example regressions for model 1. The solid line is the line of best fit. The dotted lines have the same slope as the

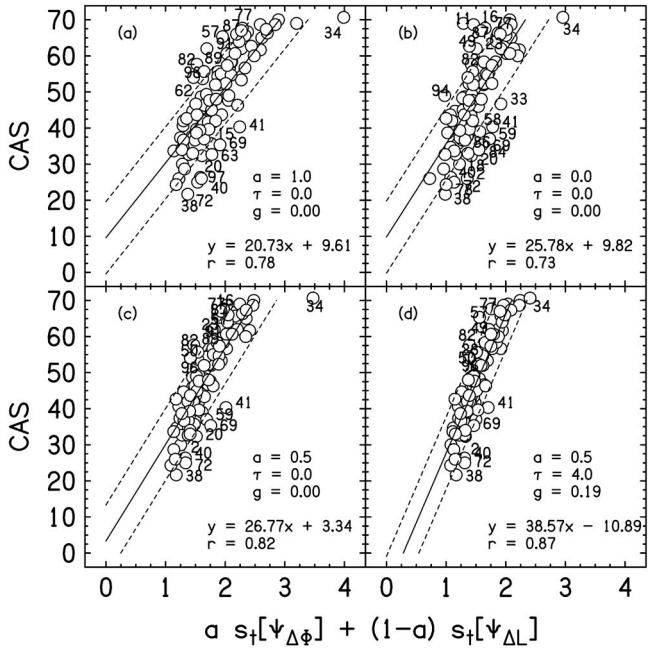


FIG. 3. Example linear regressions for model 1. Laterality compression is applied to all the noises. The solid line is the line of best fit; the equation and value of τ are reported. The dotted lines have the same slope as the solid line, but have intercepts that differ by ±10 CAS. Noises that fall outside the dotted lines are numbered by the serial number in Fig. 1. The plots show the advantages of (c) using both IPD and ILD, and (d) envelope weighting and temporal averaging.

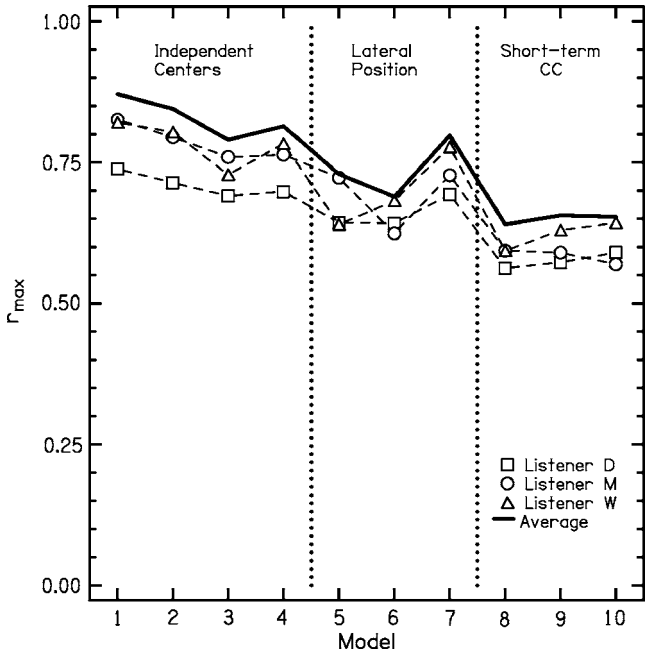


FIG. 4. The comparison of CAS scores for the 14-Hz noises of Experiment 1 with ten models. The value of r_{max} shows the correlation between the experimental CAS scores for the 100 noises and predictions by each model, optimized by adjusting the model parameters. The solid line represents a fit to the data of the average listener. It is not the average of r_{max} averaged over the listeners.

solid line but are displaced vertically by plus or minus 10 CAS units. Noises falling outside of the dotted lines region are numbered on the plot. Figure 3(a) shows model 1 using only the laterality-compressed IPD ($a=1$) without exponential averaging and threshold weighting. Figure 3(b) shows model 1 using only the laterality-compressed ILD ($a=0$). Figure 1 shows that noise 57 has large fluctuations in IPD and ILD but experiments, including those of Article I, showed that listeners found it relatively difficult to detect the incoherence in this noise. By contrast, Fig. 3(a) shows stimulus 57 to the left of the line of best fit. This shows that when the laterality compression is included in the model, the fluctuations are actually comparatively small, which is more in line with the detection data. Figures 3(c) and 3(d) show equal weighting of IPD and ILD ($a=0.5$), respectively, without and with temporal averaging and envelope weighting.

The linear correlation coefficient, r, was used to compare the results of the regressions. The maximum r, r_{max}, was found by independently varying all the free parameters over a reasonable space. For example, model 1 has three free parameters a, τ, and g. The range of a was 0 to 1 with a 0.01 increment; the range of τ was 0 to 10 ms with a 0.5-ms increment (tests with larger values of τ will be described later); the range of g was 0 to 0.5 with an increment of 0.01. Therefore, for model 1, 400 000 linear regressions were performed ($100\times20\times50\times4$ listeners). For the threshold models 4 and 7, the range of h was 0 to 10 with a 0.25 increment. (Recall that the laterality-compressed IPD and ILD are on a scale of −10 to 10.) For threshold model 10, the range of h was 0 to 2 with a 0.01 increment. A power law regression equation was also used to fit the data, but it did not improve the correlation between the experimental data and the models.

1. Comparison of model types

The results of the regressions are shown in Fig. 4 for the best combination (largest r) of all the parameters for each model. Figure 4 presents results for individual listeners and for the average listener. Therefore, the r_{max} of the averaged data is not the average r_{max} of the listeners. It is interesting that the most successful models agree better with the average listener than they do with any single listener. A similar result was found by Isabelle and Colburn (2004) in modeling MLDs for reproducible noise. Given that the models are simple signal processing algorithms whereas human listeners have complicated individual tendencies, that is the sort of result that one would expect from a model that correctly represents the general population.

Figure 4 shows that the models of the independent-interaural-difference type (models 1–4) were more successful than the lateral-position types (models 5–7) with the exception that model 7 had a larger r_{max} than model 3. Least successful were the STCC models (models 8–10). Since the STCC depends entirely on $\Delta\Phi(t)$, this may be evidence for an important contribution of $\Delta L(t)$ to incoherence detection.

Model 1 had the largest r_{max} for all three listeners and for the averaged data. For the averaged data, $r_{max}=0.87$. The performance of model 2 is very similar to that of model 1, except that the r_{max} is always slightly smaller for model 2.

Table I shows the values of the free parameters that maximized r for the 14-Hz bandwidth modeling. Table I

TABLE I. Values of free parameters that optimize r in modeling the detection results of Experiment 1 with 100 noises with a bandwidth of 14 Hz. Parameter τ is the exponential window time constant. Parameters a and b weight IPD and ILD contributions, with extremes of 1 and 0 equivalent to IPD only and ILD only, respectively. Parameter g is the envelope threshold for discounting IPD. The IPD is ignored if the envelope in either the left or right channel is less than g times the overall rms value. Parameter h is a threshold for models that measure the duration of time the function is greater than the threshold value. Threshold is lateral position for models 4 and 7; it is deviation from perfect coherence for model 10.

Model	Listener	τ (ms)	a,b	g	h
d_1	D	3.0	0.53	0.17	⋯
	M	5.0	0.45	0.21	⋯
	W	0.5	0.53	0.23	⋯
	Ave	4.0	0.50	0.19	⋯
d_2	D	3.0	0.52	0.17	⋯
	M	4.0	0.46	0.23	⋯
	W	0.5	0.54	0.23	⋯
	Ave	3.0	0.50	0.23	⋯
d_3	D	1.0	0.62	0.04	⋯
	M	4.5	0.44	0.15	⋯
	W	6.5	0.51	0.12	⋯
	Ave	3.5	0.50	0.12	⋯
d_4	D	2.0	0.78	0.04	4.00
	M	0.0	0.48	0.11	3.75
	W	0.5	0.59	0.12	3.25
	Ave	0.5	0.66	0.04	3.75
d_5	D	1.0	0.00	0.07	⋯
	M	3.5	0.00	0.12	⋯
	W	3.5	0.00	0.12	⋯
	Ave	3.0	0.00	0.11	⋯
d_6	D	1.0	0.97	0.04	⋯
	M	2.0	0.11	0.15	⋯
	W	0.0	0.42	0.15	⋯
	Ave	0.5	0.91	0.04	⋯
d_7	D	2.0	0.99	0.04	4.00
	M	1.0	0.96	0.04	3.75
	W	0.0	0.54	0.24	2.50
	Ave	1.0	0.88	0.04	3.75
d_8	D	0.0	⋯	0.28	⋯
	M	0.0	⋯	0.28	⋯
	W	0.0	⋯	0.27	⋯
	Ave	0.0	⋯	0.28	⋯
d_9	D	6.0	⋯	0.11	⋯
	M	6.0	⋯	0.11	⋯
	W	6.0	⋯	0.27	⋯
	Ave	6.0	⋯	0.11	⋯
d_{10}	D	7.5	⋯	0.00	0.06
	M	0.0	⋯	0.00	0.06
	W	6.5	⋯	0.00	0.08
	Ave	6.0	⋯	0.00	0.06

shows that the parameters are similar for different listeners over the different types of models (independent-centers, lateral-position, and STCC). Consequently, the fits to the average listener shown in Fig. 4 are meaningful. Table I also shows that fitting parameters that optimize r are similar across models, to the extent that the models permit them to be compared.

2. Optimized parameters for model 1

The most successful model was model 1, and Fig. 5 shows how the free parameters covary to maximize r for the average listener in that model. Figure 5 was generated by varying two parameters and keeping the other constant at the optimum value. Plots of a vs τ assume that $g=0.19$, plots of g vs τ assume that $a=0.50$, and plots of a vs g assume that $\tau=4$.

The results of Fig. 5 can be summarized as follows:

a. Integration time. The greatest r occurs for an integration time of $\tau=4$ ms. However, r was quite insensitive to τ over the 0–10 ms range tested in detail. Apparently, the characteristic stimulus fluctuations, expected to be of order 1/14 s, are slow enough that they do not challenge a system with time constants of this order. Longer integration times will be discussed later.

b. Critical envelope weighting. According to the regression analysis, the best critical envelope weight for model 1 is $g=0.19$, though the r_{max} is insensitive to g in this vicinity (approximately $0<g<0.3$). This result indicates that there is a modest benefit on the average of ignoring phase differences that coincide with a particularly small envelope in one or both of the ears.

A greater benefit from envelope weighting is seen when one tries to predict detection for individual waveforms. The weighting omits large phase fluctuations that occur during the onset and offset of the stimulus, where the temporal shaping is applied and the envelope is small. One would expect that even large phase fluctuations during these times would often be missed by the listener because they occur at the very beginning or end of the stimulus. This benefit can also be seen in the lower panels of Figure 3. In Fig. 3(c) (without envelope weighting) there were nine stimuli to the right of the rightmost dotted line. In Fig. 3(d) (with envelope weighting), two stimuli moved within the dotted line region and the rest of these points moved close to the dotted line.

The envelope weighting applied in Eq. (9) is a simple on/off type. Other envelope weighting functions were tried—linear envelope weighting of the IPD and squared envelope weighting of the IPD—but including these functions led to lower r values than did the on/off type envelope weighting.

c. Relative IPD-ILD contributions. The regression analysis for the average listener data found that the best value of a for model 1 is 0.50, which means that transformed IPD and ILD values contribute equally to the sensation of incoherence. Because the transformed values were scaled by Eqs. (7) and (8), we interpret this equality to mean that the scaling correctly represents the relative perceptual importance of IPD and ILD. Because the scaling was derived from Yost's steady-state sine experiment results, we conclude that it is valid to extend the results of steady-state measurements to the case of slowly fluctuating interaural differences.

3. Optimized parameters for other models

a. Longer integration times. Longer integration times were also tested for models 1 and 3–10 (model 2 was omitted because the results were so similar to model 1). Longer times were tested because the oscillating coherence experiments of Grantham and Wightman (1978) led to binaural time constants as long as 64 ms (−3 dB response at 2.5 Hz), an effect commonly called "binaural sluggishness." MLD experiments using a masker with temporally varying coherence (Grantham and Wightman, 1979) led to time constants that were even longer. Although the phenomenon of binaural

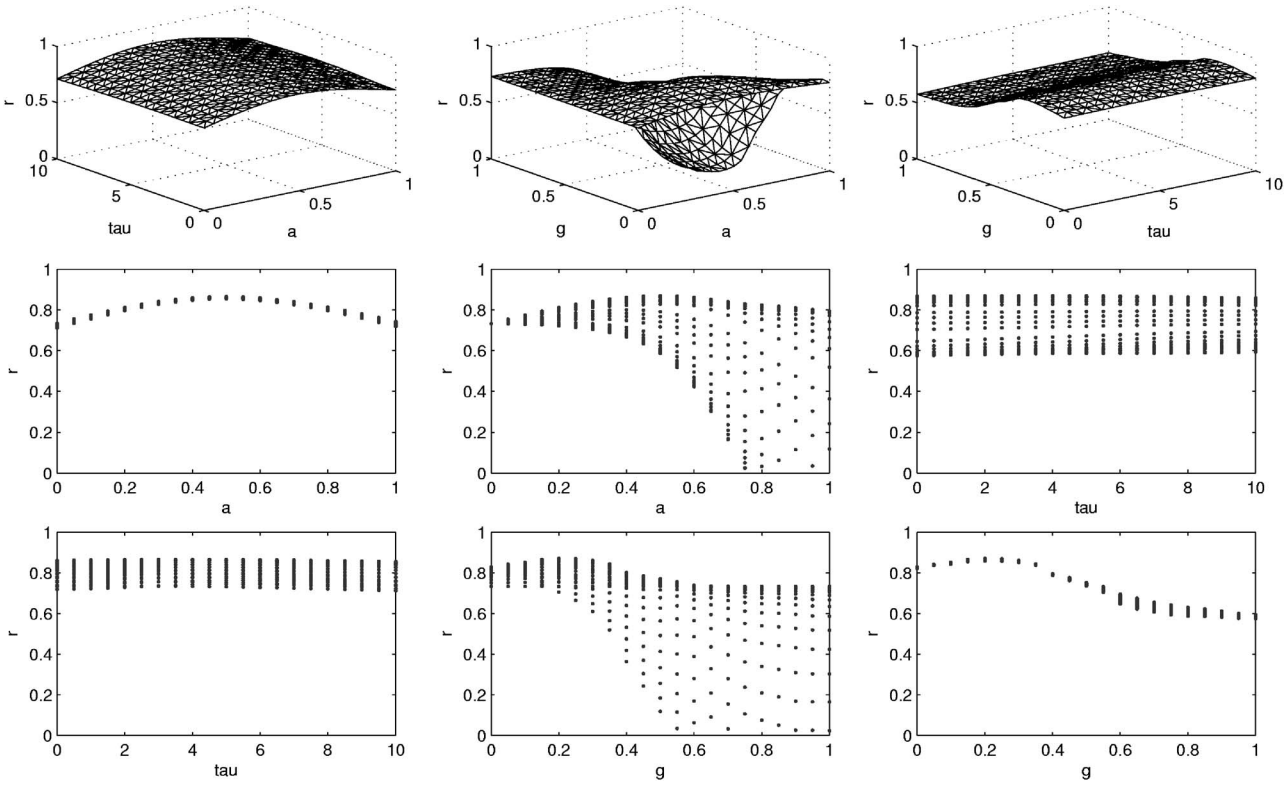


FIG. 5. The free parameter surfaces for fitting model 1 to the average listener data from Experiment 1—100 noises with 14-Hz bandwidth. In the upper-left panel, r is plotted against a and τ for $g=0.19$. In the upper-middle panel, r is plotted against a and g for $\tau=4$ ms. In the upper-right panel, r is plotted against g and τ for $a=0.5$ ms. The two panels below an upper panel flatten one of the free parameter dimensions. The variations of free parameters create smooth surfaces. At this bandwidth, the dependence upon τ is negligible.

sluggishness probably does not indicate an inertia affecting all binaural temporal variations (Hall *et al.*, 1998), we performed spot checks at 50-ms intervals (25 ms for model 1), trying to fit CAS data with fluctuations that had longer values of τ applied to the stimuli. Figure 6 shows that the value of $r_{\max}$ decreases monotonically for increasing values of τ for models 1 and 3–7. For model 1, the value of $r_{\max}$ dropped from $r_{\max}\approx 0.9$ for $\tau=0$ ms, to $r_{\max}\approx 0.6$ for $\tau=150$ ms. We conclude from Fig. 6 that there is no useful role for binaural sluggishness. For models 1 through 7, incorporating sluggishness through large τ leads to worse agreement with the experiment. For cross-correlation models 8–10, increasing values of τ lead to negligible change in agreement. However, we note that the approximation made in Eq. (24) assumed a short-analysis window. The validity of this assumption and the use of large values of τ will be addressed in Sec. VI.

b. Order of operations. The calculations described above applied temporal averaging to the physical stimulus, then applied laterality compression. However, it is not clear that this is the correct order of operations. Therefore, the models were rerun, first applying laterality compression and then temporal averaging. The integration times used were both the fine-scale (0–10 ms in 0.5-ms steps) and the longer times (50, 100, 150 ms). It was found that the order of operations did not matter for τ ranging from 0 to 10 ms in that the value of r changed by less than 0.01. Reversing the order of operations for the longer integration times always led to smaller r values compared to those in Fig. 6. The reduction could be as much as 0.2. Therefore, the best model applies laterality compression to signals that have been temporally averaged at a previous stage of processing.

c. Lateral-position models. Table I shows that lateral-position model 5 favors the transformed ILD over the transformed IPD in fitting the average listener data ($b=0$). The other lateral-position models, 6 and 7, mostly favor the transformed IPD ($b\approx 1$). This is in contrast to the independent-centers models (1–4), that weigh IPD and ILD as equally important. However, a lateral-position model that uses only IPD ($b=1$) or only ILD ($b=0$) is equivalent to an independent-centers model that uses only IPD ($a=1$) or only ILD ($a=0$). For example, an independent-centers model that incorporates IPD fluctuations separately must lead to an r value that is at least as large as the r for the lateral-position model with $b=0$ or 1. Therefore, model 1 must perform as least as well as model 5 (where b equals zero) in Fig. 4.

d. Short-term cross-correlation. The STCC models (models 8–10) correlated least well with the data, possibly because only the IPD is used in these decision statistics. Model 10 produced an interesting result in that the optimum threshold magnitude is approximately 0.06 (Table I), which means that listeners detected brief decorrelations from unity at coherence values of 0.94. This is larger than the jnds found by Gabriel and Colburn (1981). However, that work measured the jnd for the entire duration of the stimulus and not decorrelations over a short time interval. Therefore, 0.94 seems like a reasonable result. Even though the STCC models correlate with the data least well, it appears that the approximation that yields Eq. (24) is not a bad approximation.

4. The advantage of preprocessing

The results of modeling the data with the preprocessing removed (no temporal averaging, no laterality compression,

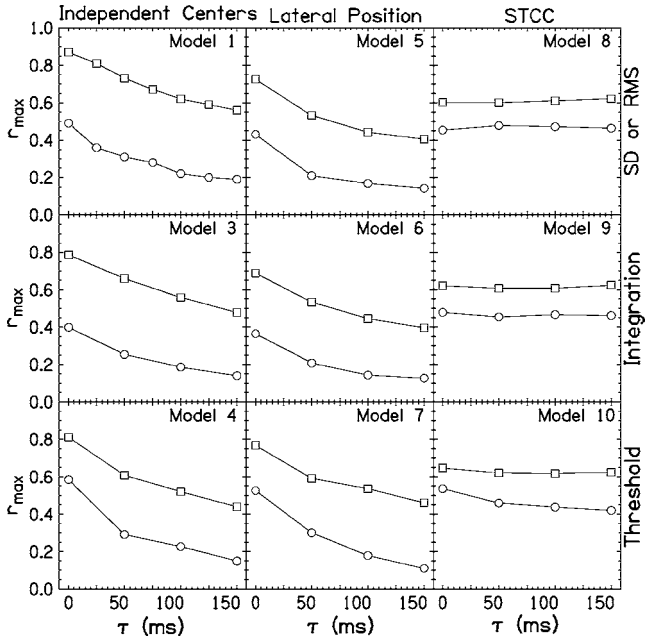


FIG. 6. The results of using long integration times for models 1 and 3–10. The squares show r_{max} for the 14-Hz data (Experiment 1), the circles show r_{max} for the 108-Hz data (Experiment 2). The top row is for standard-deviation models 1 and 5 or rms model 8. The middle row is for models that integrate absolute values. The bottom row is for threshold models.

and no critical envelope weighting) yielded r_{max}=0.69 for the independent-centers model for the averaged listener data to be compared with 0.87 with preprocessing included, i.e., model 1. Thus, these preprocessing assumptions prove beneficial in our modeling attempts. Without preprocessing r was maximized by weighting ILD and IPD fluctuations by the ratio of 0.14 dB/deg. Yost and Hafter (1987) reported that the trading of intensity and phase should be 0.10 dB/deg for interaural phases less than 90° and should be 0.08 or 0.10 dB/deg for interaural phases greater than 90°. Our ratio is higher than that of Yost and Hafter, but not much higher. The difference may arise because our experiment has dynamic fluctuations, whereas Yost and Hafter analyzed static interaural differences.

IV. EXPERIMENT 2: 108-Hz BANDWIDTH

After testing the ten models against the 14-Hz bandwidth noises and coming to some preliminary conclusions, we wondered how the models would perform for the wider bandwidth of 108 Hz.

A. Method

The 100 noises used in Experiment 2 are described in Fig. 7 along with the means, standard deviations, and correlation of interaural parameters. It was the same collection from which particular noises were selected in Article I. The same three listeners participated and the same procedure was used.

B. Results

The results of Experiment 2 are shown in Fig. 8, entirely parallel to Fig. 2 for Experiment 1. For Experiment 2 the

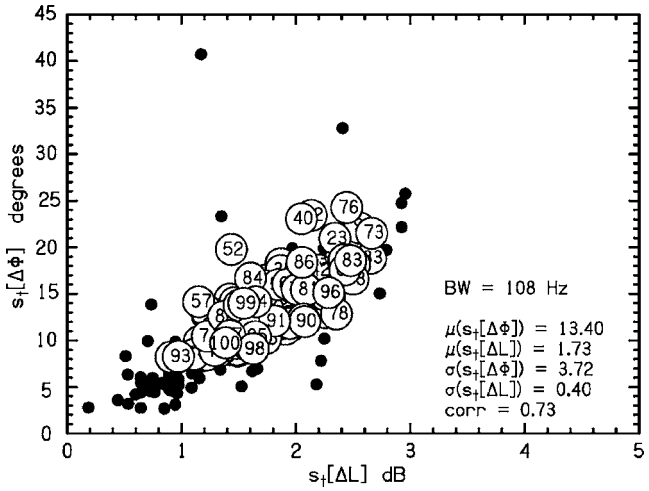


FIG. 7. Fluctuations of IPD vs fluctuations of ILD for the 100 reproducible noises with a 108-Hz bandwidth used in Experiment 2. Each noise is labeled by a serial number. The means, standard deviations, and IPD-ILD correlation of the distributions are reported. The means remained about the same as in Fig. 1, but the standard deviations decreased. Closed symbols replot the data of Fig. 1 for comparison.

percentage of correct responses shows a ceiling effect but the CAS does not. Unlike Experiment 1, where the ceiling was reached for particular noises for all the listeners, the ceiling for P_c in Experiment 2 was a factor for listeners D and M, but not necessarily W.

1. Comparison of model types

The ten models tested with Experiment 1 were also tested with Experiment 2. The results of the regression analysis for Experiment 2 are shown in Fig. 9, parallel to Fig. 4 for Experiment 1. The values of r_{max} are smaller than in Fig. 4 because there is less variation in detectability for a band as wide as 108 Hz compared to a band with a 14-Hz width. Model 4 gave the largest r_{max} for all three listeners and for the averaged data. For the averaged data, r_{max}=0.59. It is

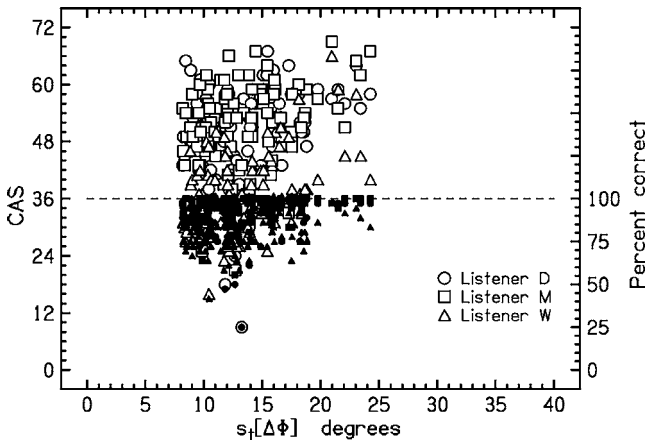


FIG. 8. All the detection data for the 100 noises from Experiment 2 for three listeners, D, M, and W, are plotted twice, once as the number correct—on a scale from 0 to 36, (0 to 100%) and once as CAS—on a scale from 0 to 72. The data are plotted as a function of the standard deviation of the interaural phase as in Fig. 2.

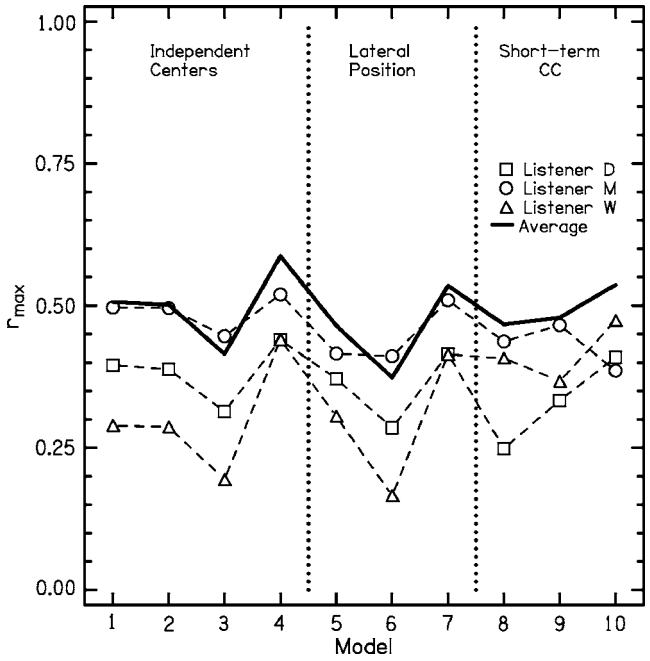


FIG. 9. The comparison of CAS scores for the 108-Hz noises of Experiment 2 with ten models. The value of r_{max} shows the correlation between the experimental CAS scores for the 100 noises and the best fit for each model, optimized by adjusting the model parameters. The solid line represents a fit to the data of the average listener. The values of r_{max} are not as high as for the 14-Hz data shown in Fig. 3.

possible that r values are small because the stimulus bandwidth is wider than the relevant auditory analysis filter. Auditory filtering will be discussed later.

Table II shows the values of the free parameters that maximized r for the 108-Hz bandwidth modeling. Table II shows that the parameters for the models at the 108-Hz bandwidth are mostly similar for different listeners. Consequently, the fits to the average listener shown in Fig. 9 are meaningful, although less convincing than the 14-Hz bandwidth fits.

Figure 9 shows that the ten models all perform about equally well. For the average listener, the highest r value, 0.59, is not much greater than the lowest, 0.37. As for Experiment 1, the most successful model is of the independent-centers type, but the threshold models (4, 7, and 10) outperform the others—even model 1, the most successful model in Experiment 1. The reason for this may be that models 4 and 7 have an extra free parameter, but model 10 has the same number of free parameters as model 1. This could be evidence that a threshold statistic is used for detecting incoherence in larger bandwidth stimuli. Again, the models account better for the average listener than for any individual listener, with a few exceptions.

2. Optimized parameters for model 4

Figure 10 shows how the four free parameters covary in the exploration of model 4, the best model from Experiment 2. Each panel shows how two free parameters change while two others are kept constant. As in Fig. 5, the constant parameters are set equal to the free parameters that yield r_{max}. When kept constant, $a=0.48$, $\tau=0.5$ ms, $g=0.15$, and $h=6.0$.

Figure 10(a) shows a vs τ. It shows that model 4 is fairly insensitive to changes in a, but there is a peak near $a=0.5$, consistent with the modeling from Experiment 1. Also, the calculation leads to positive r values only when τ is rather brief, 1.5 ms or less. Figure 10(b), a vs g, again shows that the model is fairly insensitive to a. Also, the largest values of r occur for g near 0.15.

Figure 10(c) shows that the optimum h is insensitive to different values of a. Also, it is possible to see the sharp drop off of r for values of h greater than 6. Figure 10(d) shows that the best value of τ is 0.5 ms and small values of g lead to the highest r values.

TABLE II. Values of free parameters that optimize r in modeling the detection results of Experiment 2 with 100 noises with a bandwidth of 108 Hz. Parameters are defined in the caption to Table I.

Model	Listener	τ (ms)	a,b	g	h
d_1	D	4.5	0.71	0.02	⋯
	M	0.5	0.50	0.02	⋯
	W	0.5	0.69	0.03	⋯
	Ave	1.5	0.54	0.03	⋯
d_2	D	4.5	0.68	0.02	⋯
	M	0.5	0.46	0.02	⋯
	W	0.5	0.63	0.03	⋯
	Ave	1.5	0.49	0.03	⋯
d_3	D	2.5	0.47	0.06	⋯
	M	1.0	0.29	0.02	⋯
	W	1.0	0.45	0.00	⋯
	Ave	1.5	0.36	0.03	⋯
d_4	D	0.0	0.60	0.18	5.50
	M	0.0	0.42	0.15	6.50
	W	0.0	0.60	0.17	7.00
	Ave	0.5	0.48	0.15	6.00
d_5	D	1.5	0.00	0.10	⋯
	M	1.5	0.00	0.00	⋯
	W	0.5	0.00	0.08	⋯
	Ave	1.0	0.00	0.08	⋯
d_6	D	5.5	1.00	0.03	⋯
	M	1.0	0.00	0.00	⋯
	W	6.0	0.17	0.15	⋯
	Ave	1.0	0.00	0.00	⋯
d_7	D	0.0	0.34	0.19	4.25
	M	0.5	0.38	0.26	4.00
	W	2.0	0.86	0.00	6.50
	Ave	1.0	0.97	0.01	5.75
d_8	D	0.0	⋯	0.04	⋯
	M	0.0	⋯	0.45	⋯
	W	0.0	⋯	0.00	⋯
	Ave	0.0	⋯	0.04	⋯
d_9	D	0.0	⋯	0.18	⋯
	M	0.0	⋯	0.06	⋯
	W	0.0	⋯	0.06	⋯
	Ave	0.0	⋯	0.06	⋯
d_{10}	D	0.0	⋯	0.24	0.13
	M	0.0	⋯	0.09	0.24
	W	0.0	⋯	0.11	0.30
	Ave	0.0	⋯	0.09	0.23

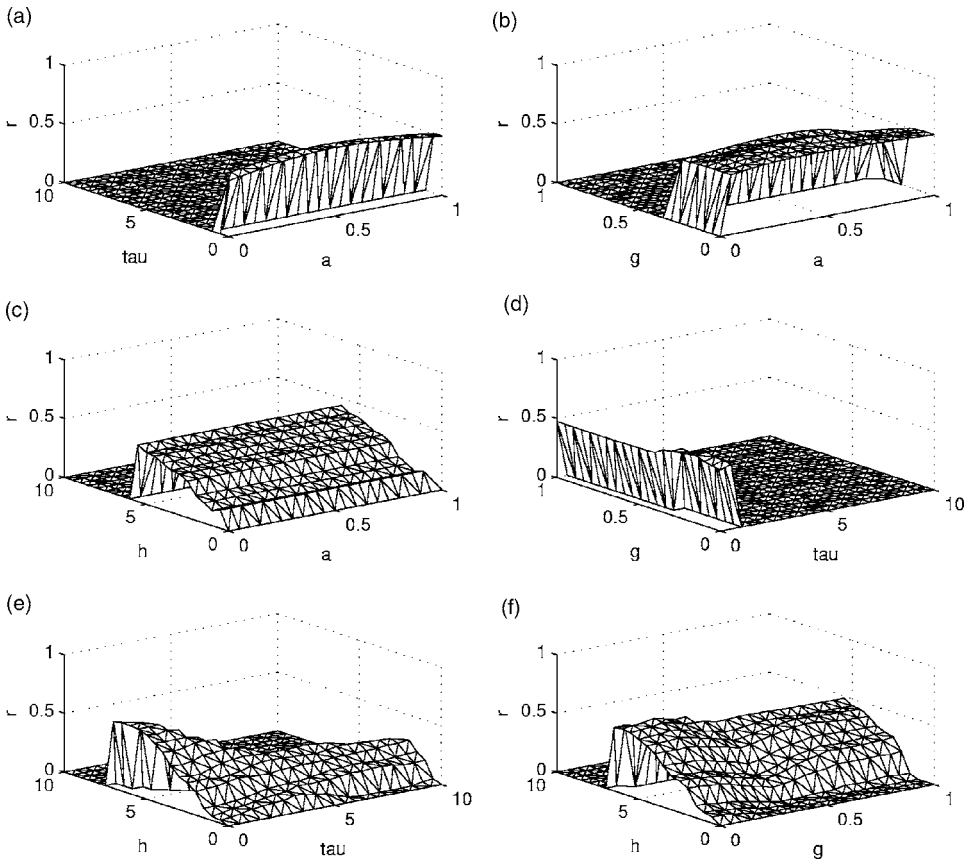


FIG. 10. The free parameter surfaces for fitting model 4 to the average listener data from Experiment 2 with 108-Hz bandwidth. Fixed parameters, not appearing along the axes, were given the optimum values for the average listener for model 4 in Table II. Results for model 4 show a strong interaction between h and g, a strong interaction between h and τ, and insensitivity to a.

Figure 10(e) shows that for no temporal averaging ($\tau=0$), model 4 leads to positive r values for all values of h up to 7.5. However, as the fluctuations become smoother due to larger values of τ (fewer peaks above a high threshold), the model leads to negative r values for h greater than 3.

Finally, Fig. 10(f) shows a rather strong interaction between h and g. The largest values of r occur when h is large and g is near 0.15.

3. Optimized parameters for other models

a. Longer integration times and order of operations. Figure 6 shows r values for Experiment 2 (108-Hz bandwidth) modeled with larger values of τ. As for Experiment 1 (14-Hz bandwidth), the data are described best by integration times less than 10 ms. Also, changing the order of operations showed that the best results occurred for temporal averaging followed by laterality compression, as in Experiment 1.

b. Lateral-position models. Table II shows that the lateral-position models (models 5–7) usually fit the Experiment 2 data best when the values of b are near 0 or 1. A similar result appeared in Table I for Experiment 1. Therefore, these models are most successful when they use only IPD or ILD information. However, a lateral-position model that makes no use of one of the interaural differences is indistinguishable from an independent-centers model. In fact, lateral-position models with $b=1$ are identical to independent-centers models with $a=1$. Therefore, in the case of the 108-Hz bandwidth data, it seems that independent centers may again be the better type of model, even if there is little distinction between models 1–7 by the values of $r_{\max}$.

c. Auditory filtering. A bandwidth of 108 Hz is 8% smaller than a Munich critical band at 500 Hz (Zwicker and Terhardt, 1980), but it is 37% larger than a Cambridge band (Moore and Glasberg, 1983). A possible role for auditory filtering was tested by centering a Cambridge gammatone filter on 500 Hz to filter the noise with 84-Hz bandwidth. Such filtering never increased the r values. Instead the r values decreased by as much as 0.1. As noted in Article I, it does not seem possible to understand our experimental results using a model in which information is confined to a critical band. A similar conclusion in connection with the MLD was reached by Evilsizer *et al.* (2002).

4. The advantage of preprocessing

The preprocessing assumptions of laterality compression, temporal averaging, and critical envelope weighting were removed to gauge their effect in Experiment 2. For model 1 and average listener data from Experiment 2, $r_{\max}=0.47$ without preprocessing can be compared with $r_{\max}=0.50$ with preprocessing included. As for Experiment 1, the preprocessing assumptions improved the agreement between model and data, but the improvement was much smaller than in Experiment 1. The best fit without preprocessing was obtained by weighting ILD and IPD fluctuations in the ratio of 0.08 dB/deg, which impressively matches the ratio suggested in the review by Yost and Hafter (1987), 0.08–0.10 dB/deg.

V. REPRODUCIBILITY

Prior to Experiment 1, a similar experiment was performed with two differences. First, the values of coherence among the 100 noises varied from 0.969 to 0.998, to be compared with 0.992 in Experiment 1. Second, listener D in Experiment 1 was replaced by listener E, female, age 19, and

well-practiced in incoherence detection. This experiment will be called Experiment 0. The distributions of phase and level fluctuations in Experiment 0 were almost identical to those seen in Fig. 1. A regression between the CAS data and the coherence values yielded an $r=0.48$ for the average listener.

The results of Experiment 0 were essentially the same as Experiment 1. Again model 1 emerged as the best model for all listeners and for the average listener with $r_{max}=0.89$, compared to 0.87 for Experiment 1. For Experiment 0, the values of the free parameters were $\tau=0.5$ ms (vs 4.0 for Experiment 1), $a=0.43$ (vs 0.50 for Experiment 1), and $g=0.17$ (vs 0.19 for Experiment 1). The r_{max} values, as a function of the ten models, for Experiment 0 correlated with those from Experiment 1, as shown in Fig. 4, at 0.90. Other notable results were also consistent between Experiments 0 and 1. Again, the independent-centers models outperformed the lateral-position models, which, in turn, outperformed the STCC models. Again, the model fits to data were insensitive to changes in τ, and there was an important advantage to preprocessing. The value added by the results presented in this section is to demonstrate that major results from Experiment 1 were reproducible with stimuli with different values of coherence but similar fluctuation statistics.

VI. SHORT-TERM CROSS-CORRELATION REVISITED

Articles I, II, and this article used noises that have a fixed value of long-term cross-correlation of 0.992. This value was calculated by cross-correlating the physical signals, x_R and x_L, over the entire duration of the stimulus. The major difference between the long-term and short-term cross-correlation is that the long-term cross-correlation calculation yields a single value whereas the short-term cross-correlation calculation yields a function of time. Models 8–10 approximated a STCC model by using the cosine of the IPD.

In this section, the STCC is computed as a function of time directly from an equation of the form of Eq. (22) without use of the cosine approximation. Again, a linear regression is used to compare models and data, leading to correlation coefficients r.

A. Physiological transformations

To bring the STCC model into line with current models of binaural processing, physiologically motivated stimulus transformations were also included in the model. These calculations were performed to test the possibility that a model using some form of STCC, as it appears in the auditory system after peripheral processing, might be able to fit the experimental results as well as our best model based on interaural fluctuations.

1. Auditory filtering

Breebaart and Kohlrausch (2001) found that using a fourth-order gammatone filter to approximate auditory filtering changed the value of coherence of noises with a bandwidth as small as 10 Hz. Therefore, we tested such a filter in our analysis, expecting it to change the effective coherence from a constant value of 0.992. The center frequency of the filter was 500 Hz.

2. Cochlear compression and rectification

Bernstein *et al.* (1999) were able to account for MLD data from Eddins and Barber (1998) and from Bernstein and Trahiotis (1996), by applying envelope compression and square-law half-wave rectification to simulate the cochlea. Bernstein *et al.* (1999) found that a compression exponent of approximately 0.2 could describe the data. This exponent, together with a square-law rectifier, corresponds to an exponent of 0.4 together with a half-wave rectifier, which agrees with the exponent derived by Oxenham and Moore (1995) and used by van de Par and Kohlrausch (1998).

The analysis in this section used envelope compression of the form

$$x'(t) = [E(t)]^{p-1}x(t), \tag{29}$$

where p is the compression exponent, a value between zero and one. Consistent with previous studies we used $p=0.4$ for both the right and left channels.

After cochlear compression, half-wave rectification was applied to the stimulus to represent the response of haircells. In the first calculation, only the positive portions of the waveform were retained; in the second, only the negative portions. The change from positive to negative led to negligible changes in r values, changes of less than 0.002.

3. Temporal averaging

The duration of the rectangular window Δt in the analog to Eq. (22) was varied from 10 to 300 ms. This range encompasses the values of binaural sluggishness that have been found in varied coherence experiments (e.g., Grantham and Wightman, 1979). The short-term cross-correlation function was calculated for each instant, t, of the 500-ms duration of a noise. When t was less than Δt, the window extended back in time only to the beginning of the noise.

4. Decision making

Since the STCC is a function of time, a mathematical operation was needed to calculate a final decision statistic to compare against the psychological data. The calculations used to obtain a statistic were similar to those used in models 8, 9, and 10—rms deviation, integration, and threshold. The threshold values tested in the threshold models were $h=0.001$, 0.01, 0.025, 0.05, and 0.075.

B. Results and discussion

The results of the regression between the STCC statistics, as calculated with model physiology, and the CAS for the average listener can be seen in Fig. 11. The solid lines are for the 14-Hz bandwidth and the dotted lines are for the 108-Hz bandwidth. Values of r are plotted for two values of the threshold level, $h=0.025$ and $h=0.05$, representative parameters for the best performing threshold models.

For the 14-Hz data, the best model was the threshold model with $h=0.05$ ($r=0.78$), shown in the bottom panel of

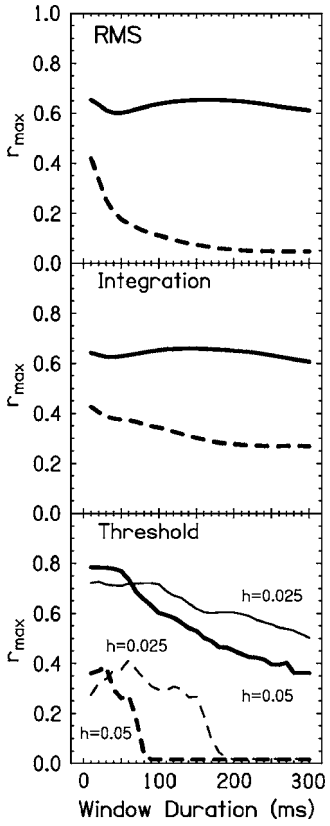


FIG. 11. Performance of the three types of models (rms, integration, and threshold) for the physiologically motivated short-term cross-correlation models for different window durations. Each model incorporates auditory filtering, cochlear compression, and half-wave rectification. The solid lines represent the correlations between the model and the 14-Hz data (Experiment 1). The dashed lines represent the correlations for the 108-Hz data (Experiment 2). There are two threshold calculations for each bandwidth, with $h=0.025$ (thin) or $h=0.05$ (thick).

Fig. 11. However, as noted in Sec. IV, the threshold model might outperform the other models simply because it has an extra degree of freedom. For the 108-Hz data, the best model was the integration model. However, for this bandwidth, there was little difference between the best integration model ($r=0.43$), the best rms model ($r=0.42$), and the best threshold model with $h=0.025$ ($r=0.41$).

As shown in Fig. 11, the best fit to the data used window durations less than 100 ms. The shortest window duration, 10 ms, was optimum in all but one case—the threshold model for the 108 Hz, which had an optimum window of 60 ms for $h=0.025$ and an optimum window of 30 ms for $h=0.05$. Once again, the long integration times that are suggested by binaural sluggishness experiments were not supported by our modeling of incoherence detection data.

In summary, STCC models that include physiologically-based transformations can describe incoherence detection data better than models 8, 9, and 10, which do not have those transformations. However, no STCC model could account for incoherence detection data quite as well as models based on interaural fluctuations. For a bandwidth of 14 Hz, the most successful STCC model produced an r value of 0.78 to be compared with model 1, which led to $r=0.87$. For a bandwidth of 108 Hz, the most successful STCC model produced an r value of 0.43 to be compared with model 4, which led to $r=0.59$.

VII. GENERAL DISCUSSION

The goal of the experiments reported in this article was to discover the signal characteristics and the binaural perceptual operations that enable a listener to detect small amounts of interaural incoherence in bands of noise.

A. Summary

Experiments 1 and 2 used 100 reproducible noises, with bandwidths of 14 and 108 Hz, respectively, to test models of interaural incoherence detection. Each model was rated by varying its parameters to find the best agreement with detection data.

The results of Experiment 1 were the most informative. They showed that the independent-IPD/ILD-centers models outperformed lateral-position models, and that the lateral-position models outperformed the short-term cross-correlation models. The results of Experiment 2 were less informative, mainly because the wider bandwidth led to interaural fluctuations that varied less across different noises. In the comparison of Experiment 2 data with model predictions, the data were less well fitted by the models, the data discriminated among models less clearly, and the best performing models had one or two more parameters than the other models, which may have made the comparison unfair.

Because of the greater power of Experiment 1, with the narrow bandwidth, the general discussion will be mainly concerned with the comparison of models with the results of Experiment 1. The most interesting comparison is between model types.

The first comparison is between models 1–4, which treat fluctuations in IPD and ILD by independent centers, and models 5–7, which consider a fluctuation of the lateral position of the image. Figure 4 for 14-Hz bandwidth shows that the independent-centers models outperform the lateral-position models for all three listeners and for the average listener. Experiment 0 supports this conclusion. Thus, models of the independent processing type are favored unambiguously.

Tables I and II show that values of the IPD-ILD trading parameter b for the lateral-position models were often near 1 or 0, so that only the IPD or only the ILD contributes to detection. That result means that in the optimizing process the lateral-position models become unstable and become equivalent to independent-binaural-centers models. These problems with lateral-position models favor the independent processing model by default. Also, there is some possible support for models that use IPD and ILD independently in the observation that multiple images can be tracked over short durations (Hafter and Jeffress, 1968; Ruotolo *et al.*, 1979).

In a second comparison, the independent-centers models, models 1–4, also performed better than the STCC models 8–10 and physiology-motivated STCC models tested in Sec. VI. This result continues a pattern: Article I showed that the long-term cross-correlation of the physical signals, as

measured over the full, 500-ms stimulus duration, was an inadequate predictor of incoherence detection. Article II showed that the cross-correlation of the physical signals, as measured over short durations of 25–100 ms, was also inadequate. The present article shows that cross-correlation of transformed signals is also relatively unsuccessful compared to fluctuation detection models.

In the end, the best of the best-fitting models was of the following form: The binaural system detects incoherence on the basis of fluctuations in independently processed IPD and ILD channels, as though IPD and ILD were encoded at different centers without regard for the relative timing of their fluctuations. Nevertheless, there is a residual IPD-ILD interaction in that an IPD fluctuation has no effect if the envelope in the left or right channel becomes smaller than about 20% of the rms envelope value. The IPD and ILD fluctuations are temporally averaged by an exponential window with a time constant less than 5 ms. The time-averaged fluctuations are then laterality compressed. The laterality compression factors found in steady-state experiments on lateral position turn out to be adequate to describe the laterality compression of fluctuations. The processing centers register laterality-compressed fluctuations in IPD and ILD as measured by standard deviations over time. The registered fluctuations are added on a laterality scale at a more central site to form a decision statistic used to detect incoherence. For narrow bands near 500 Hz, time-averaged and laterality-compressed IPD and ILD fluctuations are added with approximately equal weight. As the bandwidth grows, different noises with a given interaural coherence have increasingly similar fluctuations, and the detection of incoherence can be predicted with increasing reliability by the value of coherence itself.

The above paragraph specifying the best binaural model is based on a literal interpretation of the correlation coefficient values, although these values often varied little with model parameters. We also doubt that it is really possible to say that combining independent IPD and ILD fluctuations (model 1) is appreciably more successful than combining independent IPD and ILD mean square fluctuations (model 2). (See Fig. 4.) Further, it is always possible that other models, not tested, might do better. Also, model 1 cannot be proved to be the best for a bandwidth as large as, or larger than, a critical band.

The conclusion that incoherence is detected on the basis of IPD and ILD fluctuations contradicts the conclusions of Breebaart *et al.* (1999) and Breebaart and Kohlrausch (2001) concerning the similar problem of NoSπ detection. In those articles it was pointed out that the distributions of the IPD and the ILD do not vary with noise bandwidth. By contrast, NoSπ detection shows considerable bandwidth dependence. Thus, it was argued, detection is unlikely to be mediated by IPD or ILD fluctuations. Our distribution calculations agree numerically in the sense that the mean standard deviations of IPD and ILD, as shown in Figs. 1 and 7, hardly change when the bandwidth is increased by a factor of 8. However, the variance of the fluctuations among different noise samples, as shown in those figures, depends greatly on bandwidth. We think it possible that the large variance in fluctuations across different samples of noise having small bandwidth, particularly the occurrence of especially large fluctuations, is responsible for the observed bandwidth dependence of detection.

This conjecture, based on the width of the *ensemble* distribution, appears to answer the objection to fluctuation models from the work of Breebaart and his colleagues. It may also help to explain the large individual differences observed in NoSπ detection for narrow bands (Bernstein *et al.*, 1998; Buss *et al.*, 2007) because it suggests that good detection performance requires recognizing the signal in atypical epochs of the stimuli.

B. Binaural processing

Like the experiments of Articles I and II, the experiments presented here conclude that long-term coherence inadequately predicts incoherence detection when the bandwidth is narrow. Long-term coherence may be adequate in the wideband limit. Three results from Article I and the present article indicate features of the wideband limit. These three trends with increasing bandwidth are: (1) the variance among different noises of the fluctuations of IPD and the fluctuations of ILD decreases, (2) the ability of listeners to detect incoherence varies less among different noise samples, and (3) different models of incoherence detection make predictions that are increasing similar, consistent with the prediction of Domnitz and Colburn (1976).

An important difference between the experiments with 14-Hz bandwidth and experiments with 108-Hz bandwidth is the speed of the fluctuations. On the basis of our experiments and modeling, we would agree with Zurek and Durlach (1987) about the advantage of slow fluctuations, but we would not agree that binaural sluggishness plays a role. Instead, our calculations suggest that the binaural system responds rapidly, with a time constant of the order of milliseconds. The best fitting model found an insensitivity to τ in the region of 4 ms for the 14-Hz bandwidth. This time constant is not inconsistent with our matched-noises experiment in Article I wherein the slow fluctuations at 14-Hz bandwidth proved advantageous compared to the fluctuations at 108 Hz. It is commensurate with modulation transfer functions seen in such monaural tasks as the detection of amplitude modulation of broadband noise (Viemeister, 1979). Recently, Stellmack *et al.* (2005) measured temporal modulation transfer functions with time constants of 1 ms for monaural and 1.3 ms for interaural modulation.

By contrast, binaural sluggishness is associated with time constants of tens, or even hundreds, of milliseconds. As suggested in the last paragraph of Hall *et al.* (1998), binaural sluggishness seems to arise in situations where both the masker and the signal plus masker contain dynamical interaural cues. If the masker is interaurally stable the binaural system can take advantage of events in brief epochs. The detection of a small amount of incoherence as a contrast to a diotic noise, as in our experiments, is well modeled as a stable masker (No) and a noise-like signal with a different phase relationship. A rapid response for such a task is consistent with other experiments cited by Hall *et al.*

The best integration time less than 5 ms can be compared with the integration time of 300 ms found to be best in

the loudness meter model of localization as calculated by Hartmann and Constan (2002). Thus, it seems that the binaural auditory system is capable of employing either short or long integration times depending on which better suits the task. When the task is to lateralize an image based on a binaural cue (loudness meter) the integration time is long. When the task is to detect rapid fluctuations in binaural cues, as in the present article, the time is short. A similar point of view was taken with respect to monaural listening by Eddins and Green (1995) wherein integration times of several hundreds of milliseconds are possible for detecting the presence of a signal but times as short as several milliseconds are possible for detecting rapid signal variations.

VIII. CONCLUSION

Experiments of this article studied the detection of small amounts of interaural incoherence in noise bands near 500 Hz. The goal of the experiments was to test different binaural detection models: independent-IPD/ILD, lateral-position, and short-term cross-correlation. Several different transformations were included in the analysis: temporal averaging, laterality compression, and critical envelope weighting. The parameters of these transformations were systematically optimized in the tests of the models. The strongest test came from experiments with a 14-Hz bandwidth. There it was found that the best model independently added the standard deviations of transformed IPD and ILD. This model outperformed a variety of plausible short-term cross-correlation models, even when the cross-correlation models incorporated auditory filtering, cochlear compression, and half-wave rectification.

The nature of incoherence detection depends on the stimulus bandwidth. In the limit of extreme narrow bands the interaural parameters vary extremely slowly, and one imagines that listeners can track the lateral positions indicated by the interaural differences, or by their combination, to detect incoherence. At the other extreme, where the bandwidth is considerably larger than a critical band, the incoherence detection data do not distinguish between different models. In the intermediate range, near 10 Hz, home to the most dramatic MLD results, our experiments indicate that the binaural system is not only sensitive to a running average of the interaural differences, as reflected in the perception of lateral position, but also makes use of the separate fluctuations of the IPD and ILD to detect interaural incoherence.

ACKNOWLEDGMENTS

We are grateful to Dr. H. S. Colburn, Dr. N. I. Durlach, Dr. A. Kohlrausch, Dr. S. van de Par, and Dr. C. Trahiotis for useful discussions about coherence. This work was supported in part by the National Institute on Deafness and Other Communicative Disorders, Grant No. DC00181.

[1]This idea was first suggested by Dr. H. S. Colburn in 2004.

Bernstein, L. R., and Trahiotis, C. (**1992**). "Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies," J. Acoust. Soc. Am. **91**, 306–316.

Bernstein, L. R., and Trahiotis, C. (**1996**). "The normalized correlation: Accounting for binaural detection across center frequency," J. Acoust. Soc. Am. **100**, 3774–3784.

Bernstein, L. R., Trahiotis, C., and Hyde, E. L. (**1998**). "Inter-individual differences in binaural detection of low-frequency tonal signals masked by narrowband or broadband noise," J. Acoust. Soc. Am. **103**, 2069–2078.

Bernstein, L. R., van de Par, S., and Trahiotis, C. (**1999**). "The normalized interaural correlation: Accounting for NoSπ thresholds obtained with Gaussian and low-noise masking noise," J. Acoust. Soc. Am. **106**, 870–876.

Breebaart, J., and Kohlrausch, A. (**2001**). "The influence of interaural stimulus uncertainty on binaural signal detection," J. Acoust. Soc. Am. **109**, 331–345.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**1999**). "The contribution of static and dynamically varying ITDs and IIDs to binaural detection," J. Acoust. Soc. Am. **106**, 979–992.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001a**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001b**). "Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters," J. Acoust. Soc. Am. **110**, 1089–1104.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001c**). "Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters," J. Acoust. Soc. Am. **110**, 1105–1117.

Buss, E., Hall, J. W., and Grose, J. H. (**2007**). "Individual differences in the masking level difference with a narrowband masker at 500 or 2000 Hz." J. Acoust. Soc. Am. **121**, 411–419.

Colburn, H. S., Isabelle, S. K., and Tollin, D. J. (**1997**). "Modelling binaural detection performance for individual masker waveforms." in *Binaural and Spatial Hearing*, edited by R. H. Gilkey and T. Anderson (Erlbaum, Englewood Cliffs, NJ).

Domnitz, R. H., and Colburn, H. S. (**1976**). "Analysis of binaural detection models for dependence on interaural target parameters," J. Acoust. Soc. Am. **59**, 598–601.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Durlach, N. I., Colburn, H. S., and Trahiotis, C. (**1986**). "Interaural correlation discrimination. II. Relation to binaural unmasking," J. Acoust. Soc. Am. **79**, 1548–1556.

Eddins, D. A., and Barber, L. E. (**1998**). "The influence of stimulus envelope and fine structure on the binaural masking level difference," J. Acoust. Soc. Am. **103**, 2578–2589.

Eddins, D. A., and Green, D. M. (**1995**). "Temporal integration and temporal resolution," *Handbook of Perception and Cognition—Hearing*, 2nd ed., edited by B. C. J. Moore (Academic, San Diego), pp. 207–242.

Egan, J. P., Schulman, A. I., and Greenberg, G. Z. (**1959**). "Operating characteristics determined by binary decision and by ratings," J. Acoust. Soc. Am. **31**, 768–773.

Evilsizer, M. E., Gilkey, R. H., Mason, C. R., Colburn, H. S., and Carney, L. H. (**2002**). "Binaural detection with narrowband and wideband reproducible noise maskers: I. Results for human," J. Acoust. Soc. Am. **111**, 336–345.

Gabriel, K. J., and Colburn, H. S. (**1981**). "Interaural correlation discrimination. I. Bandwidth and level dependence," J. Acoust. Soc. Am. **69**, 1394–1401.

Gilkey, R. H., Robinson, D. E., and Hanna, T. E. (**1985**). "Effects of masker waveform and signal-to-masker phase relation on diotic and dichotic masking by reproducible noise," J. Acoust. Soc. Am. **78**, 1207–1219.

Goupell, M. J., and Hartmann, W. M. (**2006**). "Interaural fluctuations and the detection of interaural incoherence: Bandwidth effects," J. Acoust. Soc. Am. **119**, 3971–3986.

Goupell, M. J., and Hartmann, W. M. (**2007**). "Interaural fluctuations and the detection of interaural incoherent. II. Brief duration noises," J. Acoust. Soc. Am. **121**, 2127–2136.

Grantham, D. W., and Wightman, F. L. (**1978**). "Detectability of varying interaural temporal differences," J. Acoust. Soc. Am. **63**, 511–523.

Grantham, D. W., and Wightman, F. L. (**1979**). "Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation," J. Acoust. Soc. Am. **65**, 1509–1517.

Hafter, E. R. (**1971**). "Quantitative evaluation of a lateralization model of masking-level differences," J. Acoust. Soc. Am. **50**, 1116–1122.

Hafter, E. R., and Jeffress, L. A. (**1968**). "Two-image lateralization of tones and clicks," J. Acoust. Soc. Am. **44**, 563–569.

Hall, J. W., Grose, J. H., and Hartmann, W. M. (**1998**). "The masking level

difference in low-noise noise," J. Acoust. Soc. Am. **103**, 2573–2577.
Hartmann, W. M., and Constan, Z. A. (**2002**). "Interaural level differences and the level meter model," J. Acoust. Soc. Am. **112**, 1037–1045.
Isabelle, S. K., and Colburn, H. S. (**1987**). "Effects of target phase in narrowband frozen noise detection data," J. Acoust. Soc. Am. **82**, S109.
Isabelle, S. K., and Colburn, H. S. (**1991**). "Detection of tones in reproducible narrowband noise," J. Acoust. Soc. Am. **89**, 352–359.
Isabelle, S. K., and Colburn, H. S. (**2004**). "Binaural detection of tones masked by reproducible noise: Experiment and models," Report BU-HRC 04–01.
Jeffress, J. A., Blodgett, H. C., Sandel, T. T., and Wood, C. L. (**1956**). "Masking of tonal signals," J. Acoust. Soc. Am. **28**, 416–426.
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.
Osman, E. (**1971**). "A correlation model of binaural masking level differences," J. Acoust. Soc. Am. **50**, 1494–1511.
Oxenham, A. J., and Moore, B. C. J. (**1995**). "Additivity of masking in normally hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **98**, 1921–1934.
Ruotolo, B. R., Stern, R. M., and Colburn, H. S. (**1979**). "Discrimination of symmetric time-intensity traded binaural stimuli," J. Acoust. Soc. Am. **66**, 1733–1737.
Schulman, A. I., and Mitchell, R. R. (**1966**). "Operating characteristics from yes-no and forced-choice procedures." J. Acoust. Soc. Am. **40**, 473–477.
Stellmack, M. A., Viemeister, N. F., and Byrne, A. J. (**2005**). "Monaural and interaural temporal modulation transfer functions measured with 5-kHz carriers," J. Acoust. Soc. Am. **118**, 2507–2518.
van de Par, S., and Kohlrausch, A. (**1998**). "Diotic and dichotic detection using multiplied noise maskers," J. Acoust. Soc. Am. **103**, 2100–2110.
Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.
Webster, F. A. (**1951**). "The influence of interaural phase on masked thresholds. I. The role of interaural time-deviation," J. Acoust. Soc. Am. **23**, 452–462.
Yost, W. A. (**1981**). "Lateral position of sinusoids presented with interaural intensive and temporal differences," J. Acoust. Soc. Am. **70**, 397–409.
Yost, W. A., and Hafter, E. R. (**1987**). "Lateralization" in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer, New York), pp. 49–84.
Zurek, P. M., and Durlach, N. I. (**1987**). "Masker-bandwidth dependence in homophasic and antiphasic tone detection," J. Acoust. Soc. Am. **81**, 459–464.
Zwicker, E., and Terhardt, E. (**1980**). "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency," J. Acoust. Soc. Am. **68**, 1523–1525.

Frequency modulation detection with simultaneous amplitude modulation by cochlear implant users

Xin Luo[a] and Qian-Jie Fu
Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 16 August 2006; revised 29 May 2007)

To better represent fine structure cues in cochlear implants (CIs), recent research has proposed varying the stimulation rate based on slowly varying frequency modulation (FM) information. The present study investigated the abilities of CI users to detect FM with simultaneous amplitude modulation (AM). FM detection thresholds (FMDTs) for 10-Hz sinusoidal FM and upward frequency sweeps were measured as a function of standard frequency (75–1000 Hz). Three AM conditions were tested, including (1) No AM, (2) 20-Hz Sinusoidal AM (SAM) with modulation depths of 10%, 20%, or 30%, and (3) Noise AM (NAM), in which the amplitude was randomly and uniformly varied over a range of 1, 2, or 3 dB, relative to the reference amplitude. Results showed that FMDTs worsened with increasing standard frequencies, and were lower for sinusoidal FM than for upward frequency sweeps. Simultaneous AM significantly interfered with FM detection; FMDTs were significantly poorer with simultaneous NAM than with SAM. Besides, sinusoidal FMDTs significantly worsened when the starting phase of simultaneous SAM was randomized. These results suggest that FM and AM in CI partly share a common loudness-based coding mechanism and the feasibility of "FM+AM" strategies for CI speech processing may be limited. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2751258]

PACS number(s): 43.66.Fe, 43.66.Hg, 43.66.Ts [AJO] Pages: 1046–1054

I. INTRODUCTION

Given the success of multichannel cochlear implants (CIs) for speech recognition in quiet, CI research has increasingly focused on improving patient performance for more challenging listening tasks, such as speech understanding in noise (e.g., Friesen *et al.*, 2001; Fu and Nogaki, 2005; Fu *et al.*, 1998; Stickney *et al.*, 2004), music appreciation (e.g., Kong *et al.* 2004, 2005; McDermott, 2004), speaker and voice gender identification (e.g., Fu *et al.* 2004, 2005; Vongphoe and Zeng, 2005), and vocal emotion recognition (Luo *et al.*, 2006). The generally poor CI performance in these tasks is thought to be due to the limited spectrotemporal information provided by the CI device (Wilson *et al.*, 2005). Much research has been directed at restoring fine structure cues in CIs, including high stimulation rates, which improve temporal coding and mimic the spontaneous neural firing activity observed in the normal cochlea (e.g., Rubinstein *et al.*, 1999), and "current steering" (e.g., Townshend *et al.*, 1987; Wilson *et al.*, 1994; Donaldson *et al.*, 2005), in which current is simultaneously delivered to adjacent electrodes to create "virtual channels," thereby increasing the number of spectral channels provided by the CI device.

Most present-day CI speech processors only transmit amplitude modulation (AM) information (e.g., AM rates and depths) by modulating the pulse train amplitudes according to the temporal envelope in each spectral channel. However, fine structure information is mostly found in the frequency modulation (FM) spectrum (e.g., the transitions of pitch and formant frequencies). Both the slowly varying temporal envelope and the fast varying fine structure of an acoustic signal can be derived from the Hilbert transform (e.g., Smith *et al.*, 2002). Zeng *et al.* (2005) proposed to vary the stimulation rate to encode the FM information of input acoustic signals, as a complement to the AM information encoded by varying the stimulation level (i.e., frequency and amplitude modulation encoding, or FAME). While strategies such as FAME may transmit additional fine structure cues, their success depends on CI users' ability to perceive FM information encoded by the time-varying stimulation rate, especially in the presence of AM information encoded by the temporal envelope.

In electric hearing, pitch perception produced by varying the stimulation rate (temporal rate pitch) is independent of that produced by varying the location of the stimulated electrode (place pitch) (e.g., Tong *et al.*, 1982; McKay *et al.*, 2000). Psychophysical studies using single-electrode stimulation have shown that CI users' temporal sensitivity declines sharply above 300-Hz stimulation rate (e.g., Shannon, 1983; Zeng, 2002), although some CI users exhibit good temporal resolution up to 1000 Hz (Wilson *et al.*, 2000). Extremely high stimulation rates (e.g., up to 12 kHz) may produce percepts in some CI users that may not be directly related to pitch (Landsberger and McKay, 2005). Chen and Zeng (2004) recently found that, although FM detection thresholds (FMDTs) in electric hearing generally worsened with increasing standard frequencies, some CI users were able to access FM information for standard frequencies as high as 1000 Hz.

In normal hearing (NH), two mechanisms are thought to encode FM information (e.g., Moore and Sek, 1996). At low

[a]Electronic mail: xluo@hei.org

 0001-4966/2007/122(2)/1046/9/$23.00

TABLE I. Relevant information for cochlear implant subjects who participated in the present experiments.

Subject	Age	Gender	Etiology	Prosthesis	Strategy	Years with prosthesis
S1	63	F	Genetic	N-24	ACE	3
S2	75	M	Noise induced	N-22	SPEAK	9
S3	64	M	Trauma/unknown	N-22	SPEAK	15
S4	49	M	Trauma	N-22	SPEAK	13
S5	67	M	Hereditary	N-22	SPEAK	14

carrier frequencies (below about 4 kHz) and low FM rates (below about 10 Hz), FM is encoded by a "temporal" mechanism, based on the phase locking between the auditory nerve and the stimulus frequency (Siebert, 1970; Goldstein and Srulovicz, 1977); temporal firing patterns in the auditory nerve contain instantaneous frequency information. At higher carrier frequencies or FM rates, FM is encoded by a "place" mechanism, in which changes in the excitation pattern dominate FM coding (e.g., Zwicker, 1956; Moore and Sek, 1994). In the place mechanism, changes in stimulus frequency are transformed into amplitude fluctuations of the peripheral auditory filter outputs; in this scenario, FM is encoded similarly to AM. Several experiments with NH listeners suggest that, indeed, FM and AM are not encoded by completely independent mechanisms. For example, Demany and Semal (1986) were the first to show that AM is best detected at high modulation rates but that FM is best detected at low modulation rates. They also showed that the threshold for detecting modulation and identifying whether it was FM or AM was higher than that for just detecting the modulation, but only at relatively high modulation rates. Similarly, Edwards and Viemeister (1994) found that for 1-kHz carrier frequency and modulation rates below 64 Hz, modulation was similarly encoded for AM and FM stimuli at low modulation depths; only when modulation was highly detectable (i.e., at large modulation depths) could FM stimuli be discriminated from AM stimuli set at equally detectable levels. Other studies have shown that FMDTs worsen in the presence of simultaneous AM (e.g., Moore and Sek, 1996). Simultaneous AM has also been shown to be more disruptive to FM detection in hearing-impaired (HI) listeners than in NH listeners (e.g., Grant, 1987; Moore and Skrodzka, 2002), presumably because HI listeners largely rely on excitation pattern cues to detect FM, while the cochlear filtering associated with hearing loss is much broader. It is also worth noting that in modulation detection interference, the detection of AM at a target carrier frequency was adversely affected by simultaneous FM at remote masker carrier frequencies and vice versa, which partly resulted from the perceptual auditory grouping phenomenon (Moore *et al.*, 1991).

When FM is encoded by changes in stimulation rate on a single electrode, CI users must rely on purely temporal cues for FM detection (e.g., phase locking between the auditory nerve firing pattern and the stimulus FM rate). Unlike previous NH studies (in which the temporal and place code cannot be easily separated), FM detection in CI users allows for independent study of the temporal coding mechanism. McKay and Carlyon (1999) found that when listening to amplitude-modulated current pulse trains on a single electrode, CI users achieved dual temporal pitch percepts, one of which corresponded to the AM rate while the other corresponded to the carrier stimulation rate; as the AM modulation depth was increased, the perceived pitch gradually switched from the carrier stimulation rate to the AM rate (McKay *et al.*, 1995). It is unclear how simultaneous AM may affect FM detection in electric hearing, given that the coding mechanisms for AM and FM stimuli in electric hearing are both temporal. FM detection with simultaneous AM in CI users may provide additional insight into modulation coding mechanisms in NH, HI, and CI users, and whether there are modulation parameters that are held in common between AM and FM detection. As a practical matter, it is important to know the limits of FM detection in the presence of AM when designing "FM+AM" CI speech processing strategies (e.g., Zeng *et al.*, 2005).

The present study measured FM sensitivity in the presence of simultaneous AM in five adult CI users, as a function of standard frequency. In the first experiment, FMDTs were measured for sinusoidal FM (modulation rate fixed at 10 Hz) and for upward frequency sweeps under three AM conditions: (1) No AM, (2) 20-Hz Sinusoidal AM (SAM), and (3) Noise AM (NAM). 10-Hz sinusoidal FM and 20-Hz sinusoidal AM were used to avoid aliasing at the lowest standard frequency (75 Hz). In the second experiment, FMDTs for 10-Hz sinusoidal FM were also measured when the starting phase of simultaneous 20-Hz SAM was randomized across stimulation intervals. Compared to the in-phase SAM (as in the first experiment), the random-phase SAM produced a similar temporal pitch, but random loudness fluctuations. The effect of different starting phases between simultaneous SAM on sinusoidal FM detection may indicate whether AM interferes with FM only in the temporal pitch domain, or in the loudness domain as well.

II. EXPERIMENT 1

A. Methods

1. Subjects

Five postlingually deafened CI subjects, including four Nucleus-22 users (male) and one Nucleus-24 user (female), participated in the present experiments. All subjects had many years' experience with their devices, as well as extensive experience with speech and psychophysical experiments. Table I shows the relevant demographic details for the participating subjects. Informed consent was obtained from all subjects, all of whom were paid for their participation.

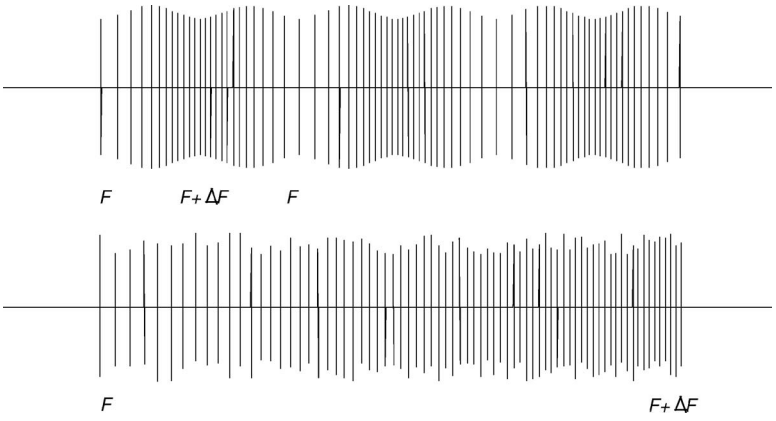


FIG. 1. Two examples of experimental stimuli. The upper panel shows a stimulus with sinusoidal FM (10 Hz) and simultaneous sinusoidal AM (20 Hz; 30% modulation depth), while the lower panel shows a stimulus with upward frequency sweep and simultaneous noise AM (−3-dB noise level).

2. Stimuli

All stimuli were presented in BP+1 stimulation mode and delivered to a middle electrode pair (10,12) via custom research interface (HEINRI: Wygonski and Robert, 2001). The reference stimulation level for both modulated and unmodulated stimuli corresponded to 50% of the estimated electrode dynamic range (i.e., a comfortable listening level). Only one electrode location and one reference stimulation level were tested, as CI users' FMDTs have been shown to be independent of electrode location and stimulation level (Chen and Zeng, 2004). FMDTs were obtained for five standard frequencies (75, 125, 250, 500, and 1000 Hz). All stimuli were 300-ms, biphasic pulsatile trains. For each pulse, the phase duration was 200 μs for the 75-, 125-, and 250-Hz frequency conditions, and 100 μs for the 500- and 1000-Hz frequency conditions; the interphase gap was fixed at 45 μs.

FMDTs were measured for both sinusoidal FM and upward frequency sweeps. For the sinusoidal FM conditions, the FM modulation rate was fixed at 10 Hz, i.e., low enough to avoid aliasing effects even at the lowest standard frequency (75 Hz). There were three complete FM periods for the 300-ms stimuli with the 10-Hz sinusoidal FM. In each period of the sinusoidal FM, the instantaneous stimulation frequency began at the standard frequency F_{STD}, increased to $F_{\mathrm{STD}}+\Delta F$ in the first half period, and decreased back to F_{STD} in the second half period. The minimal detectable peak-to-valley frequency difference ΔF was defined as the detection threshold for the sinusoidal FM. For the upward frequency sweeps, the instantaneous stimulation frequency began at the standard frequency F_{STD}, and increased as a linear function of time to $F_{\mathrm{STD}}+\Delta F$ at the end of the stimulation interval; the minimal detectable frequency change ΔF was defined as the detection threshold for the upward frequency sweeps. Downward frequency sweeps were not tested, as the direction of the frequency sweeps has been shown to have no significant effect on CI users' FMDTs (Chen and Zeng, 2004).

FM detection was also measured in the presence of simultaneous sinusoidal AM (SAM) and noise AM (NAM). For simultaneous SAM, the reference stimulation level was modulated by a 20-Hz sine wave with relative modulation depths of 10%, 20%, or 30%. The starting phase of the sinusoidal AM was the same as for the sinusoidal FM; note that the modulation rate for SAM (20 Hz) was twice that of the sinusoidal FM (10 Hz). The different AM and FM rates were not selected to maximize the interference between AM and FM, but rather to investigate whether simultaneous SAM (even with different rates from FM) would affect FMDTs. For simultaneous NAM, the instantaneous stimulation level for each pulse was varied by a random value that was uniformly distributed between 0 (i.e., no change in amplitude) and n dB (where $n=-1$, -2, or -3); note that the random fluctuations of instantaneous amplitude were different across NAM stimulation intervals. Because the loudness of AM stimuli can be different from that of unmodulated stimuli, all AM stimuli were loudness-balanced to steady state pulse trains at the reference stimulation level, for each standard carrier frequency (see Sec. II A 3). Figure 1 shows two examples of the experimental stimuli; the upper panel shows a stimulus with sinusoidal FM (10 Hz) and simultaneous SAM (20 Hz; 30% modulation depth), while the lower panel shows a stimulus with upward frequency sweep and simultaneous NAM (−3-dB noise level).

In summary, there were five standard frequencies (75, 125, 250, 500, and 1000 Hz) $\times 2$ FM types (10-Hz sinusoidal FM and upward frequency sweep) $\times 7$ detailed AM conditions [no AM, 20-Hz SAM (10%, 20%, and 30% modulation depth), and NAM (−1, −2, and −3-dB noise level)], resulting in a total of 70 experimental conditions.

3. Procedures

All experiments were conducted via custom software and custom Nucleus research interface (Shannon *et al.*, 1990; Wygonski and Robert, 2001); stimuli were directly presented to subjects' implant devices via a custom research interface, thereby bypassing subjects' clinically assigned speech processors. Before beginning the FM detection experiments, electrode dynamic ranges (DRs) were estimated for all standard frequencies, using unmodulated stimuli. A counting method was used to estimate absolute detection thresholds, similar to clinical fitting procedures. Beginning at a subthreshold level, the current amplitude was increased until subjects were able to correctly count the number of stimuli; the amplitude was then reduced until subjects could no longer correctly count the number of stimuli. These ascending and descending sequences were repeated several times to obtain absolute detection thresholds. A method of limits was

used to measure maximal comfortable levels (MCLs), defined as the maximum stimulation level that subjects could comfortably listen to for an extended period of time (e.g., during an experiment). Subjects pressed a mouse button to slowly increase the current amplitude until achieving MCL; MCLs were measured several times to ensure reliable levels. The estimated DR for each standard frequency was calculated as the difference in current level (in linear μA) between threshold and MCL. For each standard frequency condition in the FM detection experiments, the stimulation level was set to 50% of DR (calculated in linear μA).

Before beginning the FM detection experiments, AM stimuli were loudness-balanced to steady-state, unmodulated stimuli (i.e., no AM or FM, presented at 50% of DR) for each standard frequency, using a two-alternative forced-choice (2AFC), double-staircase procedure (Jesteadt, 1980; Zeng and Turner, 1991); depending on the sequence, the adaptation rule was 2-down/1-up or 2-up/1-down. The standard stimulus was an unmodulated, steady-state pulse train, and the reference amplitude of the AM stimulus was adjusted according to subject response, in 0.8-dB steps for the first 4 reversals and in 0.4-dB steps thereafter. The sequences terminated after 12 reversals or 60 trials with at least 8 reversals. The amplitudes of the final 8 reversals were averaged for each sequence; the mean values from both sequences were then averaged to obtain the loudness-balanced amplitudes for AM stimuli. Once loudness-balanced amplitudes for all AM stimuli (at all standard frequencies) were obtained, these levels were used for the subsequent FM detection experiments. Note that the above-mentioned loudness-balance procedure controlled only for the overall loudness of AM stimuli, and not for loudness fluctuations within AM stimuli.

For all conditions, FMDTs were measured using an adaptive 3AFC procedure (3-down/1-up), converging on the frequency difference in FM that produced 79.4% correct (Levitt, 1971). In each trial, two stimulation intervals (randomly selected) contained stimuli with no FM and one interval contained a stimulus with FM. Subjects were asked to choose which interval was different, and the parameter ΔF was adjusted according to subject response. The starting value of ΔF was set to 80%–100% of the standard frequency, such that subject responses in the first 3 trials were always correct. The initial step size was $\sim$25% of the standard frequency, and was reduced to $\sim$10% of the standard frequency after the first 4 reversals. The adaptive run terminated after 12 reversals or after 60 trials with a minimum of 8 reversals. FMDTs were calculated as the average frequency difference across the final 8 reversals. During each run, the simultaneous AM type, AM depth, and standard frequency were held constant. For each condition, a minimum of six runs was completed by each subject. The test order of experimental conditions was randomized across subjects.

B. Results

Figure 2 shows mean FMDTs (averaged across subjects) in terms of relative ΔF to the standard frequency (i.e., $\Delta F/F$) for sinusoidal FM (modulation rate fixed at 10 Hz; upper two panels) and for upward frequency sweeps (lower two panels), obtained with different amounts of simultaneous SAM (left two panels) or NAM (right two panels), as a function of standard frequency.

For 10-Hz sinusoidal FM without simultaneous AM, mean FMDTs (averaged across subjects) increased from 9 Hz at the 75-Hz standard frequency to 276 Hz at the 1000-Hz standard frequency (the corresponding $\Delta F/F$ ranged from 0.12 to 0.28). FMDTs for 10-Hz sinusoidal FM generally worsened with increasing amounts of simultaneous SAM. For example, the mean FMDT with simultaneous SAM of 30% modulation depth (averaged across subjects and standard frequencies) was 1.33 times higher than that with no AM. With simultaneous NAM, FMDTs were generally higher than those with simultaneous SAM. For example, the mean FMDT for the −3-dB NAM condition (averaged across subjects and standard frequencies) was 1.76 times higher than that with no AM. As the depth of simultaneous NAM was increased, FMDTs worsened. Similar patterns of results were observed for upward frequency sweeps. Note that FMDTs were generally higher for upward frequency sweeps than for 10-Hz sinusoidal FM, for most experimental conditions. For example, for conditions without AM, the mean FMDT for upward frequency sweeps (averaged across subjects and standard frequencies) was 1.35 times higher than that for 10-Hz sinusoidal FM.

Two-way repeated measures analyses of variance [(ANOVAs); with standard frequency and amount of simultaneous AM as factors] were performed on the $\Delta F/F$ data for each combination of AM and FM type (i.e., the data shown in each panel of Fig. 2). For 10-Hz sinusoidal FM with simultaneous SAM (upper left panel), $\Delta F/F$ was significantly affected by the standard frequency [$F(4,48)=11.5, p<0.001$] and by the amount of SAM [$F(3,48)=7.2, p=0.005$]. There was a significant interaction between the standard frequency and amount of SAM [$F(12,48)=2.0, p=0.046$, power of analysis: 0.49]. *Post hoc* Bonferroni *t*-tests showed that $\Delta F/F$ was significantly higher for standard frequencies of 500 and 1000 Hz, relative to 75 and 125 Hz ($p<0.05$); there was no significant difference in $\Delta F/F$ between other standard frequencies ($p>0.05$). *Post hoc* analyses also showed that $\Delta F/F$ significantly worsened (relative to that without AM) only when the modulation depth of simultaneous SAM was 20% or 30% ($p<0.05$); further analyses revealed that simultaneous SAM significantly affected $\Delta F/F$ only for the 75- and 500-Hz standard frequencies.

For 10-Hz sinusoidal FM with simultaneous NAM (upper right panel), $\Delta F/F$ was significantly affected by the standard frequency [$F(4,48)=18.5, p<0.001$] and by the amount of NAM [$F(3,48)=31.7, p<0.001$]. There was a significant interaction between the standard frequency and amount of NAM [$F(12,48)=4.9, p<0.001$]. *Post hoc* Bonferroni *t*-tests showed that $\Delta F/F$ was significantly different between most standard frequencies, except for 75 vs 125 Hz, 125 vs 1000 Hz, 250 vs 500 Hz, and 500 vs 1000 Hz ($p>0.05$). *Post hoc* analyses also showed that $\Delta F/F$ significantly worsened (relative to that without AM) for all levels of simultaneous NAM ($p<0.05$).

A similar pattern in the analysis results was found be-

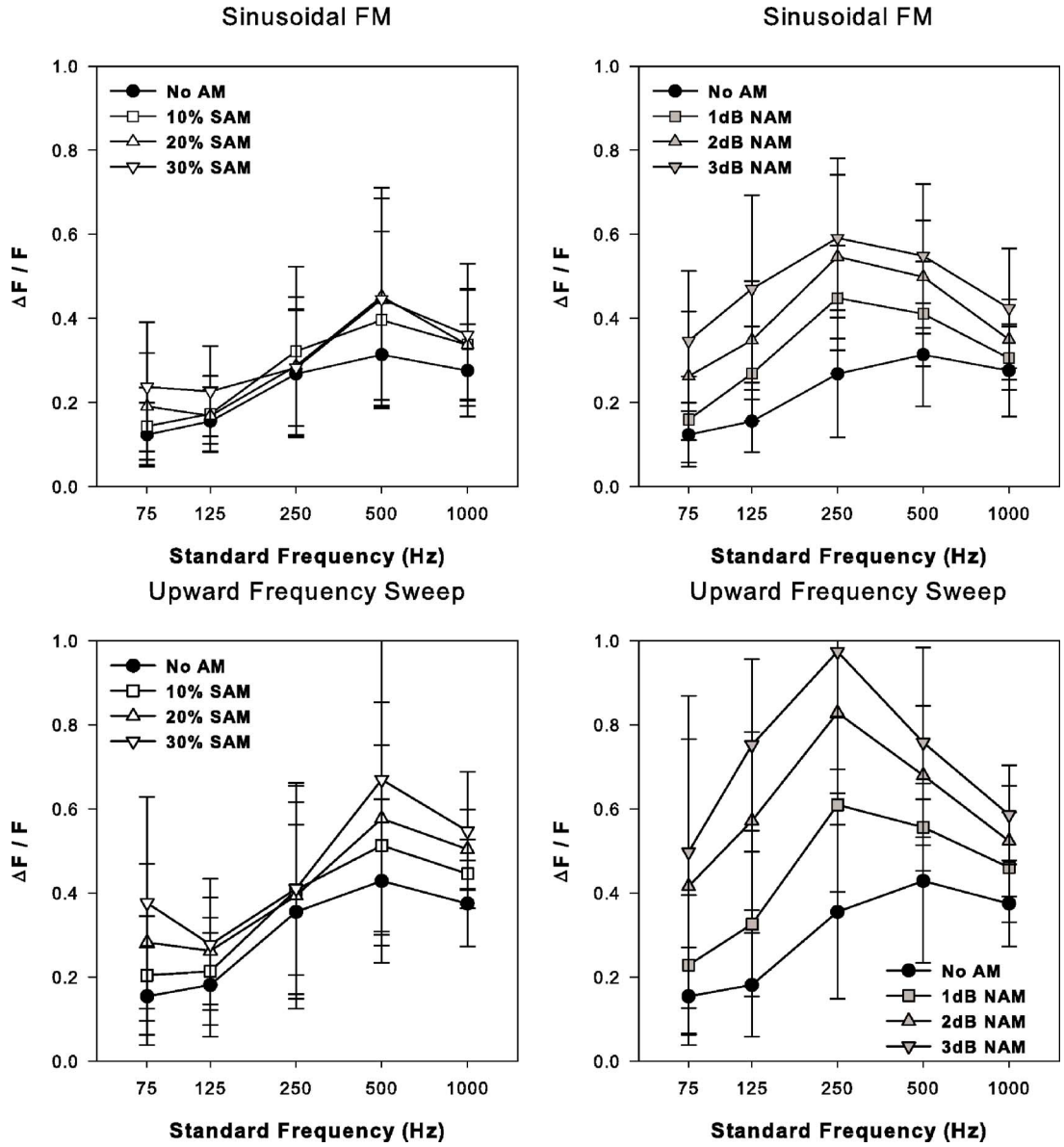


FIG. 2. (Color online) Mean FM detection thresholds (averaged across subjects) relative to the standard frequency (i.e., $\Delta F/F$), as a function of standard frequency. The upper two panels show $\Delta F/F$ for sinusoidal FM (modulation rate fixed at 10 Hz), while the lower two panels show $\Delta F/F$ for upward frequency sweeps. The left two panels show $\Delta F/F$ obtained with different amounts of simultaneous SAM, while the right two panels show $\Delta F/F$ obtained with different amounts of simultaneous NAM. The error bars represent 1 s.d.

tween 10-Hz sinusoidal FM and upward frequency sweeps. For upward frequency sweeps with simultaneous SAM (lower left panel), $\Delta F/F$ was significantly affected by the standard frequency [$F(4,48)=5.9, p=0.004$] and by the amount of SAM [$F(3,48)=23.6, p<0.001$]. There was a significant interaction between the standard frequency and amount of simultaneous SAM [$F(12,48)=2.0, p=0.043$, power of analysis: 0.50]. *Post hoc* Bonferroni *t*-tests showed that $\Delta F/F$ was significantly higher for standard frequency of 500 Hz, relative to 75 and 125 Hz ($p<0.05$); there was no significant difference in $\Delta F/F$ between other standard frequencies ($p>0.05$). *Post hoc* analyses also showed that $\Delta F/F$ significantly worsened (relative to that without AM) only when the modulation depth of simultaneous SAM was 20% or 30% ($p<0.05$); the 30% SAM modulation depth also produced significantly higher $\Delta F/F$ values than those with the 10% SAM modulation depth ($p<0.05$).

Simultaneous NAM produced a similar pattern of results for 10-Hz sinusoidal FM and upward frequency sweeps, although the effect of NAM on $\Delta F/F$ was much more significant for upward frequency sweeps than for 10-Hz sinusoidal FM. For upward frequency sweeps with simultaneous NAM (lower right panel), $\Delta F/F$ was significantly affected by the standard frequency [$F(4,48)=9.7, p<0.001$] and by the amount of NAM [$F(3,48)=53.0, p<0.001$]. There was a significant interaction between the standard frequency and amount of NAM [$F(12,48)=2.7, p=0.008$]. *Post hoc* Bonferroni *t*-tests showed that $\Delta F/F$ was significantly different between only some standard frequencies: 75 vs 250 Hz, 75 vs 500 Hz, and 125 vs 250 Hz ($p<0.05$); there was no significant difference in $\Delta F/F$ between other standard frequencies. *Post hoc* analyses also showed that $\Delta F/F$ significantly worsened (relative to that without AM) for all levels of simultaneous NAM ($p<0.05$); $\Delta F/F$ was significantly

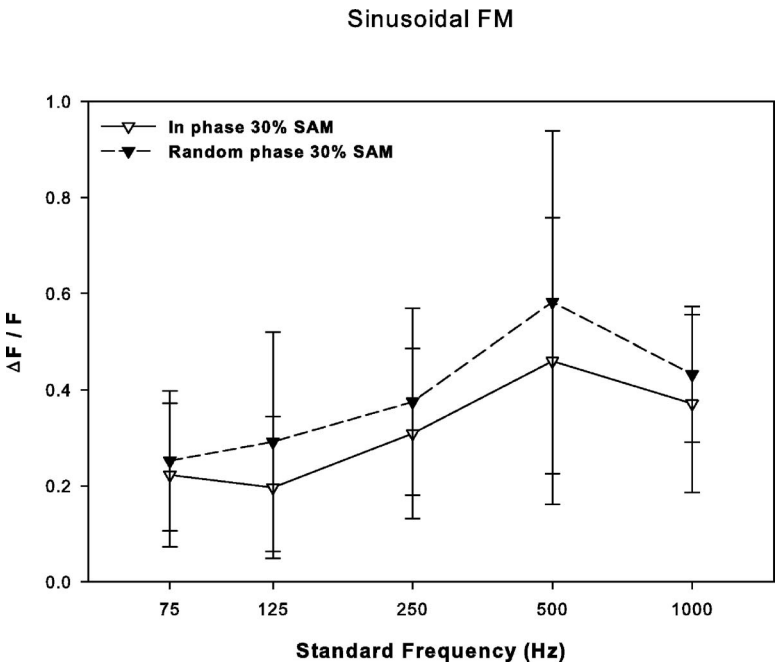


FIG. 3. Mean FM detection thresholds for 10-Hz sinusoidal FM (averaged across subjects) relative to the standard frequency (i.e., $\Delta F/F$), obtained when the simultaneous 20-Hz SAM (30% modulation depth) was in-phase or random-phase relative to the sinusoidal FM, as a function of standard frequency. The error bars represent 1 s.d.

different between all levels of NAM ($p<0.05$), except for between −2 and −3-dB NAM ($p>0.05$).

III. EXPERIMENT 2

A. Methods

The same five CI subjects from Experiment 1 participated in Experiment 2. FMDTs for 10-Hz sinusoidal FM were measured in the presence of 20-Hz simultaneous SAM (30% modulation depth); the starting phase of the SAM was either the same as the sinusoidal FM (in-phase SAM, the same as in Experiment 1) or randomized across stimulation intervals (random-phase SAM). Thus, there was a total of 10 experimental conditions in Experiment 2: 5 standard frequencies (75, 125, 250, 500, and 1000 Hz) ×2 AM conditions [20-Hz 30% SAM (in-phase and random-phase)]. The other relevant stimulus parameters, as well as the testing procedures, were the same as in Experiment 1.

B. Results

Figure 3 shows mean FMDTs (averaged across subjects) for 10-Hz sinusoidal FM (in terms of $\Delta F/F$) obtained when the simultaneous 20-Hz SAM (30% modulation depth) was either in-phase with the sinusoidal FM or random phase, as a function of standard frequency. A two-way repeated measures ANOVA (with standard frequency and starting phase of simultaneous SAM as factors) showed that $\Delta F/F$ was significantly affected by the standard frequency [$F(4,16)=7.4, p=0.001$] and by the starting phase of simultaneous SAM [$F(1,16)=28.5, p=0.006$]. There was no significant interaction between the standard frequency and starting phase of simultaneous SAM [$F(4,16)=1.2, p=0.37$]. *Post hoc* Bonferroni *t*-tests showed that $\Delta F/F$ was significantly higher for standard frequency of 500 Hz, relative to 75 and 125 Hz ($p<0.05$); there was no significant difference in $\Delta F/F$ between other standard frequencies ($p>0.05$). *Post hoc* analyses also showed that $\Delta F/F$ significantly worsened when the starting phase of simultaneous SAM was randomized across stimulation intervals ($p<0.05$); further analyses revealed that the random-phase SAM significantly affected $\Delta F/F$ only for the 125- and 500-Hz standard frequencies.

IV. GENERAL DISCUSSION

The results from the present study show that CI users were able to detect FM at relatively low stimulation rates, consistent with previous findings (Chen and Zeng, 2004); FMDTs gradually increased as the standard frequency was increased. More important, FM detection by CI users significantly worsened in the presence of simultaneous AM. These results suggest that AM and FM in CI at least partly share a common coding mechanism, and that fine structure cues encoded by dynamically changing the stimulation rate may be largely masked by simultaneous temporal envelope modulation.

There are some notable differences in terms of procedures and results between the present study and the Chen and Zeng (2004) study. First, in the Chen and Zeng study, FMDTs were measured with 1 dB of amplitude roving, comparable to the −1-dB NAM condition tested in the present study. Second, FMDTs for sinusoidal FM were calculated differently between the two studies. For sinusoidal FM, Chen and Zeng (2004) calculated the peak-to-average frequency difference to determine the FMDT, while the present study calculated the peak-to-valley difference; thus the frequency difference in the present study was double that of Chen and Zeng (2004). For comparison purposes, FMDTs for sinusoidal FM from Chen and Zeng (2004) were doubled. After this scaling adjustment, data from the −1-dB NAM condition in the present study were compared to those from the Chen and Zeng study. At low standard frequencies, FMDTs were generally lower for the present study than for the Chen and Zeng study. For example, at the 75-Hz standard frequency, the mean FMDT for upward frequency sweeps was 17 Hz in

the present study, and 22 Hz in the Chen and Zeng study; similarly, the mean FMDT for 10-Hz sinusoidal FM was 12 Hz in the present study, and 20 Hz in the Chen and Zeng study. However, at high standard frequencies, the FMDTs in the present study were generally higher than those in Chen and Zeng (2004). For example, at the 1000-Hz standard frequency, the mean FMDT for upward frequency sweeps was 460 Hz in the present study, and 361 Hz in the Chen and Zeng study; similarly, the mean FMDT for 10-Hz sinusoidal FM was 305 Hz in the present study, and 200 Hz in the Chen and Zeng study. Thus, in the present study, as the standard frequency was increased, the FMDTs were more sharply elevated than in Chen and Zeng (2004). These different results may be due to intersubject differences in terms of overall FM sensitivity. Compared to the three CI subjects tested by Chen and Zeng (2004), the five subjects tested in the present study were less sensitive to FM at high standard frequencies. Interestingly, FMDTs from both studies were lower for 10-Hz sinusoidal FM than for upward frequency sweeps. Chen and Zeng (2004) suggested that the relatively faster changes in stimulation rate and the multiple opportunities to listen for peak changes in stimulation rate with sinusoidal FM might have contributed to its lower detection thresholds. It is important to note that both studies showed large intersubject variability in FMDTs.

The most important finding of the present study is that CI users' FM detection was adversely affected by simultaneous AM, similar to results found with NH and HI listeners (e.g., Grant, 1987; Moore and Sek, 1996; Moore and Skrodzka, 2002). Because simultaneous AM interfered with CI users' FM detection, the present results suggest that in electric hearing, FM (as encoded by stimulation rate) may be processed (in part) by the same mechanism as AM (as encoded by stimulation level). In the following, we discuss possible common mechanisms for AM and FM coding in CI.

AM may interfere with FM detection in the temporal pitch domain. Varying the stimulation rate (i.e., FM coding) at relatively low standard frequencies will evoke temporal rate pitch in implant users (e.g., Shannon, 1983; Zeng, 2002), while sinusoidally amplitude-modulated electrical pulse trains (i.e., AM coding) will produce temporal envelope pitch determined by the AM rate (e.g., McKay *et al.*, 1994). With both temporal rate and envelope pitch mixed together, listeners may have difficulty discerning between these two temporal pitch percepts. When the modulation depth of simultaneous SAM was increased, the evoked temporal envelope pitch percept became stronger, and more greatly interfered with FM detection. CI subjects' ability to selectively listen to particular temporal fluctuation rates may have contributed to the lesser effect of the 20-Hz SAM than the NAM on the detection of 10-Hz sinusoidal FM. However, it should be noted that the tested 20-Hz AM rate is below the lower limit of voice pitch, and is very different from the tested stimulation rates, thus the present results may not be used to testify the interference between AM and FM in the temporal pitch domain.

AM may also interfere with FM detection in the loudness domain. For high FM rates, NH listeners may detect FM using loudness cues from AM-like modulations in adjacent peripheral frequency channels (e.g., Moore and Sek, 1994). In the single-electrode FM detection experiments conducted in the present study (and in Chen and Zeng, 2004), multiple frequency channels were not available. It is possible that, even for single-electrode stimulation, subjects may use the loudness fluctuations associated with changes in stimulation rate to detect FM (especially at high standard frequencies), and that these loudness cues may be overwhelmed by the amplitude fluctuations associated with AM. Note that only relatively large amounts of simultaneous SAM interfered with FM detection, and that simultaneous NAM produced greater interference with FM detection. Compared to SAM, NAM produced greater variability in amplitude and theoretically would have produced greater masking effects if loudness cues were driving both AM and FM detection. Chen and Zeng (2004) hypothesized that loudness cues associated with FM would be largely masked by the 1 dB of amplitude roving they incorporated in the FM signal. However, in the present study, while 1 dB of simultaneous NAM produced some interference with FM detection, 2 and 3 dB of NAM produced even greater interference.

When the starting phase of simultaneous SAM was randomized across stimulation intervals, FMDTs for sinusoidal FM were significantly elevated. Compared to in-phase SAM, random-phase SAM would produce similar amounts of interference with FM detection in the temporal pitch domain, and much greater interference in the loudness domain. Therefore, the poorer FMDTs with random-phase SAM suggest that FM and AM at least partly share a common loudness-based coding mechanism in CI listeners. It is also possible that different AM rates of simultaneous SAM would produce different amounts of interference with the detection of 10-Hz sinusoidal FM.

It is interesting to note that, for both types of simultaneous AM, FM sensitivity (in terms of $\Delta F/F$) was poorest at some intermediate standard frequency, which was higher for SAM (500 Hz) than for NAM (250 Hz). This is consistent with the effects of stimulation rate on loudness being greatest at high rates (McKay and McDermott, 1998), which allows the residual loudness cues associated with FM to rescue performance at high standard frequencies. In addition, assuming that loudness is integrated over some finite duration, simultaneous NAM, in which the amplitude was roved on a pulse-by-pulse basis, may have been less effective in producing random loudness fluctuations at high standard frequencies. This might account for the improved performance for standard frequencies above 250 Hz in the NAM conditions.

The present results raise important considerations for the design of "FM+AM" CI speech processing strategies (Zeng *et al.*, 2005). For such strategies, AM information (i.e., temporal envelope cues, encoded by varying the stimulation level over time) will be simultaneously presented with FM information (i.e., fine structure cues, encoded by varying the stimulation rate over time). The present results suggest that AM information will significantly interfere with FM information. Larger amounts of simultaneous AM will cause greater interference, and aperiodic AM has an even more pronounced effect. It is also possible (but less likely) that simultaneous FM may interfere with AM detection by CI

users. Therefore, in "FM+AM" CI speech processing strategies, there is a potential tradeoff between the reception of simultaneously presented AM and FM information.

V. CONCLUSIONS

FM detection thresholds for sinusoidal FM (modulation rate fixed at 10 Hz) and for upward frequency sweeps were measured with different amounts and types of simultaneous AM (SAM or NAM) in five adult CI subjects, as a function of standard frequency. FM sensitivity was significantly affected by the standard frequency, FM type, and AM condition. For all AM conditions and FM types, FMDTs significantly worsened with increasing standard frequencies. For all AM conditions and standard frequencies, FMDTs were higher for upward frequency sweeps than for 10-Hz sinusoidal FM. For all standard frequencies and FM types, FMDTs significantly worsened as the amount of simultaneous AM was increased; simultaneous NAM produced greater interference with FM detection than simultaneous SAM. FMDTs for sinusoidal FM significantly worsened when the starting phase of simultaneous SAM was randomized. The results suggest that in electric hearing, FM (as encoded by varying the stimulation rate) and AM (as encoded by varying the stimulation level) at least partly share a common loudness-based coding mechanism. The adverse effects of simultaneous AM on FM detection may limit the feasibility of "FM+AM" strategies for CI speech processing.

ACKNOWLEDGMENTS

We are grateful to all subjects for their participation in these experiments. We thank John J. Galvin III for editorial assistance. We would also like to thank Dr. Robert V. Shannon and two anonymous reviewers for their constructive comments on an earlier version of this paper. Research was supported in part by NIH (R01-DC004993 and R03-DC008192).

Chen, H. -B., and Zeng, F. -G. (**2004**). "Frequency modulation detection in cochlear implant subjects," J. Acoust. Soc. Am. **116**, 2269–2277.

Demany, L., and Semal, C. (**1986**). "On the detection of amplitude modulation and frequency modulation at low modulation frequencies," Acustica **61**, 243–255.

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

Edwards, B. W., and Viemeister, N. F. (**1994**). "Frequency modulation versus amplitude modulation discrimination: Evidence for a second frequency modulation encoding mechanism," J. Acoust. Soc. Am. **96**, 733–740.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X.-S. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q. -J., Chinchilla, S., and Galvin, J. J., III (**2004**). "The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users," J. Assoc. Res. Otolaryngol. **5**, 253–260.

Fu, Q. -J., Chinchilla, S., Nogaki, G., and Galvin, J. J., III (**2005**). "Voice gender identification by cochlear implant users: The role of spectral and temporal resolution," J. Acoust. Soc. Am. **118**, 1711–1718.

Fu, Q. -J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Fu, Q. -J., Shannon, R. V., and Wang, X. -S. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.

Goldstein, J. L., and Srulovicz, P. (**1977**). "Auditory-nerve spike intervals as an adequate basis for aural frequency measurement," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 337–347.

Grant, K. W. (**1987**). "Frequency modulation detection by normally hearing and profoundly hearing-impaired listeners," J. Speech Hear. Res. **30**, 558–563.

Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgements," Percept. Psychophys. **28**, 85–88.

Kong, Y. -Y., Cruz, R., Jones, J. A., and Zeng, F. -G. (**2004**). "Music perception with temporal cues in acoustic and electric hearing," Ear Hear. **25**, 173–185.

Kong, Y. -Y., Stickney, G. S., and Zeng, F. -G. (**2005**). "Speech and melody recognition in binaurally combined acoustic and electric hearing," J. Acoust. Soc. Am. **117**, 1351–1361.

Landsberger, D. M., and McKay, C. M. (**2005**). "Perceptual differences between low and high rates of stimulation on single electrodes for cochlear implantees," J. Acoust. Soc. Am. **117**, 319–327.

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

Luo, X., Fu, Q. -J., and Galvin, J. J., III (**2006**). "Vocal emotion recognition with cochlear implants," in *Proceedings of the International Conference on Spoken Language*. Processing, 2006, Pittsburgh, PA, pp. 1830–1833.

McDermott, H. J. (**2004**). "Music perception with cochlear implants: A review," Trends Amplif. **8**, 49–82.

McKay, C. M., and Carlyon, R. P. (**1999**). "Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains," J. Acoust. Soc. Am. **105**, 347–357.

McKay, C. M., and McDermott, H. J. (**1998**). "Loudness perception with pulsatile electrical stimulation: The effect of interpulse intervals," J. Acoust. Soc. Am. **104**, 1061–1074.

McKay, C. M., McDermott, H. J., and Carlyon, R. P. (**2000**). "Place and temporal cues in pitch perception: Are they truly independent?," ARLO **1**, 25–30.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). "Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees," J. Acoust. Soc. Am. **96**, 2664–2673.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1995**). "Pitch matching of amplitude-modulated current pulse trains by cochlear implantees: The effect of modulation depth," J. Acoust. Soc. Am. **97**, 1777–1785.

Moore, B. C. J., Glasberg, B. R., Gaunt, T., and Child, T. (**1991**). "Across-channel masking of changes in modulation depth for amplitude- and frequency-modulated signals," Q. J. Exp. Psychol. A **43**, 327–347.

Moore, B. C. J., and Sek, A. (**1994**). "Effects of carrier frequency and background noise on the detection of mixed modulation," J. Acoust. Soc. Am. **96**, 741–751.

Moore, B. C. J., and Sek, A. (**1996**). "Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking," J. Acoust. Soc. Am. **100**, 2320–2331.

Moore, B. C. J., and Skrodzka, E. (**2002**). "Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation," J. Acoust. Soc. Am. **111**, 327–335.

Rubinstein, J. T., Wilson, B. S., Finley, C. C., and Abbas, P. J. (**1999**). "Pseudospontaneous activity: Stochastic independence of auditory nerve fibers with electrical stimulation," Hear. Res. **127**, 108–118.

Shannon, R. V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," Hear. Res. **11**, 157–189.

Shannon, R. V., Adams, D. D., Ferrel, R. L., Palumbo, R. L., and Grandgenett, M. (**1990**). "A computer interface for psychophysical and speech research with the Nucleus cochlear implant," J. Acoust. Soc. Am. **87**, 905–907.

Siebert, W. M. (**1970**). "Frequency discrimination in the auditory system: Place or periodicity mechanisms," Proc. IEEE **58**, 723–730.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.

Stickney, G. S., Zeng, F.-G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Tong, Y. C., Clark, G. M., Blamey, P. J., Busby, P. A., and Dowell, R. C. (**1982**). "Psychophysical studies for two multiple-channel cochlear implant patients," J. Acoust. Soc. Am. **71**, 153–160.

Townshend, B., Cotter, N., Van Compernolle, D., and White, R. L. (**1987**).

"Pitch perception by cochlear implant subjects," J. Acoust. Soc. Am. **82**, 106–115.
Vongphoe, M., and Zeng, F. -G. (**2005**). "Speaker recognition with temporal cues in acoustic and electric hearing," J. Acoust. Soc. Am. **118**, 1055–1061.
Wilson, B. S., Lawson, D. T., Zerbi, M., and Finley, C. C. (**1994**). "Recent developments with the CIS strategies," in *Advances in Cochlear Implants*, edited by I. J. Hochmair-Desoyer and E. S. Hochmair (Manz, Vienna).
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.
Wilson, B. S., Wolford, R. D., and Lawson, D. T. (**2000**). "Speech processors for auditory prostheses," Center for Auditory Prosthesis Research, Research Triangle Park, NC, pp. 1–61.
Wygonski, J., and Robert, M. E. (**2001**). "HEI nucleus research interface specification," House Ear Institute.
Zeng, F. -G. (**2002**). "Temporal pitch in electric hearing," Hear. Res. **174**, 101–106.
Zeng, F. -G., Nie, K. B., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C.-G., and Cao, K. -L. (**2005**). "Speech recognition with amplitude and frequency modulations," Proc. Natl. Acad. Sci. U.S.A. **102**, 2293–2298.
Zeng, F.-G., and Turner, C. W. (**1991**). "Binaural loudness matches in unilaterally impaired listeners," Q. J. Exp. Psychol. A **43**, 565–583.
Zwicker, E. (**1956**). "Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs," "(The elementary bases to the regulation of information capacity of the hearing)," Acustica **6**, 356–381.

Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information

Kathryn Hopkins[a] and Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

(Received 15 November 2006; revised 24 April 2007; accepted 22 May 2007)

The ability of normally hearing and hearing-impaired subjects to use temporal fine structure information in complex tones was measured. Subjects were required to discriminate a harmonic complex tone from a tone in which all components were shifted upwards by the same amount in Hz, in a three-alternative, forced-choice task. The tones either contained five equal-amplitude components (non-shaped stimuli) or contained many components, but were passed through a fixed bandpass filter to reduce excitation pattern changes (shaped stimuli). Components were centered at nominal harmonic numbers (N) 7, 11, and 18. For the shaped stimuli, hearing-impaired subjects performed much more poorly than normally hearing subjects, with most of the former scoring no better than chance when $N=11$ or 18, suggesting that they could not access the temporal fine structure information. Performance for the hearing-impaired subjects was significantly improved for the non-shaped stimuli, presumably because they could benefit from spectral cues. It is proposed that normal-hearing subjects can use temporal fine structure information provided the spacing between fine structure peaks is not too small relative to the envelope period, but subjects with moderate cochlear hearing loss make little use of temporal fine structure information for unresolved components. [DOI: 10.1121/1.2749457]

PACS number(s): 43.66.Fe, 43.66.Sr [AJO] Pages: 1055–1068

I. INTRODUCTION

The frequency of a sine wave could be coded by the place of excitation on the basilar membrane (von Békésy, 1960; Zwislocki and Nguyen, 1999) or by the temporal information in the auditory nerve (phase locking), which is widely believed to be available for frequencies up to about 5 kHz (Rose *et al.*, 1967), although the exact upper limit varies across species (Palmer and Russell, 1986). Psychoacoustic evidence from humans suggests that the temporal mechanism plays an important role for frequencies up to about 5 kHz (Moore, 1973; Moore, 2003; Plack and Oxenham, 2005), although Heinz *et al.* (2001) have shown computationally that temporal information in auditory-nerve firing patterns is sufficient to account for psychophysical frequency discrimination performance in humans for frequencies up to at least 10 kHz.

It seems likely that temporal information is used to code the frequencies of resolved, lower harmonics in complex tones (Hartmann and Doty, 1996; Moore *et al.*, 1984, 2006c), either to allow extraction of the frequencies of constituent tones to be fed into a central "pattern recognizer" (Goldstein, 1973; Terhardt, 1974) or to establish a common time interval across harmonics that can be used to deduce the fundamental frequency ($F0$) of the complex (Moore, 1982; van Noorden, 1982).

Unresolved harmonics interact on the basilar membrane to form a complex periodic waveform with a period of $1/F0$. Complexes with only unresolved harmonics have a pitch corresponding to their $F0$, so this period must be extracted by temporal mechanisms. The temporal information has two forms. "Temporal fine structure" refers to the rapidly fluctuating variations in amplitude of the waveform. "Envelope" refers to slower modulation superimposed on this fine structure, which occurs at a rate equal to $F0$. Periodicity extraction from a complex waveform could occur by measurement of the time period between corresponding peaks in the envelope of the waveform, or by measurement of the time interval between corresponding peaks in the temporal fine structure. The use of temporal fine structure cues would allow greater accuracy in coding waveform periodicity (Moore *et al.*, 2006a; 2006b), but would require greater precision in temporal coding. For waveforms with temporal fine structure frequencies greater than 4–5 kHz (the frequency above which use of phase-locking information appears to break down in humans), periodicity extraction must be on the basis of envelope repetition rate alone. For lower temporal fine structure frequencies, either coding strategy is possible.

To assess the role of temporal fine structure, de Boer (1956) and Schouten *et al.* (1962) obtained pitch matches to "frequency-shifted" complex tones, derived from harmonic complexes by shifting each component upwards (or downwards) by the same amount in Hertz. The envelope repetition rate was the same as for the original harmonic complex, but the time intervals between peaks in the fine structure were smaller (or larger). The matched pitch was found to shift with the shift of the component frequencies, suggesting that temporal fine structure plays an important role in the pitch perception of complex tones. However, Moore and Moore (2003b) argued that these results might have been influenced by shifts in the spectrum or excitation pattern produced by

[a]Electronic mail: kh311@cam.ac.uk

the frequency shift of the components. To eliminate this effect, Moore and Moore (2003b) conducted an experiment similar to those of de Boer (1956) and Schouten *et al.* (1962), but both the test and matching tones were spectrally shaped so that they evoked very similar excitation patterns on the basilar membrane when no resolved components were present. They used complexes with components that were resolved (low relative to *F*0), unresolved (high relative to *F*0), or intermediate, but probably containing mainly unresolved harmonics (Moore *et al.*, 2006b). For the resolved condition, matching tones were made up of harmonics in a different spectral region to the test tone to prevent comparisons between the frequencies of individual components. For the intermediate and resolved conditions, subjects matched the inharmonic tones to harmonic tones with higher envelope repetition rates. This suggests that temporal fine structure information does play a role in the pitch perception of complex tones with intermediate harmonic numbers, as the complexes in the intermediate condition contained harmonics that were probably unresolved and no pitch shift would be expected if only envelope cues were used. In contrast, subjects matched harmonic and inharmonic tones with the same envelope repetition rate for the unresolved condition. This suggests that only envelope cues were used to compare the pitch of these tones, and that temporal fine structure cues did not play a role, even when temporal fine structure fluctuations were below the frequency at which phase locking is thought to break down.

An alternative explanation for the pitch shifts observed in the intermediate condition by Moore and Moore (2003b) is that subjects matched the partially resolved harmonics in the test and matching stimuli rather than using temporal fine structure cues. One paper has presented evidence that harmonics up to the 11th or 12th may be resolved (Bernstein and Oxenham, 2003), although other work suggests that only harmonics below the eighth are resolvable (Moore and Ohgushi, 1993; Moore *et al.*, 2006c; Plomp, 1964). To address this issue, Moore *et al.* (2006b) measured difference limens for the *F*0 (*F*0 DLs) of three-component complex tones using normal-hearing listeners. The nominal frequency of the center component was fixed, but the harmonic number of that component, *N*, was varied. For example, for a center frequency of 2000 Hz, a three-component complex with $N=9$ would have components with frequencies 1777.8, 2000, and 2222.2 Hz ($F0=222.2$ Hz). A complex with $N=10$ would have components of 1800, 2000, and 2200 Hz ($F0=200$ Hz). Discrimination was tested when all components had the same starting phase (cosine phase) and when the phase of the central component was shifted by 90 deg (alternating phase). It was argued that the presence of a phase effect would indicate that the components were not resolved (Bernstein and Oxenham, 2005; Houtsma and Smurzynski, 1990; Moore, 1977; Shackleton and Carlyon, 1994).

For complexes with *N* equal to 6 or 7, there was no significant phase effect. However, *F*0 DLs were smaller for complexes presented in cosine phase than for those presented in alternating phase when *N* was greater than 8. The phase effect suggests that components were not resolved for complexes with *N* of 8 or more. This suggests in turn that the pitch shifts observed by Moore and Moore (2003b) for complexes with intermediate harmonic numbers were due to use of temporal fine structure information rather than comparison of excitation patterns of partially resolved components.

Listeners with cochlear hearing loss usually show a poor ability to discriminate the pitch of complex sounds, even when the sounds are presented well above the detection threshold (Moore and Carlyon, 2005). These deficits can be partly attributed to damage to outer hair cells resulting in a broadening of the auditory filters (Glasberg and Moore, 1986; Liberman and Kiang, 1978; Pick *et al.*, 1977). This would lead to a reduced ability to resolve the partials in complex tones and to more complex waveforms at the outputs of the auditory filters (Rosen, 1987). Auditory filter broadening cannot, however, explain all deficits associated with cochlear hearing loss. Moore and Peters (1992) found only a weak correlation between pure-tone frequency discrimination and auditory filter sharpness at a particular frequency, though recent work by Bernstein and Oxenham (2006) showed a correlation between frequency selectivity in hearing-impaired subjects, and the position of the transition in *F*0 discrimination ability from good to poor as the number of the lowest harmonic was increased from low values towards higher values.

One animal study suggested that cochlear damage may lead to a deficit in phase locking, which would impair the ability to use temporal fine structure information (Woolf *et al.*, 1981). However, another study showed no phase locking deficits in guinea pigs with kanamycin-induced outer hair cell damage (Harrison and Evans, 1979). These contradictory results may reflect a species difference or a difference in the methods used to induce cochlear damage. It is unclear whether humans with hearing impairments suffer phase-locking deficits, but if they do, then a reduced ability to use temporal fine structure information would be expected. A phase-locking deficit is not the only pathology that may result in a reduced ability to make use of temporal fine structure information, however. A change in cochlear tuning caused by outer hair cell damage can result in a change in the phase response properties of auditory filters. It has been suggested that temporal fine structure information may be extracted by cross correlation of auditory filter outputs—a particular phase shift in response between adjacent places along the basilar membrane would indicate a particular signal frequency (Loeb *et al.*, 1983; Shamma, 1985). Changes in filter phase response through a loss of the active mechanism might disrupt this cue and reduce the ability to extract temporal fine structure information.

A number of psychoacoustic and speech perceptual studies suggest that hearing-impaired listeners have an impaired ability to use temporal fine structure information (Buss *et al.*, 2004; Lacher-Fougère and Demany, 1998, 2005; Lorenzi *et al.*, 2006; Moore and Moore, 2003a; Moore and Skrodzka, 2002; Moore *et al.*, 2006a).

Moore and Moore (2003a) showed that hearing-impaired subjects relied more than normal-hearing subjects on spectral cues to discriminate the *F*0 of complex tones. They also showed that hearing-impaired subjects had similar *F*0 DLs for tones with intermediate harmonics and with high

harmonics. This contrasts with results for normal-hearing subjects, where smaller $F0$ DLs were found for tones with intermediate harmonic numbers. Moore and Moore (2003a) suggested that this difference might have occurred because the hearing-impaired subjects could not use temporal fine structure information and relied instead on envelope or spectral cues, which remained similar as the harmonic number increased.

Further evidence suggesting that the ability to use temporal fine structure information is reduced for hearing-impaired listeners comes from recent research by Moore *et al.* (2006a). $F0$ DLs were measured for hearing-impaired listeners using the same procedures and stimuli as in Moore *et al.* (2006b). For a center frequency of 2 kHz, $F0$ DLs tended to improve with increasing N, while for normally hearing subjects $F0$ DLs worsened with increasing N in the range $N=8$ to 13. The pattern of results for the hearing-impaired subjects was similar to that found for normal-hearing subjects at a center frequency of 5 kHz, a frequency for which phase locking is believed to be absent. The authors interpreted this result as evidence that hearing-impaired subjects were not using temporal fine structure information even at frequencies where phase locking is believed to be robust in normally hearing subjects.

Here, we attempted to determine more directly the extent to which hearing-impaired listeners can use temporal fine structure information, by measuring the discrimination of harmonic and frequency-shifted complex tones under conditions where temporal fine structure cues were available, but envelope and spectral cues were limited or absent.

II. RATIONALE

Moore and Moore (2003b) described an experiment in which a harmonic complex tone was matched in pitch to a frequency-shifted tone; the two tones were bandpass filtered so that they evoked very similar excitation patterns when components were unresolved. For frequency-shifted complexes containing only components with high frequencies relative to $F0$, no pitch shift was measured for normal-hearing subjects, suggesting that temporal fine structure information in these stimuli was inaccessible. If hearing-impaired subjects cannot use temporal fine structure information, even for complexes with lower harmonic numbers, then no pitch shift for these frequency-shifted complexes would be expected. However, when the experiment was attempted with hearing-impaired listeners, results were very erratic. Hearing-impaired listeners could not make consistent pitch matches, so no conclusions about their ability to use fine structure information could be drawn.

Here, similar stimuli were used, but a three-interval forced-choice task rather than a pitch-matching task was chosen, as this was expected to be easier for non-musically trained hearing-impaired listeners to perform. Subjects were required to discriminate frequency-shifted complexes from harmonic complexes; as in Moore and Moore (2003b), stimuli were bandpass filtered so that they would evoke almost the same excitation pattern on the basilar membrane when only unresolved components were present. If the hearing-impaired subjects are unable to use temporal fine structure information, they should be unable to discriminate the frequency shifted and non-frequency-shifted complexes.

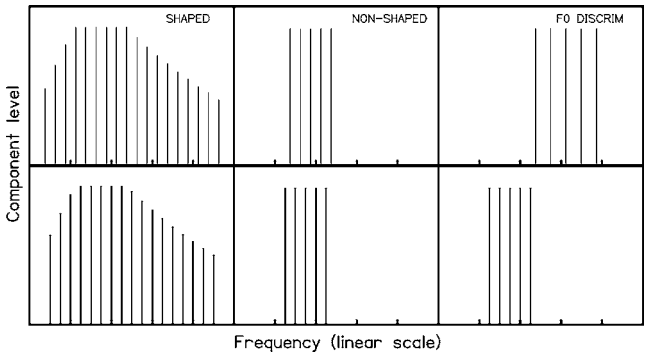


FIG. 1. Schematic spectra of stimuli for shaped, non-shaped, and $F0$-discrim stimulus types. Reference (unshifted) stimuli are shown in the bottom row, and shifted stimuli are shown in the top row. The $F0$ was 400 Hz and N was equal to 11.

A concern with this method is that, if hearing-impaired subjects do indeed perform this task very poorly, this could be because they have failed to understand the task or because performance was limited by cognitive factors, such as the ability to remember the three stimuli within a trial. Cognitive factors were a potential concern since some of the hearing-impaired subjects included in the present study were elderly. To exclude this interpretation, and to further investigate the cues used for discrimination by normal-hearing and hearing-impaired subjects, two additional types of stimuli were used. The three stimuli types are illustrated in Fig. 1.

For stimulus type 1 ("shaped"), the stimuli to be discriminated were the bandpass filtered harmonic and frequency-shifted tones described above. The components of the frequency-shifted tone had the same spacing as for the harmonic comparison tone, but each component was shifted upwards by the same amount in Hertz. For stimuli with only unresolved components (high frequencies relative to $F0$), the main cue for discrimination was changes in temporal fine structure, although changes in envelope might have been usable to a small extent (this is discussed in more detail later). For conditions with lower frequency components relative to $F0$, the task could be performed by comparing the frequencies of individual resolved components.

For stimulus type 2 ("non-shaped"), subjects discriminated harmonic and inharmonic tones similar to those for the shaped condition, except that complexes were made up of five equal-amplitude components only, meaning that upwards shifts in frequency were accompanied by an upward shift in the "center of gravity" of the excitation pattern, even when harmonics were unresolved.

For stimulus type 3 ("$F0$-discrim"), subjects were required to detect a change in $F0$ of a five-component harmonic complex. $F0$ DLs have been measured many times previously with similar stimuli (Arehart, 1994; Moore and Glasberg, 1988; Moore and Peters, 1992), and it has been found that even severely hearing-impaired subjects can perform the task if the difference in $F0$ is made large enough. Envelope, temporal fine structure, and spectral cues were all available for these stimuli.

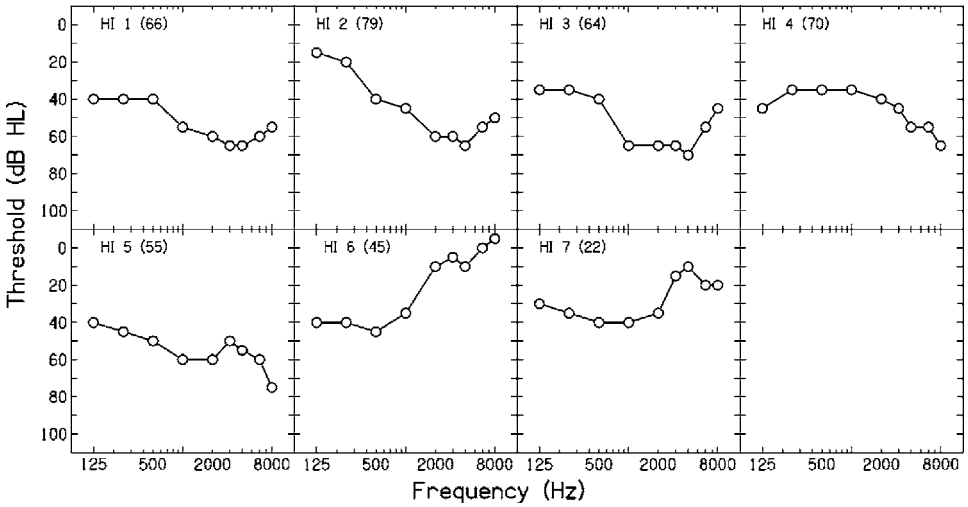


FIG. 2. Air conduction thresholds for the test ears of the hearing-impaired subjects. The age of each subject is also shown.

An additional reason for using the non-shaped and *F*0-discrim stimuli was to prevent the hearing-impaired subjects from becoming unduly discouraged; we found in pilot studies that the performance of these subjects with shaped stimuli was often very poor.

III. METHODS

A. Subjects

Nine subjects with normal hearing and seven hearing-impaired subjects were recruited for this experiment. Five of the nine normal-hearing subjects and four of the seven hearing-impaired subjects had previous experience of psychoacoustical experiments, and one of the normal-hearing listeners and two of the hearing-impaired listeners were musically trained. Normal-hearing subjects were between 20 and 22 years of age and had absolute thresholds of less than 15 dB HL in their test ears at all audiometric frequencies. Audiograms of the test ears of the hearing-impaired subjects are shown in Fig. 2. These were measured using a Grason-Stadler GSI 61 audiometer with Telephonics TDH 50-P earphones.

The age of each subject is indicated in the figure. All hearing-impaired subjects had air-bone gaps of less than 15 dB at octave frequencies between 500 and 4000 Hz, indicating that there was no large conductive component of the hearing loss. Hearing-impaired subjects were tested for cochlear dead regions for octave frequencies between 500 and 4000 Hz, using the method described by Moore *et al.* (2004). No dead regions were found in any subject. Two of the hearing-impaired subjects (HI 6 and HI 7) had little or no hearing loss at high frequencies.

The ear with the smaller variation in pure-tone thresholds across frequency was chosen as the test ear for each subject. For HI 6, the worse-hearing ear was used as the test ear, so it was necessary to prevent "cross hearing." Masking noise generated with an IVIE IE-20B pink noise generator was presented to the nontest ear at a level of 30.6 dB/ERB_N at 1000 Hz [where ERB_N refers to the equivalent rectangular bandwidth of the auditory filter for young, normally hearing subjects tested at moderate sound levels, Glasberg and Moore (1990)]. This noise level was calculated from the maximum stimulus level per ERB_N in the test ear. The level reaching the contralateral ear was determined by subtracting 40 dB from this value, which is the interaural attenuation measured for the headphones that were used. The final level of the contralateral masker was calculated by adding 6 dB to this level to ensure masking.

Five normal-hearing subjects were tested with shaped stimuli (NH 1–NH 5), five with non-shaped stimuli (NH 6–NH 10), and four with the *F*0-discrim stimuli (NH 6, NH 7, NH 9, and NH 10). All six hearing-impaired subjects were tested using shaped stimuli, four were tested using non-shaped stimuli (HI 3, HI 4, HI 6, and HI 7), and three were tested using *F*0-discrim stimuli (HI 1, HI 2, and HI 5). The order in which conditions were tested was randomized for each subject.

Subjects were paid for their time, apart from one of the authors, HI 7. Subjects underwent training until performance appeared to be stable; this usually took 1 hour, but took a longer time for some subjects.

B. Stimuli

Nominal *F*0's of 100, 200, and 400 Hz were used. For normal-hearing subjects, discrimination was measured with $N=7$, 11, and 18 for each *F*0, making nine conditions for each stimulus type. The condition with $F0=400$ Hz and $N=18$ was not tested for five of the seven hearing-impaired subjects (HI 1–HI 5), as residual hearing in this high-frequency region (around 7200 Hz) was very poor.

1. Stimulus type 1 (shaped): Discrimination of harmonic and frequency-shifted complexes spectrally shaped to reduce excitation pattern cues

A trial consisted of three intervals, two containing a harmonic complex and one containing a frequency-shifted complex. For the harmonic complex, multiple harmonics of the *F*0 were added, each starting in sine phase. The inharmonic complex was formed in the same way as the harmonic complex except that each component was shifted upwards in frequency by the same amount in Hertz. For example, for a harmonic complex that contained (among others) components with frequencies of 600, 700, and 800 Hz, the corre-

TABLE I. Largest differences in excitation level between harmonic and frequency-shifted shaped tones using the maximum frequency shift of 0.5*F*0. Excitation patterns were calculated using a model for normal hearing (Moore *et al.*, 1997).

*F*0	*N*	Largest difference in excitation level (dB)
100	7	4.4
100	11	3.7
100	18	1.9
200	7	5.9
200	11	4.4
200	18	2
400	7	6.3
400	11	4.1
400	18	2.1

sponding inharmonic complex, with a shift of 20 Hz, would contain components with frequencies of 620, 720, and 820 Hz. The amplitudes of the harmonics were defined using a bandpass filter function with a central flat region with a width of 5*F*0 and skirts that decreased in level at a rate of 30 dB/octave; see Fig. 1.

The maximum shift was 0.5*F*0 Hz. Excitation patterns for harmonic and frequency-shifted complexes with the largest possible frequency shift (0.5*F*0) were calculated using the model proposed by Moore *et al.* (1997), to check that differences were small when the components in the passband were unresolved. Table I shows the largest difference in excitation level for the harmonic and frequency-shifted complexes for each condition. Examples of excitation patterns for $N=7$, 11, and 18 are shown in Fig. 3. The largest excitation-pattern differences were about 4 dB for $N=11$ and about 2 dB for $N=18$. However, the model used to calculate these differences applies to normal-hearing subjects; excitation-pattern differences for hearing-impaired subjects would be smaller than calculated using this model, at least in frequency regions of hearing loss, probably by a factor of 2 or more, since the auditory filters typically broaden with increasing hearing loss (Glasberg and Moore, 1986).

This means that the changes in excitation level for the hearing-impaired subjects would typically have been 2 dB or less. Thresholds for detecting a change in excitation level in a restricted frequency region are typically 2–4 dB (Moore *et al.*, 1989). Hence, the excitation-pattern changes would have been barely, if at all, detectable for the hearing-impaired subjects.

Note that when $N=7$, components would be resolved for normally hearing subjects, so the excitation pattern is clearly different for the harmonic complex (solid line) and the frequency-shifted complex (dotted line). This could allow discrimination using excitation pattern cues, although such cues would be less effective for hearing-impaired subjects.

2. Stimulus type 2 (non-shaped): Discrimination of harmonic and frequency-shifted complexes without spectral shaping

Stimuli were synthesized in a similar way as for the shaped stimuli. However, harmonic and frequency-shifted complexes were made up of only five equal-amplitude components. No flanking components were presented. Consequently, in addition to temporal fine structure information, a change in excitation pattern could be used as a cue to identify the frequency-shifted complex, even when components were unresolved.

3. Stimulus type 3 (F0-discrim): Discrimination of the F0 of a harmonic complex

Harmonic complexes were formed from five harmonics added in sine phase. Two intervals contained a complex tone with *F*0 equal to the nominal value (100, 200, or 400 Hz), and the other interval contained a complex tone with *F*0 equal to the nominal value plus δ. Envelope, spectral, and temporal fine structure cues could all be used for this stimulus type.

4. Threshold equalizing noise (TEN)

Threshold equalizing noise (TEN) (Moore *et al.*, 2000), extending from 200 to 16 000 Hz, was used to mask combination tones for all task types, and, for the shaped stimuli, to help ensure that the audible parts of the excitation patterns evoked by the harmonic and frequency-shifted tones were the same. TEN was chosen because it is designed to give equal masked thresholds (in dB SPL) across frequency for

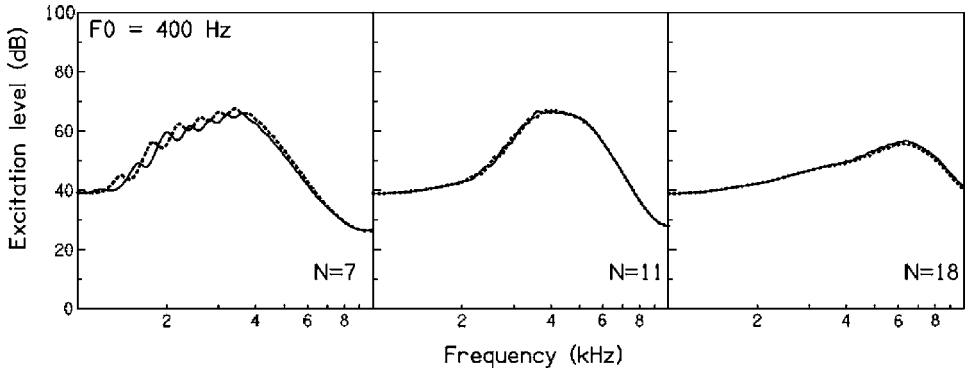


FIG. 3. Excitation patterns (Moore *et al.*, 1997) for shaped stimuli with *F*0=400 Hz and $N=7$, 11, and 18, presented in pink noise with a spectrum level of 18 dB at 1000 Hz. This noise was designed to give roughly the same excitation pattern as the TEN noise used in the experiment (see the text). The patterns are plotted only over the frequency range where the shaped stimuli produced excitation comparable to or above that produced by the noise. Patterns for harmonic and frequency-shifted stimuli are plotted as solid and dotted lines, respectively. The frequency shift was 0.5*F*0 Hz (the maximum shift).

TABLE II. Conditions for which the nonadaptive procedure (na) was used for the normal-hearing subjects. Stimulus types shaped (SH) and non-shaped (N-SH) are shown only, as all subjects completed the adaptive procedure for all conditions in the *F*0-discrim task.

*F*0	*N*	NH 1 SH	NH 2 SH	NH 3 SH	NH 4 SH	NH 5 SH	NH 6 N-SH	NH 7 N-SH	NH 8 N-SH	NH 9 N-SH	NH 10 N-SH
100	7										
100	11										
100	18		na	na	na	na					
200	7										
200	11										
200	18		na	na	na	na					
400	7										
400	11										
400	18	na	na	na	na	na		na			na

subjects with normal hearing, and approximately equal masked thresholds for subjects with cochlear hearing loss, but without dead regions (Moore *et al.*, 2000). This meant that we could be confident that combination tones of a particular level would be masked irrespective of their frequency. For normal-hearing subjects, the TEN level at 1 kHz was set to 20 dB/ERB_N below the level of the most intense component in the complex tones. For hearing-impaired subjects, the TEN level was set to 30 dB/ERB_N below the level of the most intense component, to prevent the noise from being uncomfortably loud while still providing sufficient masking.

C. Signal generation

Stimuli were produced by a Tucker Davis Technologies (TDT) system II, using a 16-bit digital-to-analog converter (TDT DA4) with a sampling rate of 50 kHz. Levels of the tones and the TEN were controlled independently using two TDT PA4 attenuators; stimuli were low-pass filtered at 20 kHz using Kemo (VBF8) dual variable filters. Stimuli were presented via Sennheiser HD580 headphones in a double-walled sound-attenuating chamber.

For normal-hearing listeners, each complex was presented at an overall level of 65 dB SPL. Complexes were presented to hearing-impaired subjects at a sensation level of 20 dB, except for subjects HI 1 and HI 7 who found this level to be too quiet. For them, a level of 30 dB SL was used. The audiograms of the hearing-impaired subjects were used to calculate sensation levels at the frequency corresponding to *N* for each condition. Linear interpolation between audiometric frequencies and conversion to dB SPL allowed the appropriate level to be calculated. Two hearing-impaired subjects (HI 6 and HI 7) had normal audiometric thresholds at some frequencies, so for some conditions, a level of 20 or 30 dB SL would give levels much lower than those presented to normal-hearing subjects. For these conditions, an overall level of 65 dB SPL was used, which was the same as the level used for normal-hearing subjects.

D. Procedure

The same experimental procedure was used for all stimulus types. When two stimulus types were tested using the same subject, they were interleaved and the subject was not informed of which stimulus type was being presented.

A trial consisted of three successive stimuli, indicated by lights on a response box. Each stimulus was 540 ms long, including 20-ms raised-cosine onset and offset ramps. The interstimulus interval was 200 ms. Two intervals contained the same stimulus and the third, chosen at random, contained a different one. Subjects were instructed to press the button corresponding to the interval that sounded different. Feedback was given after every trial via lights on the response box.

Initially, we tried to use an adaptive procedure to estimate "thresholds" for discrimination for each stimulus type. We assumed that discriminability would increase monotonically with increasing frequency shift, even for the shaped stimuli, since, for normally hearing subjects, the pitch shift increases monotonically with increasing frequency shift up to at least 0.25*F*0 (Moore and Moore, 2003b). This assumption was confirmed by the orderly nature of the adaptive tracks obtained for conditions where performance was relatively good. However, especially for the shaped stimuli, performance was sometimes too poor for the adaptive procedure to be used, since the procedure called for a change larger than the maximum possible value (a shift of 0.5*F*0 for the shaped stimuli). In such cases, performance was measured with the shift fixed at 0.5*F*0. Details of the subjects and conditions where the nonadaptive procedure was used are shown in Tables II and III for the normal-hearing and hearing-impaired subjects, respectively. Column headings SH and N-SH refer to shaped and non-shaped stimulus types, and the abbreviation "na" indicates conditions for which the nonadaptive procedure was used. The two types of procedure are described below.

1. Adaptive procedure

The tracking variable (the frequency shift of the components for the shaped and non-shaped stimuli or the value of δ for the *F*0-discrim stimuli) was varied adaptively using a three-down, one-up tracking procedure (Levitt, 1971). It was increased by a factor *K* after one incorrect response and was decreased by the same factor after three consecutive correct responses. For the first four turnpoints, *K* equaled 1.414 and for the last eight turnpoints, it was reduced to 1.189. Twelve turnpoints were obtained and the geometric mean of the frequency differences at the last eight was taken to be the threshold corresponding to 79.4% correct. The standard de-

TABLE III. As Table II, but for the hearing-impaired subjects. Dashes indicate conditions that were not tested.

		HI 1	HI 2	HI 3		HI 4		HI 5	HI 6		HI 7	
*F*0	*N*	SH	SH	SH	N-SH	SH	N-SH	SH	SH	N-SH	SH	N-SH
100	7			na		na		na	na			
100	11	na	na	na		na	na	na	na		na	
100	18	na	na	na	na	na	na	na	na		na	
200	7	na	na	na		na	na	na	na			
200	11	na	na	na	na	na		na				
200	18	na	na	na	na	na	na	na	na		na	na
400	7	na	na	na	na	na		na				
400	11	na	na	na	na	na		na				
400	18	- - -	- - -	- - -	- - -	- - -	- - -	- - -	na		na	

viation of the logarithms of the turnpoint values was also calculated. If this standard deviation was greater than 0.2 then results of the run were discarded and the condition was repeated.

Each condition was tested three times and the geometric mean of the threshold estimates was calculated. The standard deviation of the log values of the threshold estimates was determined, and if it was greater than 0.15, then an extra run was obtained, and the final threshold was taken as the geometric mean of all four runs.

If a value of 0.5*F*0 was reached after the fourth turnpoint for the shaped and non-shaped stimuli, then the run was aborted. Before the fourth turnpoint, the frequency difference was allowed to reach 0.5*F*0 four times before the run was aborted, as the first four turnpoints were not used in calculation of the run geometric mean. When a run was aborted, the subject was later retested on the same condition, using the adaptive procedure. If a subject was consistently unable to complete runs using the adaptive procedure, the nonadaptive procedure described below was used. It should be noted that the abortion of some runs, but acceptance of others for the same condition, might have led to a bias to accept runs for which performance was better; thus, the thresholds estimated using the adaptive procedure may be underestimates of the "true" thresholds in cases where the threshold approached 0.5*F*0.

2. Nonadaptive procedure

This procedure was only used for shaped and non-shaped stimulus types, as all subjects could complete the adaptive procedure for the *F*0-discrim stimuli for all conditions.

Subjects were given the same instructions as for the adaptive procedure. They selected the different interval from three alternatives and feedback was given as before. A run consisted of 55 trials, with the last 50 trials being used to calculate a percent-correct value. The frequency shift was fixed at the maximum difference of 0.5*F*0. Subjects completed five nonadaptive runs. The mean of the percent-correct values for each run was used to estimate the final-percent correct value.

E. Statistics

To allow results from the two types of procedure (adaptive and nonadaptive) to be compared, thresholds from the adaptive procedure and percent-correct values from the nonadaptive procedure were converted to d' values (Green and Swets, 1974). Conversion was by use of a table of d' values for *m*-alternative forced-choice procedures (Hacker and Ratcliff, 1979). The adaptive procedure tracked the 79.4% correct point on the psychometric function, which corresponds to a d' of 1.63 for a three-alternative, forced-choice task. When the adaptive procedure was used, the d' value that would have been measured for a difference of 0.5*F*0 Hz was calculated by dividing 1.63 by the threshold measured in the adaptive procedure, and multiplying this value by 0.5*F*0. This method for calculating d' assumes that d' is proportional to the frequency shift in Hz (Nelson and Freyman, 1986). However, this assumption is not critical for interpreting the results. This method sometimes yielded values of d' that were much larger than would normally be encountered as, for some conditions, a frequency difference of 0.5*F*0 Hz was much larger than the threshold value. Such large values of d' would be difficult or impossible to measure in practice, and should not be taken too literally. The main point here is that the calculated d' values are inversely proportional to the estimated threshold values, so that large d' values indicate good performance. Table IV summarizes results for conditions for which the adaptive procedure was completed by all of the normally hearing subjects. The thresholds in Hertz, measured using the adaptive procedure are shown, as well as the calculated d' values, so the two can be related.

Statistical tests were performed on the square root of the absolute d' values, as this transformation gave a roughly uniform variance (across subjects in the case of data for the normal-hearing subjects and within subjects in the case of the data for the hearing-impaired subjects) across conditions. Values that were negative before the transformation were multiplied by −1 after the transformation to restore their sign.

IV. RESULTS

Mean d' values for the normal-hearing subjects for the shaped and non-shaped stimuli are shown in Fig. 4. Open and filled symbols denote results for the shaped and non-

TABLE IV. Summary of results for normal-hearing subjects for conditions where all subjects completed the adaptive procedure. Geometric means of the thresholds measured using the adaptive procedure are shown in Hertz. The d' values that would have been measured for discrimination of tones with a frequency shift of 0.5 $F0$ Hz were estimated by extrapolation from these thresholds (see Sec. III E for details).

		shaped		non-shaped		$F0$-discrim	
$F0$	N	Hz	d'	Hz	d'	Hz	d'
100	7	5.0	18.1	4.2	19.8	0.67	123.1
100	11	14.6	7.0	9.7	8.8	0.82	99.8
100	18			22.5	3.7	1.6	50.9
200	7	6.8	26.8	6.9	24.4	1.5	111.7
200	11	16.2	12.1	22.3	7.5	1.8	91.7
200	18			39.3	4.2	2.8	59.2
400	7	11.70	32.7	18.9	19.9	2.6	137.4
400	11	31.9	13.7	33.2	10.1	3.9	85.0
400	18					10.5	31.7

shaped stimuli, respectively. Values of d' are plotted on a square-root scale, as the standard deviation of the d' values was roughly proportional to the square root of their magnitude. Some conditions yielded d' values that were not significantly different from zero (chance performance; see later for details of how this was determined). These points were assigned a different symbol (a square) and were plotted at zero. No error bars are shown for these points.

Figure 4 shows that performance worsened as N increased. This was true for all $F0$'s tested and for both stimulus types. There was little difference between d' values for the two stimulus types when $N=7$ or 11, but when $N=18$, performance was better for the non-shaped stimuli than for the shaped stimuli. A within-subjects, repeated measures analysis of variance (ANOVA) was performed with factors $F0$ (100, 200, and 400 Hz), N (7, 11, and 18), and stimulus type (shaped and non-shaped). Throughout this paper, the degrees of freedom used to calculate p values for each factor were corrected using the Greenhouse-Geisser correction. The effect of N was significant $[F(2,16)=136.9; p<0.001]$, but the main effects of $F0$ and stimulus type were not. The interaction between N and stimulus type was significant $[F(1.26,16)=7.9;\ p=0.014]$, and *post hoc* tests of the effect of stimulus type using the least-significant-differences (LSD) test (using the pooled error term from the ANOVA) (Keppel, 1991) showed that the effect of stimulus type was significant for $N=18$ ($p=0.024$), but not for $N=7$ or 11 ($p=0.462$ and 0.757, respectively). The implications of these results will be discussed later.

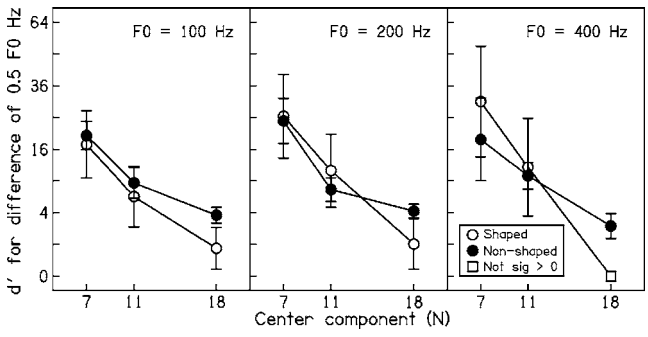


FIG. 4. Mean d' values for normal-hearing subjects for discrimination of harmonic and frequency-shifted complex tones, plotted as a function of N. Each panel shows results for one $F0$. Open and filled circles show d' values for shaped and non-shaped stimuli, respectively. Error bars show ±1 standard deviation of the mean. The square symbol indicates that the d' value is not significantly different from zero. This point is plotted at zero.

The results for the hearing-impaired listeners showed marked individual differences, so further analysis was performed on individual subject data only. Data for the four subjects who were tested using shaped and non-shaped stimuli are shown in Fig. 5. Symbols have the same meaning as for Fig. 4. The data for the three other hearing-impaired subjects who were tested with shaped stimuli (HI 1, HI 2, and HI 5) are not plotted, as they performed very poorly with those stimuli; performance was close to the chance level for most conditions, as described in more detail later. Subjects HI 3 and HI 4 showed better performance in all conditions for non-shaped than for shaped stimuli. HI 6 showed a similar pattern for $F0=100$ and 200 Hz, but for $F0=400$ Hz the pattern of results was more like that seen for normal-hearing subjects, with similar performance for the two stimulus types when $N=7$ or 11 but better performance for the non-shaped stimuli for $N=18$. The fact that HI 6 showed similar performance to the normal-hearing subjects for $F0=400$ Hz probably reflects the near-normal absolute thresholds of HI 6 for frequencies of 2000 Hz and above. HI 7 showed better performance for the non-shaped than for the shaped stimuli for all N for $F0=100$ Hz, but for $F0=200$ and 400 Hz the pattern of performance for HI 7 was similar to that for the normal-hearing subjects. Again, this probably reflects the fact that HI 7 had near-normal hearing for frequencies of 3000 Hz and above.

A "repeated-measures" ANOVA was performed on the individual data with the same factors as for the ANOVA performed on the data for the normal-hearing subjects. The last three estimates of d' for each condition were treated as replications. A summary of the results is shown in Table V (subjects HI 3, HI 4, HI 6, and HI 7). The degrees of freedom shown in the table are those before the Greenhouse-Geisser correction was applied; after correction, the degrees of freedom differed slightly across subjects. All p values incorporate the correction. For subjects who did not complete the conditions where $F0=400$ Hz and $N=18$, those conditions

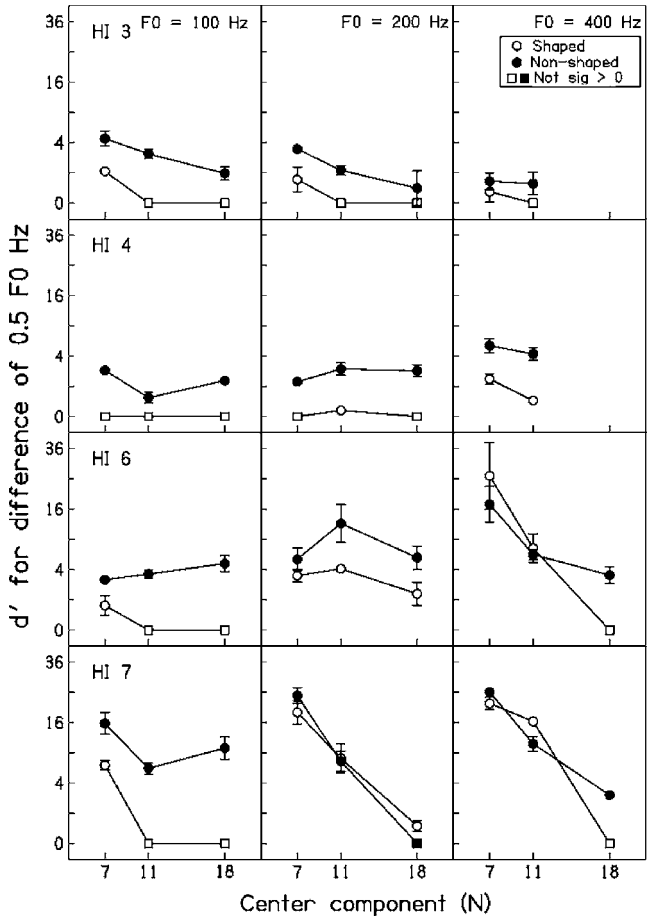


FIG. 5. As Fig. 4, but showing individual d' values for the hearing-impaired subjects. The unfilled and filled square symbols indicate d' values that are not significantly different from zero for shaped and non-shaped stimuli, respectively. These points are plotted at zero.

were treated as missing values for the purpose of the analysis. The missing values were estimated using the statistical package GENSTAT, although the ANOVAs were conducted using another package, SPSS. The main effects of *F*0 and stimulus type were significant for all subjects and the effect of *N* was significant for all subjects expect HI 4 ($p > 0.05$). Subjects HI 6 and HI 7 showed significant two-way and three-way interactions between all of the factors. HI 4 showed no significant interaction effects, and HI 3 showed a significant interaction effect between *F*0 and stimulus type. The nature of these interactions is discussed later on.

For some conditions, subjects performed very poorly. To assess whether poor performance ($d' < 0.5$) was significantly better than chance level ($d' = 0$), the standard error of the mean of the transformed d' values for that condition was calculated (see Sec. III E for details of the transformation). Performance was considered not significantly better than chance if the value zero fell within 2 standard errors of the mean of the transformed d' values for a particular condition. Conditions where performance was not significantly better than chance are shown in Table VI. Stimulus types shaped and non-shaped are denoted by column headings SH and N-SH, respectively. For HI 1–HI 4, performance for shaped stimuli was extremely poor, with discrimination significantly better than chance in only a few conditions, typically those where $N = 7$. Performance was, however, significantly better than chance for all conditions and subjects for the non-shaped condition, except for HI 7 and HI 3 when *F*0 = 200 Hz and $N = 18$.

Results for the *F*0-discrim task are shown in Fig. 6. Mean results are plotted for the four normally hearing subjects tested, and individual results are shown for the three

TABLE V. Summary of the ANOVA outcomes for hearing-impaired subjects tested with shaped and non-shaped stimuli.

	Main								Interactions					
	*F*0		*N*		Stimulus type		*F*0.N		*F*0.type		N.type		*F*0.N.type	
	$F(2,36)$	p	$F(2,36)$	p	$F(1,36)$	p	$F(4,36)$	p	$F(2,36)$	p	$F(2,36)$	p	$F(4,36)$	p
HI 3	22.06	0.015	208.28	0.003	89.5	0.011	1.37	0.359	12.02	0.042	12.60	0.051	0.83	0.482
HI 4	27.96	0.028	0.45	0.578	385.9	0.003	6.81	0.069	1.45	0.35	2.49	0.251	1.78	0.290
HI 6	21.24	0.036	35.35	0.014	45.0	0.008	31.48	0.008	28.38	0.032	31.8	0.010	15.99	0.013
HI 7	116.8	0.001	314.3	0.001	1183	0.001	52.18	0.006	120.11	0.002	103.9	0.008	14.62	0.034

TABLE VI. Summary of conditions for which performance was not significantly better than chance, as indicated by x. Only data for shaped (SH) and non-shaped (N-SH) stimuli are shown. A dash indicates that a condition was not tested for that subject.

		NH		HI 1	HI 2	HI 3		HI 4		HI 5	HI 6		HI 7	
*F*0	*N*	SH	N-SH	SH	SH	SH	N-SH	SH	N-SH	SH	SH	N-SH	SH	N-SH
100	7							x						
100	11			x	x	x		x		x	x		x	
100	18			x	x	x		x		x	x		x	
200	7							x						
200	11			x	x	x				x				
200	18				x	x	x	x		x				x
400	7				x					x				
400	11			x	x	x				x				
400	18	x		---	---	---	---	---	---	---	x		x	

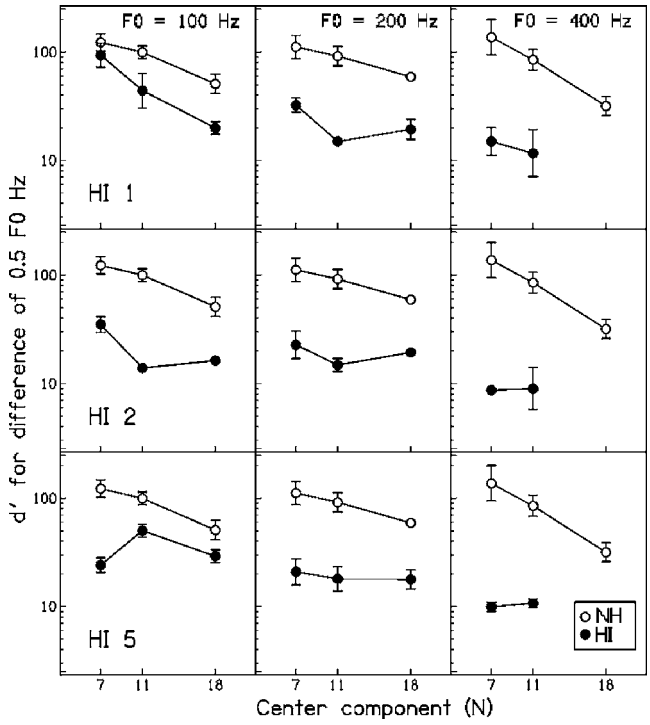


FIG. 6. d' values for $F0$ discrimination by normal-hearing and hearing-impaired subjects. Individual data for three hearing-impaired subjects are plotted (HI 1, HI 2, and HI 5, filled symbols), and average data for four normal-hearing subjects are shown in each frame for comparison (open symbols). Error bars indicate ±1 standard deviation of the mean.

hearing-impaired subjects. The hearing-impaired subjects were all able to perform the task, and could complete the adaptive procedure consistently. However, the normal-hearing subjects performed better than all of the hearing-impaired subjects tested. For the normal-hearing subjects, discrimination worsened as N increased, for all $F0$'s. The effect of N on performance was smaller for the hearing-impaired subjects, although it was marked for $F0=100$ Hz.

V. DISCUSSION

A. Data for normal-hearing subjects

The data for normal-hearing subjects for the shaped stimuli showed that performance worsened with increasing N. This is broadly consistent with the finding of Moore and Moore (2003b) that the pitch shift produced by a given frequency shift of the components decreased with increasing N, since one would expect that performance in our task would depend on the magnitude of the pitch shift produced by the frequency shift. However, Moore and Moore (2003b) found no pitch shift when shaped stimuli with $N=16$ were used, a result which they attributed to an inability to use temporal fine structure information when only very high harmonics were present. In the present study, performance for $N=18$ was poor, but was significantly better than chance when $F0=100$ and 200 Hz. There are a number of possible explanations for this apparent discrepancy.

(1) Subjects may have used remaining spectral (excitation-pattern) cues for discrimination for conditions when $F0=100$ and 200 Hz. However, this seems unlikely, given the small differences in the excitation patterns of the harmonic and frequency-shifted shaped complexes. Also, an explanation based on excitation-pattern cues does not account for why performance was not significantly better than chance for $F0=400$ Hz, where the maximum difference in excitation level was similar to when $F0=100$ and 200 Hz (see Table I).

(2) Although the envelope repetition rate was the same for the harmonic and frequency-shifted complexes, the shape of the envelope was somewhat different. It is possible that the normally hearing subjects used the change in envelope shape to discriminate the harmonic and frequency-shifted complexes. Such a change in shape would not have caused a change in low pitch, which would account for why Moore and Moore (2003b) found no pitch shift. This explanation can also account for the finding that performance was not significantly above chance for $F0=400$ Hz; it seems likely that discrimination of changes in envelope shape would be very poor for such a high $F0$, as sensitivity to modulation decreases for rates above about 120 Hz (Kohlrausch *et al.*, 2000) and discrimination of envelope shape may require detection of higher harmonics of the envelope repetition rate (Dau *et al.*, 1997). The possibility that the subjects here were able to use envelope cues is addressed in a supplementary experiment, which is described later.

(3) For complexes containing only high-frequency harmonics, temporal fine structure information may contribute to sound quality but not to low pitch. Subjects may have discriminated harmonic and frequency-shifted complexes on the basis of differences in timbre. If discrimination was based on a limited ability to use temporal fine structure information, chance performance would be expected when $F0=400$ Hz and $N=18$, as the components of this complex fall entirely into the frequency region where it is believed that phase locking is absent. The results were consistent with this interpretation.

B. Data for hearing-impaired subjects

For the hearing-impaired subjects, discrimination of shaped stimuli with $N=11$ or 18 was very poor whenever the stimuli fell in a frequency region where the hearing loss was 30 dB or more, suggesting that they could make very little use of temporal fine structure information to discriminate harmonic and frequency-shifted complexes. This is an important result, as other studies have presented data that indirectly imply an inability of subjects with cochlear hearing loss to make use of temporal fine structure information, but this has not previously been shown directly (Buss *et al.*, 2004; Lacher-Fougère and Demany, 2005; Lorenzi *et al.*, 2006; Moore and Moore, 2003a; Moore *et al.*, 2006a). Note that the poor performance of the hearing-impaired subjects for shaped stimuli with $N=11$ or 18 confirms that these subjects were not able to use excitation-pattern cues to perform the task for these conditions.

Some subjects showed above-chance performance for shaped stimuli when $N=7$. Although this may reflect a limited ability to use temporal fine structure information for the

frequencies in question, discrimination could also have been based on comparison of the frequencies of individual components, as, despite the poor frequency selectivity seen in hearing-impaired listeners, components with low harmonic numbers may have been resolved for some subjects (Moore *et al.*, 2006a).

Subjects HI 6 and HI 7 performed better than the other hearing-impaired subjects with shaped stimuli for some conditions. This can be attributed to the reduced severity of their hearing impairments. Their audiometric thresholds were within the normal range at high frequencies, and the varying deficits seen across frequency reflect this. For both subjects, performance was worse when $F0=100$ Hz, in which case the components fell into the frequency region where the hearing loss was greatest. Performance in conditions when $F0=400$ Hz was normal for both subjects, as would be predicted by their near-normal audiometric thresholds for high frequencies (see Fig. 2).

The significant interactions between factors highlighted by the ANOVAs may result from the differences across hearing-impaired subjects in the patterns of the audiometric thresholds. When components fell into frequency regions over which audiometric thresholds were nearly normal, performance was relatively good, especially for the conditions where spectral cues could be used (for the non-shaped stimuli and for the shaped stimuli when $N=7$). When components fell into a region of significant hearing loss, performance based on spectral cues was poorer as the auditory filters typically broaden with increasing hearing loss (Glasberg and Moore, 1986). Subjects HI 6 and HI 7 appeared to be able to use temporal fine structure information for some $F0$'s, which also led to significant interactions between factors, as they showed a pattern of results similar to those of normal-hearing subjects for some conditions, and a pattern of results similar to those of hearing-impaired subjects for others.

It is unlikely that the poor performance of the hearing-impaired subjects for shaped stimuli can be explained by poor understanding of the task. Performance was markedly better for the non-shaped and $F0$-discrim stimuli than for the shaped stimuli, even though subjects were not informed of the stimulus type being presented. The good performance of HI 6 and HI 7 for shaped stimuli in some conditions also indicates that, for these subjects at least, poor performance cannot be attributed to cognitive factors.

The poor performance of the three hearing-impaired subjects tested in the $F0$-discrim task is consistent with previous data (Arehart, 1994; Hoekstra and Ritsma, 1977; Moore and Peters, 1992; Moore *et al.*, 2006a). The complex tones could have been discriminated using envelope, temporal fine structure, or spectral information, or any combination of these. Poor performance could, in part, be due to poor frequency selectivity, which would reduce the number of resolved harmonics that could be directly compared. However, the results are also consistent with a loss of ability to use temporal fine structure information. $F0$ DLs for complexes with $N=11$ and 18 were similar (see Fig. 6), suggesting that envelope cues alone may have been used for both values of N, as proposed by Moore and Moore (2003a).

For the conditions of our experiment for which the components were unlikely to be resolved ($N=11$ and 18), the components always fell in frequency regions above about 900 Hz. Thus, our results do not rule out the possibility that subjects with cochlear hearing loss are able to use temporal fine structure information at lower frequencies. Indeed, there are data indicating that hearing-impaired subjects have at least some ability to use temporal fine structure cues for lateralization (Lacher-Fougère and Demany, 2005); for a review, see Moore (1998). It may be that hearing impairment has a greater effect on the ability to use temporal fine structure information at high than at low frequencies, perhaps because the precision of phase locking is reduced for frequencies above about 1000 Hz, even in the normal auditory system (Johnson, 1980; Palmer and Russell, 1986).

Up to this point, it has been assumed that normal-hearing listeners used temporal fine structure cues to discriminate harmonic and frequency-shifted tones when components were unresolved, and that a deficit in temporal fine structure processing was responsible for the poor performance by hearing-impaired subjects in these conditions. However, as previously mentioned, it is possible that envelope shape may have been used as a cue, especially when $N=18$. To investigate the extent to which this cue was used to discriminate shaped stimuli, an additional condition was tested.

VI. DISCRIMINATION OF HARMONIC AND FREQUENCY-SHIFTED TONES WITH SPECTRAL SHAPING AND COMPONENTS ADDED IN RANDOM PHASE

A. Rationale

In experiment 1 all components were added with the same starting phase (sine phase), so harmonic and frequency-shifted complexes within a trial would have had somewhat different envelope shapes. This could have contributed to discrimination ability for the shaped stimuli, particularly when $N=18$. To assess the possible role of this cue, the discrimination of spectrally shaped stimuli was measured when the components were added with random starting phase (referred to as shaped-RP stimuli). The starting phase varied across stimuli within a trial, so each stimulus had a different envelope shape, but only one (the frequency-shifted tone) had a different temporal fine structure.

B. Method

Four normal-hearing subjects were tested. Two had previously been tested with shaped stimuli (NH 1 and NH 5), and two had no previous experience of the task (NH 11 and NH 12). Three of these four subjects had musical training (NH 1, NH 11, and NH 12) and two had previous experience of psychoacoustic experiments (NH 5 and NH 12). Subjects were trained for 1 hour prior to data collection, after which performance appeared to be stable.

The shaped-RP stimuli were created in the same way as the shaped stimuli, except that components were added with randomly chosen starting phases (uniform distribution between 0 and 360°). Harmonic tones were synthesized prior to

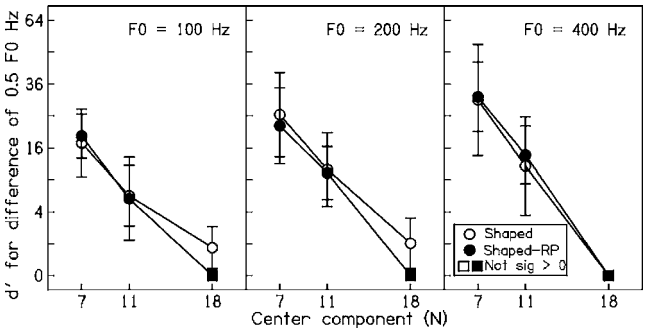


FIG. 7. The filled symbols show d' values for discrimination by normal-hearing subjects of harmonic and frequency-shifted tones with components added with random starting phases (shaped-RP stimuli). The open symbols show d' values for discrimination of shaped stimuli by normal-hearing subjects, for comparison (data from the main experiment). The square symbols indicate d' values that are not significantly different from zero. These points are plotted at zero.

the start of each run, to reduce the delay between trials. Twenty harmonic complexes were synthesized, each with a different random selection of starting phases. A trial was made up of two harmonic complexes randomly selected from these pregenerated complexes, plus a frequency-shifted complex that was synthesized "freshly" for each trial.

The same procedure was used as for the shaped stimuli. Subjects were instructed to choose the interval containing the tone that sounded the most different. Subjects were advised that, for most conditions, the correct interval would have a different pitch than the other two, and if no pitch shift could be heard, they should use the feedback lights to try and identify the quality of the sound that they should use to identify the correct interval.

All subjects were unable to complete the adaptive procedure for all $F0$'s when $N=18$, so a nonadaptive procedure was used for these conditions (see the main Methods section for details of the adaptive and nonadaptive procedures).

C. Results

Mean d' values for the shaped-RP stimuli are shown in Fig. 7, with mean d' values for the shaped stimuli replotted for comparison. For all $F0$'s, performance was very similar for the two stimulus types when $N=7$ and 11. For $N=18$ and $F0=100$ and 200 Hz, performance was worse in the shaped-RP condition than in the shaped condition and, using the same criteria as earlier, was not significantly better than chance. However, as described earlier, performance in the shaped condition with $N=18$ and $F0=100$ and 200 Hz was significantly better than chance. Therefore, the randomization of phase had a deleterious effect when $N=18$ and $F0=100$ and 200 Hz. Performance was not significantly better than chance in the condition when $F0=400$ Hz and $N=18$ for either shaped or shaped-RP stimuli.

D. Discussion

The absence of a significant difference between performance for the shaped and shaped-RP stimuli when $N=7$ and 11 suggests that envelope shape was not an important cue for the shaped stimuli for these values of N. It seems likely that no components were resolved for $N=11$ (see Fig. 3). Hence, these results suggest that, for the normal-hearing subjects, temporal fine structure cues allowed better performance than envelope shape cues, so the latter were "redundant." Similarly, the absence of a significant difference between the performance of normal-hearing subjects for shaped and non-shaped stimuli when $N=7$ and 11 indicates that subjects did not use the extra excitation-pattern cue that was available for non-shaped stimuli. In contrast, better performance for non-shaped than for shaped stimuli when $N=18$ suggests that the additional spectral cue allowed better performance for these conditions. Additionally, the better performance for shaped than for shaped-RP stimuli when $N=18$ and $F0=100$ or 200 Hz suggests that subjects used a change in envelope shape to achieve above-chance performance in these conditions for shaped stimuli. This is consistent with the idea that use of temporal fine structure information for $N=18$ was either limited or absent, as suggested by Moore and Moore (2003b).

It is of interest that the normally hearing subjects were able to use envelope-shape cues to discriminate the shaped stimuli when $N=18$ and $F0=100$ or 200 Hz, while the hearing-impaired subjects apparently were not able to use these cues, since they mostly performed close to chance for these conditions. This suggests that the hearing-impaired subjects had some deficit in the ability to process envelope cues, perhaps because the cues were subtle. Previous work has suggested that hearing-impaired subjects have a near-normal ability to detect amplitude modulation (Bacon and Gleitman, 1992; Moore *et al.*, 1992) and to discriminate changes in amplitude modulation rate (Grant, 1998), but they may show deficits in more complex tasks requiring the use of envelope cues (Lorenzi *et al.*, 1997; Sek and Moore, 2006). In any case, the results suggest that, for the shaped stimuli with $N=11$, the normally hearing subjects mainly used temporal fine structure cues to perform the task, whereas the hearing-impaired subjects were unable to use these cues.

VII. CONCLUSIONS

(1) Normal-hearing subjects appear to be able to use temporal fine structure information to discriminate inharmonic and frequency-shifted complexes with components that are unresolved, but not too high relative to the nominal $F0$.

(2) For components with frequencies that are high with respect to the spacing between components, normal-hearing subjects appear to be unable to access temporal fine structure information, consistent with the conclusions of Moore and Moore (2003b).

(3) The results suggest that hearing-impaired subjects with moderate cochlear hearing loss have very little or no ability to use temporal fine structure cues to discriminate harmonic and frequency-shifted complex tones.

ACKNOWLEDGMENTS

This work was supported by the MRC (UK). We thank Brian Glasberg for his assistance with this work and Neal Viemeister for helpful discussions. We also thank Andrew

Oxenham, Christophe Micheyl, and one anonymous reviewer for helpful comments on an earlier version of this paper.

Arehart, K. H. (**1994**). "Effects of harmonic content on complex-tone fundamental-frequency discrimination in hearing-impaired listeners," J. Acoust. Soc. Am. **95**, 3574–3585.

Bacon, S. P., and Gleitman, R. M. (**1992**). "Modulation detection in subjects with relatively flat hearing losses," J. Speech Hear. Res. **35**, 642–653.

Bernstein, J. G., and Oxenham, A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?," J. Acoust. Soc. Am. **113**, 3323–3334.

Bernstein, J. G., and Oxenham, A. J. (**2005**). "An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination," J. Acoust. Soc. Am. **117**, 3816–3831.

Bernstein, J. G., and Oxenham, A. J. (**2006**). "The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss," J. Acoust. Soc. Am. **120**, 3929–3945.

Buss, E., Hall, J. W., 3rd, and Grose, J. H. (**2004**). "Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss," Ear Hear. **25**, 242–250.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrowband carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

de Boer, E. (**1956**). "Pitch of inharmonic signals," Nature (London) **178**, 535–536.

Glasberg, B. R., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Goldstein, J. L. (**1973**). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. **54**, 1496–1516.

Grant, K. W. (**1998**). "Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 1051–1060.

Green, D. M., and Swets, J. A. (**1974**). *Signal Detection Theory and Psychophysics* (Krieger, New York).

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

Harrison, R. V., and Evans, E. F. (**1979**). "Some aspects of temporal coding by single cochlear fibres from regions of cochlear hair cell degeneration in the guinea pig," Arch. Otolaryngol. **224**, 71–78.

Hartmann, W. M., and Doty, S. L. (**1996**). "On the pitches of the components of a complex tone," J. Acoust. Soc. Am. **99**, 567–578.

Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001**). "Evalulating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve," Neural Comput. **13**, 3373–2316.

Hoekstra, A., and Ritsma, R. J. (**1977**). "Perceptive hearing loss and frequency selectivity," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).

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.

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.

Keppel, G. (**1991**). *Design and Analysis: A Researcher's Handbook* (Prentice Hall, Upper Saddle River, NJ).

Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," J. Acoust. Soc. Am. **108**, 723–734.

Lacher-Fougère, S., and Demany, L. (**1998**). "Modulation detection by normal and hearing-impaired listeners," Audiology **37**, 109–121.

Lacher-Fougère, S., and Demany, L. (**2005**). "Consequences of cochlear damage for the detection of interaural phase differences," J. Acoust. Soc. Am. **118**, 2519–2526.

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

Liberman, M. C., and Kiang, N. Y. S. (**1978**). "Acoustic trauma in cats: Cochlear pathology and auditory-nerve activity," Acta Oto-Laryngol., Suppl. **358**, 1–63.

Loeb, G. E., White, M. W., and Merzenich, M. M. (**1983**). "Spatial cross correlation: A proposed mechanism for acoustic pitch perception," Biol. Cybern. **47**, 149–163.

Lorenzi, C., Micheyl, C., Berthommier, F., and Portalier, S. (**1997**). "Modulation masking in listeners with sensorineural hearing loss," J. Speech Lang. Hear. Res. **40**, 200–207.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.

Moore, B. C. J. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**, 610–619.

Moore, B. C. J. (**1977**). "Effects of relative phase of the components on the pitch of three-component complex tones," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).

Moore, B. C. J. (**1982**). *An Introduction to the Psychology of Hearing*, 2nd ed. (Academic, London).

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, San Diego).

Moore, B. C. J., and Carlyon, R. P. (**2005**). "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in *Pitch Perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer, New York).

Moore, B. C. J., and Glasberg, B. R. (**1988**). "Effects of the relative phase of the components on the pitch discrimination of complex tones by subjects with unilateral and bilateral cochlear impairments," in *Basic Issues in Hearing*, edited by H. Duifhuis, H. Wit, and J. Horst (Academic, London).

Moore, B. C. J., and Moore, G. A. (**2003a**). "Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects," Hear. Res. **182**, 153–163.

Moore, B. C. J., and Ohgushi, K. (**1993**). "Audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **93**, 452–461.

Moore, B. C. J., and Peters, R. W. (**1992**). "Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity," J. Acoust. Soc. Am. **91**, 2881–2893.

Moore, B. C. J., and Skrodzka, E. (**2002**). "Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation," J. Acoust. Soc. Am. **111**, 327–335.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness and partial loudness," J. Audio Eng. Soc. **45**, 224–240.

Moore, B. C. J., Glasberg, B. R., and Hopkins, K. (**2006a**). "Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure," Hear. Res. **222**, 16–27.

Moore, B. C. J., Glasberg, B. R., and Shailer, M. J. (**1984**). "Frequency and intensity difference limens for harmonics within complex tones," J. Acoust. Soc. Am. **75**, 550–561.

Moore, B. C. J., Glasberg, B. R., and Stone, M. A. (**2004**). "New version of the TEN test with calibrations in dB HL," Ear Hear. **25**, 478–487.

Moore, B. C. J., Oldfield, S. R., and Dooley, G. (**1989**). "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," J. Acoust. Soc. Am. **85**, 820–836.

Moore, B. C. J., Shailer, M. J., and Schooneveldt, G. P. (**1992**). "Temporal modulation transfer functions for band-limited noise in subjects with cochlear hearing loss," Br. J. Audiol. **26**, 229–237.

Moore, B. C. J., Glasberg, B. R., Flanagan, H. J., and Adams, J. (**2006b**). "Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure," J. Acoust. Soc. Am. **119**, 480–490.

Moore, B. C. J., Glasberg, B. R., Low, K.-E., Cope, T., and Cope, W. (**2006c**). "Effects of level and frequency on the audibility of partials in inharmonic complex tones," J. Acoust. Soc. Am. **120**, 934–944.

Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcántara, J. I. (**2000**). "A test for the diagnosis of dead regions in the cochlea," Br. J. Audiol. **34**, 205–224.

Moore, G. A., and Moore, B. C. J. (**2003b**). "Perception of the low pitch of frequency-shifted complexes," J. Acoust. Soc. Am. **113**, 977–985.

Nelson, D. A., and Freyman, R. L. (**1986**). "Psychometric functions for frequency discrimination from listeners with sensorineural hearing loss,"

J. Acoust. Soc. Am. **79**, 799–805.
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.
Pick, G., Evans, E. F., and Wilson, J. P. (**1977**). "Frequency resolution in patients with hearing loss of cochlear origin," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).
Plack, C. J., and Oxenham, A. J. (**2005**). "The psychophysics of pitch," in *Pitch Perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer, New York).
Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.
Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1967**). "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey," J. Neurophysiol. **30**, 769–793.
Rosen, S. (**1987**). "Phase and the hearing impaired," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Martinus Nijhoff, Dordrecht).
Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). "Pitch of the residue," J. Acoust. Soc. Am. **34**, 1418–1424.
Sek, A., and Moore, B. C. (**2006**). "Perception of amplitude modulation by hearing-impaired listeners: The audibility of component modulation and detection of phase change in three-component modulators," J. Acoust. Soc. Am. **119**, 507–514.
Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.
Shamma, S. A. (**1985**). "Speech processing in the auditory system. II. Lateral inhibition and the central processing of speech evoked activity in the auditory nerve," J. Acoust. Soc. Am. **78**, 1622–1632.
Terhardt, E. (**1974**). "Pitch, consonance, and harmony," J. Acoust. Soc. Am. **55**, 1061–1069.
van Noorden, L. P. A. S. (**1982**). "Two-channel pitch perception," in *Music, Mind and Brain*, edited by M. Clynes (Plenum, New York).
von Békésy, G. (**1960**). *Experiments in Hearing* (McGraw-Hill, New York).
Woolf, N. K., Ryan, A. F., and Bone, R. C. (**1981**). "Neural phase-locking properties in the absence of outer hair cells," Hear. Res. **4**, 335–346.
Zwislocki, J. J., and Nguyen, N. (**1999**). "Place code for pitch: A necessary revision," Acta Oto-Laryngol. **119**, 140–145.

Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences[a)]

Ginger S. Stickney[b)]
Hearing Instrument Consultants, 3090 Bristol Street, Suite 150, Costa Mesa, California 92626

Peter F. Assmann
School of Behavioral and Brain Sciences, University of Texas at Dallas, GR41, Box 830688, Richardson, Texas 75083

Janice Chang
Hearing and Speech Research Laboratory, University of California, Irvine, 364 Medical Surgery II, Irvine, California 92697-1275

Fan-Gang Zeng[c)]
Department of Otolaryngology—Head and Neck Surgery, University of California, Irvine, 364 Medical Surgery II, Irvine, California 92697-1275

(Received 15 October 2005; revised 17 May 2007; accepted 26 May 2007)

Speech perception in the presence of another competing voice is one of the most challenging tasks for cochlear implant users. Several studies have shown that (1) the fundamental frequency (F0) is a useful cue for segregating competing speech sounds and (2) the F0 is better represented by the temporal fine structure than by the temporal envelope. However, current cochlear implant speech processing algorithms emphasize temporal envelope information and discard the temporal fine structure. In this study, speech recognition was measured as a function of the F0 separation of the target and competing sentence in normal-hearing and cochlear implant listeners. For the normal-hearing listeners, the combined sentences were processed through either a standard implant simulation or a new algorithm which additionally extracts a slowed-down version of the temporal fine structure (called Frequency-Amplitude-Modulation-Encoding). The results showed no benefit of increasing F0 separation for the cochlear implant or simulation groups. In contrast, the new algorithm resulted in gradual improvements with increasing F0 separation, similar to that found with unprocessed sentences. These results emphasize the importance of temporal fine structure for speech perception and demonstrate a potential remedy for difficulty in the perceptual segregation of competing speech sounds. [DOI: 10.1121/1.2750159]

PACS number(s): 43.66.Ts, 43.71.Ky, 43.71.Bp, 43.66.Hg [KWG] Pages: 1069–1078

I. INTRODUCTION

The fundamental frequency (F0) of voiced speech, which determines the pitch of the voice, can be a useful cue for segregating competing speech sounds (Bregman, 1990). When two voices compete, it is easier to hear what one voice is saying if the competing voice has a different pitch, or occupies a different F0 range (Bird and Darwin, 1998; Brokx and Nooteboom, 1982; Darwin and Hukin, 2000). While normal-hearing listeners are capable of using differences in the pitch of the voice to improve their performance with competing speech sounds, there are many cochlear implant users who show no benefit when two competing sentences are spoken by talkers of different genders (Stickney *et al.*, 2004).

At present, speech coding strategies used by most cochlear implants encode only the slowly varying amplitude modulations of the speech wave form (the temporal envelope), while the temporal fine structure is discarded. Therefore, the F0 can only be conveyed by the temporal modulations. Although pitch can be conveyed by the temporal envelope (Burns and Viemeister, 1976), sounds that include the temporal fine structure evoke a stronger pitch percept than sounds that preserve only the temporal envelope (Oxenham *et al.*, 2004; Smith *et al.*, 2002). The lack of explicit encoding of the temporal fine structure is one reason that speech perception in the presence of other competing voices is such a challenging task for cochlear implant users (Qin and Oxenham, 2003; Stickney *et al.*, 2004; Zeng *et al.*, 2004).

Stickney *et al.* (2004) presented normal-hearing and cochlear implant listeners with competing sentences spoken by the same or different talkers. The normal-hearing listeners were presented with natural speech or a cochlear implant simulation. The simulation transmits amplitude modulations from a series of frequency bands, and within each frequency band, the amplitude modulation is applied to a white noise carrier (Shannon *et al.*, 1995). Stickney *et al.* demonstrated that normal-hearing listeners presented with an implant

[a)]Portions of this work were presented at the 148th Meeting of the Acoustical Society of America (2004) in San Diego, CA.
[b)]Electronic mail: gsstickney@yahoo.com;
[c)]Electronic mail: fzeng@uci.edu

simulation and cochlear implant users had as much difficulty when the competing talker was a female voice as when the talker was the same male voice as the target. In contrast, normal-hearing listeners presented with natural speech did not encounter these difficulties and instead showed an improvement of 50 percentage points at a 0 dB signal to noise ratio (SNR) with the female masker compared to the same male masker. These results demonstrate that with implant simulations that extract only the envelope information within each frequency band, listeners appear to be unable to take advantage of differences in voice F0 to segregate competing speech sounds.

Some of the earlier multichannel devices, such as Cochlear Corporation's Nucleus 22-electrode device, used the F0 to modulate a pulsatile carrier during voiced speech and spectral information from one or two formants to selectively stimulate a subset of electrodes with the greatest energy. The direct coding of F0 was eventually abandoned with the introduction of a new speech processing algorithm (Continuous Interleaved Sampling) that stimulated the electrodes sequentially to avoid potential current field interactions (Wilson *et al.*, 1991). In this algorithm, F0 information could be inherently conveyed by the temporal envelope, provided the carrier pulse rate and envelope cutoff frequency were sufficiently high (Geurts and Wouters, 2001). To better represent the instantaneous temporal envelope, Rubenstein *et al.* (1999) have recommended increasing the rate at which amplitude modulation information is cycled through the electrodes (i.e., a stimulation rate greater than 2000 Hz), analogous to increasing the sampling rate. They suggest that high-rate stimulation also has the potential to reintroduce more natural stochastic effects in auditory nerve responses (see Kiang *et al.*, 1965) that would allow for a greater dynamic range. The combination of higher rates and stochastic resonance may improve speech perception in implant users as a consequence of enhancing the representation of the temporal envelope. The temporal envelope however can only convey a weak representation of pitch.

How pitch information can be better represented in cochlear implant processing by additionally modifying the carrier frequency has therefore been of great interest in recent years. Green *et al.* (2004) compared the pitch labeling of processed diphthongal glides in normal-hearing and cochlear implant listeners. In one condition, the processing of stimuli for the normal-hearing listeners involved an implant simulation with a noise carrier. In a second condition, the carrier (a sawtooth-shaped wave form) included the periodicity of the vowel. They found that the carrier which additionally coded the periodicity information improved pitch labeling for both normal-hearing listeners presented with the modified implant simulation and cochlear implant users. While pitch perception was improved with the modified processing, a more recent study by the same group (Green *et al.*, 2005) found that formant frequency discrimination and vowel recognition were adversely affected, perhaps because the modified processing disrupted spectral cues.

Although the benefits of a periodic carrier were significant with the pitch labeling task, the amount of improvement with the addition of periodicity information was small. It could be that a pitch labeling task was not sufficient for demonstrating the true benefits provided by this additional cue. Speech perception tasks relying heavily on pitch information might have demonstrated a larger effect. In a study by Lan *et al.* (2004), an implant simulation was similarly modified to include F0 information for voiced segments of Mandarin Chinese tones. They found that the pitch patterns of four Mandarin tones were more accurately identified with the modified than traditional processing. Lan *et al.* also noted improved performance for phonemes, words, and sentences. The present study examines another type of modification to the cochlear implant simulation that codes F0 indirectly by extracting the temporal fine structure of sound.

A new signal processing algorithm has recently been developed to code the temporal fine structure by means of a frequency-modulated (FM) carrier. Speech recognition performance in the presence of a competing talker was examined using the new strategy (Frequency-Amplitude-Modulation-Encoding, FAME) and compared with an implant simulation using a tone-excited vocoder (Nie *et al.*, 2005; Stickney *et al.*, 2005). Details of the new algorithm are explained in the following (see Sec. II). In both studies, the target and masker sentences were spoken by different talkers, allowing for differences in voice pitch to be captured by FM. They found that the addition of the FM significantly improved performance relative to the standard simulation. With speech maskers, performance dropped by as much as 18 percentage points with the standard simulation relative to performance in quiet. In contrast, there was no significant drop in performance when a competing talker was added for the FAME processing that included FM. This result suggests that the listeners had access to additional cues with FM, most likely F0 information, which helped them segregate the two competing talkers.

The present study tests the hypothesis that FM, added to an implant simulation, can convey sufficient F0 information such that when the mean F0 of the masker is shifted relative to the target, there will be an improvement in speech recognition. Speech recognition was measured as a function of the F0 separation of the target and masking sentence (both of which were spoken by the same talker) over several semitones in normal-hearing and cochlear implant listeners. The normal-hearing listeners were presented with natural speech, the standard cochlear implant simulation, or the FAME processing. Because FAME codes both temporal envelope and temporal fine structure cues, F0 cues should be transmitted more effectively with FAME compared to the standard implant simulation to assist in perceptually segregating the competing sentences.

II. FREQUENCY-AMPLITUDE-MODULATION-ENCODING (FAME)

The new strategy (FAME) separately extracts the slowly varying amplitude (AM) and frequency (FM) modulations within each frequency band (see Fig. 1). The FM codes the temporal fine structure of the speech wave form, whereas the AM separately codes the temporal envelope. The instantaneous amplitude of the FM carrier frequency is determined from the temporal envelope in the corresponding band. The

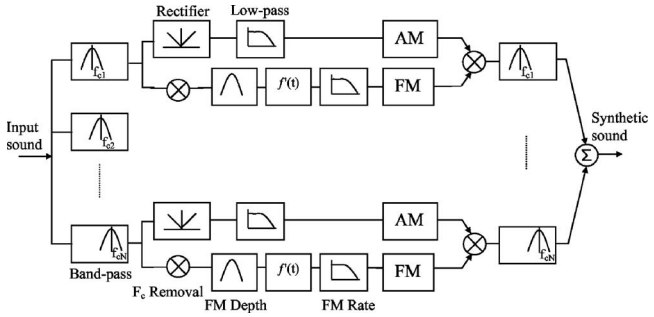


FIG. 1. Signal processing diagram for a 4-channel FAME processor.

signal is first divided into n narrowbands. The narrowband signals are then transmitted to separate AM and FM extraction pathways. The AM pathway involves full-wave rectification followed by low-pass filtering at 500 Hz to obtain the slowly varying envelope. The FM pathway involves removal of each narrowband's center frequency through phase-orthogonal demodulators (Flanagan and Golden, 1966) followed by low-pass filtering of the FM bandwidth (with a cutoff of 500 Hz) and rate (with a cutoff of 400 Hz). Limiting the FM rate is important since the eventual goal of a speech coding strategy, such as FAME, is to provide FM information that can be perceived by the majority of cochlear implant users.

As demonstrated by Chen and Zeng (2004), cochlear implant users' ability to detect a change in pitch for a frequency sweep or sinusoidal FM decreased as the standard frequency or modulation rate was increased. The delay between the AM and FM extraction pathways is adjusted prior to the combination of these two components within each subband and the signals are further bandpass filtered to remove frequency components introduced by AM and FM that fall outside of the original analysis filter's bandwidth. The waveforms from all bands are then summed to form the synthesized signal that contains the slowly varying AM and FM components.

III. EXPERIMENT

A. Methods

1. Listeners

The subjects were 49 young native English speakers, comprising undergraduate and graduate students. All subjects reported normal hearing. The subjects were recruited at the University of California, Irvine. There were seven subjects for each of the seven processing conditions. Additionally, seven cochlear implant users were recruited (see Table I for subject demographics). Subjects were compensated $10/h for their participation.

2. Test materials

Subjects were presented with IEEE sentences paired with a masker taken from the same set of sentences. All IEEE sentences in this study consisted of a subset of the 72 phonetically balanced lists of 10 sentences. The sentences were obtained from recordings by Hawley *et al.* (1999). The target sentences were spoken by a male voice (mean F0 = 108 Hz) in the presence of a different sentence spoken by the same male voice. The same masker sentence ("A large size in stockings is hard to sell") was presented on each trial to avoid confusion of the target and masker sentences.[1] The target and masker sentence had the same onset, but the masking sentence was always longer in duration. No sentences were repeated.

The stimuli were 70 sentences, with seven F0 conditions of 10 sentences each. Each sentence consisted of five keywords, for a total of 50 keywords per condition. There were six F0 conditions where the F0 contour of the masker sentence was shifted to a higher F0. The seventh condition included a natural speech masker where the F0 was neither estimated nor modified. Unlike most previous studies that have examined the effects of F0 difference on competing sentences using a steady-state (monotone) pitch, the natural F0 contour was preserved in the present study, but shifted upwards by n semitones from the average F0 measured across the entire sentence. A high-quality speech analysis-synthesis system called STRAIGHT (Kawahara, 1997, 1999) was used to estimate the F0 and resynthesize the sentence with a shifted F0. The estimated F0 contour (analyzed in 1 ms frames) was then replaced by one that was shifted up by a fixed amount compared to the average F0 of the original sentence, i.e., in each frame the measured F0 was replaced by $F0_{new} = 2^{n/12} * F0_{original}$, where $n = 0$, 3, 6, 9, 12, or 15 semitones. These conditions were labeled "semi3," "semi6,"…, and "semi15" corresponding to an F0 shift of 3, 6,…, and 15 semitones, respectively. The label semi0 represented the condition where the STRAIGHT algorithm was applied to the masker but the F0 contour was not raised, whereas the label

TABLE I. Subject demographics.

Subject	Age	Implant	Speech strategy	Duration of hearing loss (years)	Duration of deafness (years)	Duration of implant use (years)
CI1	46	Nucleus-22	SPEAK	12	12	11
CI2	69	Nucleus-24	ACE	7	7	6
CI3	70	Nucleus-24	ACE	40	14	3
CI4	61	Nucleus-22	SPEAK	52	14	12
CI5	78	Nucleus-24	CIS	35	13	1
CI6	68	Clarion CII	MPS	22	18	2
CI7	68	Clarion CII	MPS	62	58	5

"natural" was used to represent the condition where the masker was not processed with the STRAIGHT algorithm.

There were seven processing conditions. Conditions where the stimuli were not subjected to cochlear implant processing were labeled as "unprocessed." There were two conditions that used unprocessed speech: (1) unprocessed speech presented at a 0 dB signal-to-noise ratio (SNR) and (2) unprocessed speech presented at a 10 dB SNR. The remaining five conditions used standard implant simulations with the competing sentences at a 10 dB SNR. The SNR was adjusted after the application of the STRAIGHT algorithm. The SNR was calculated by first determining the rms of the unprocessed target and masker sentences (including silent periods), then scaling the masker and target to the same rms, and for the 10 dB SNR conditions, subsequently decreasing the rms of the masker sentence relative to the target sentence.

The implant simulation (AM or AM+FM) was applied after the target and masker were mixed. The AM+FM processing used the same algorithm as in the previous study by Nie *et al.* (2005). The combined target and masker stimuli were first pre-emphasized with a high-pass, first-order Butterworth filter with a cutoff frequency of 1.2 kHz. The sentences were then filtered into 4, 8, or 32 narrowbands using fourth-order elliptic bandpass. The AM and FM extraction was accomplished with fourth-order Bessel filters. The overall processing bandwidth was 80–8800 Hz. A sinusoidal carrier was used for both AM-only and AM+FM conditions. The AM-filter cutoff was set to 500 Hz, while the FM rate and depth were 400 and 500 Hz, respectively. However, for filters with bandwidths $<$500 Hz, the FM depth was set to be the same as the bandwidth of the filter. The implant simulation conditions were: (1) 4-channel AM-only processed speech; (2) 8-channel AM-only processed speech; (3) 32-channel AM-only processed speech; (4) 4-channel AM+FM-processed speech; and (5) 8-channel AM+FM-processed speech. The seven groups of normal-hearing subjects were presented with one of these conditions. The cochlear implant subjects were presented with only the unprocessed speech at a 10 dB SNR. Based on pilot data, the SNR for the implant simulation conditions was changed from 0 to 10 dB SNR. A 0 dB SNR was too difficult for several conditions with the AM-only processing and for the cochlear implant subjects.

3. Procedure

The stimuli were presented monaurally to the right ear through headphones (Sennheiser HAD 200), with subjects seated in an IAC sound booth. The level of the combined target and masker sentence was set to approximately 65 dB SPL, on average (Brüel & Kjær 2260 Investigator sound level meter; Brüel & Kjær Type 4152 artificial ear). After each stimulus was presented, subjects typed their responses using the computer keyboard and were encouraged to guess if unsure. Subjects were given as much time as needed to type their responses and were also given an opportunity to correct their spelling errors. Their responses were scored automatically based on the percentage of target sentence keywords correctly identified. Since all scoring was done with the computer program, no allowance was made for minor spelling errors.

Prior to testing, subjects were presented with two practice sessions. The first practice session presented ten unprocessed sentences in quiet. Subjects were to identify at least 85% of the keywords in order to participate in the study. No subjects were disqualified in this practice session. The second practice session consisted of seven sentences processed in the same manner as the test stimuli, with one sentence for each condition. This portion of the test was designed to simulate the test conditions so the subjects would know what to expect in the actual test session. No score was calculated for this practice session.

In the test session, each subject was presented with seventy sentences. There were ten sentences per condition for each subject. Each subject received a different set of sentences for each condition (digram-balanced Latin square design) to distribute the effects of sentence difficulty across the conditions. For example, subject 1 heard sentences 1–10 in the natural condition, while subject 2 heard the same sentences 1–10 in the condition with the masker's F0 contour shifted by 0 semitones. The order of the ten sentences in each set was randomized, as was the order of the conditions presented to the subject. Each test session lasted for approximately 20 min.

B. Results

1. Unprocessed speech

Figure 2 illustrates the wave form and corresponding F0 contours for the target and masker stimuli presented at a 0 dB SNR. The F0 contours were extracted from the sentences in isolation (i.e. prior to mixing) in 1 ms frames using the F0 estimation algorithm used by the STRAIGHT analysis system developed by Kawahara (1997). Because the stimuli shown here were not further processed with the algorithm that was used to create the cochlear implant simulations, they are referred to as "unprocessed." The panels from top to bottom show increasing differences in F0 between the target and masker sentence. Note that the F0 varies over time and that when the average F0s are the same for the target and masker, there are temporal intervals where they are well separated (i.e. the instantaneous F0 is different for the two sentences). Also, note that as the F0 shift increases, so does the size of this temporal interval where the instantaneous target and masker F0s are well separated.

Figure 3 shows the sentence recognition accuracy for the normal-hearing listeners presented with the unprocessed speech at the two SNRs. The x axis shows the F0 shift and the y axis shows the percentage of keywords correctly identified in that condition. It can be seen that at a SNR of 0 dB, performance improved as the F0 shift increased. In contrast, at a 10 dB SNR, no benefit was observed as performance was at ceiling. A mixed design analysis of variance (ANOVA) was performed with the two unprocessed speech conditions as the between-subjects factor and F0 shift as the within-subjects factor. The normal hearing listeners presented with the unprocessed speech showed higher perfor-

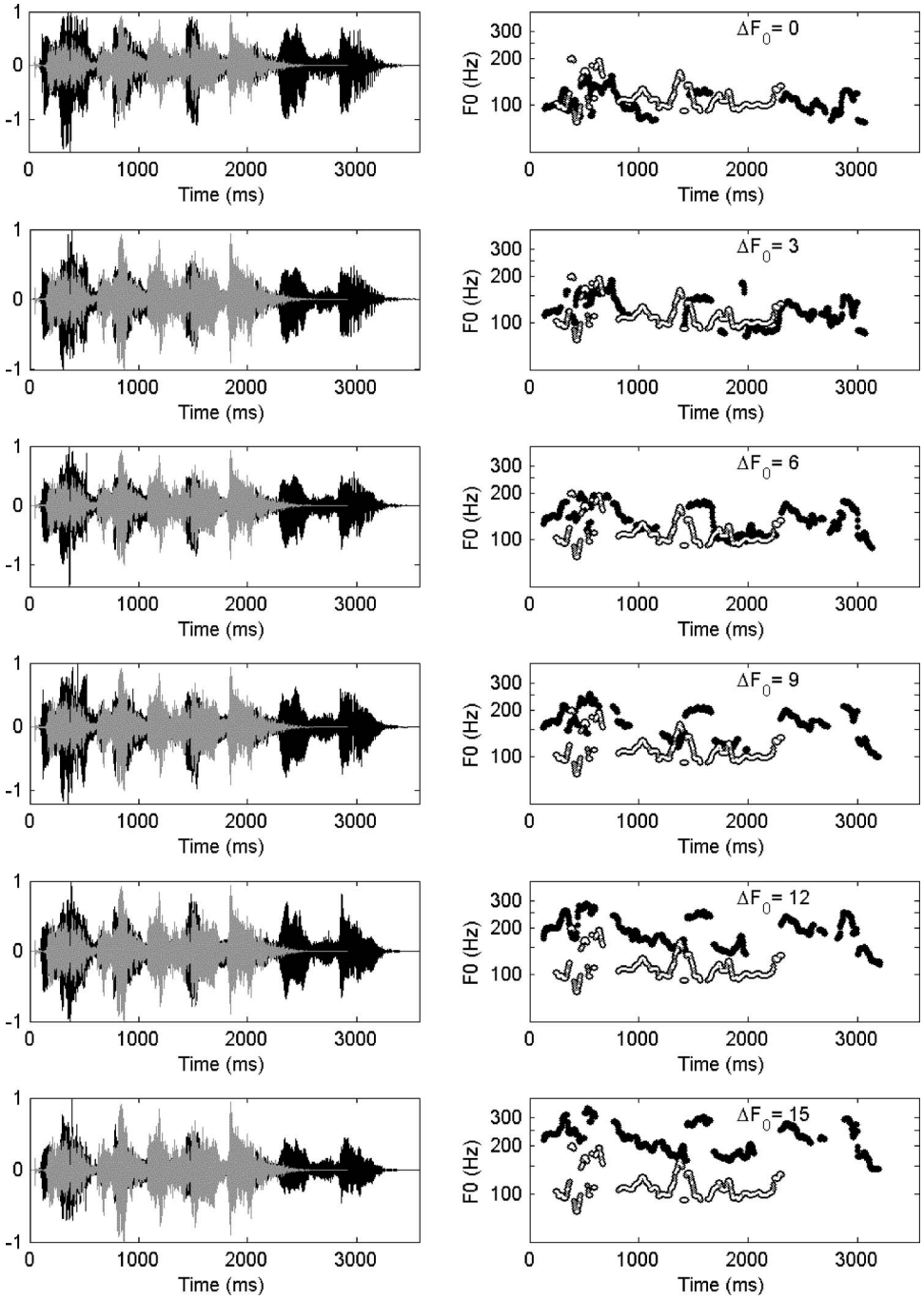


FIG. 2. Unprocessed wave forms and F0 contours for the target sentence and masker sentence. The target sentence is "The sheep were led home by a dog" and the masker sentence is "A large size in stockings is hard to sell." The wave form for the target sentence is shown in light gray and its F0 contour is represented as an unfilled line. The wave form for the masker sentence is shown in black and its F0 contour is represented as a solid black line. The F0 contours were extracted from unmixed signals that were scaled to the same rms and superimposed. The F0 contour for the masking sentence increases in frequency from the top panel to the bottom panel.

mance at the higher SNR of +10 dB [$F(1,12)=20.7, p<0.01$]. Target keyword identification generally improved as the F0 separation was increased from 0 to 6 semitones, but only when the rms level of the target and masker was matched (0 dB SNR). At a 0 dB SNR, performance improved by 12 percentage points from semi0 to semi6. However, this trend did not reach statistical significance when subjected to a Bonferroni adjustment for the six pairwise comparisons.

2. AM-only processed speech and cochlear implant performance

The left panels of Fig. 4 illustrate the estimated F0 contours for the same target and masker sentences as Fig. 2 with AM-only processing, estimated from the sentences prior to mixing but after processing. Notice that the F0 of the target and masker is relatively flat compared to the unprocessed F0 contour in Fig. 2. The sparseness of the F0 contours indicates that F0 is not as well represented in the processed versions, and that there appear to be more F0 estimation errors. What is also noticeable is that there is more overlap between the F0 components of the target and masker with AM-only processing. The reason for this greater overlap was likely due to the lack of explicit encoding of the temporal fine structure in the AM-only processed stimuli.

Figure 5 shows the results for the normal-hearing subject groups presented with 4, 8, or 32 AM-only processed channels. For purposes of comparison, the results from cochlear implant users and the normal-hearing subjects presented with the unprocessed sentences at a 10 dB SNR are

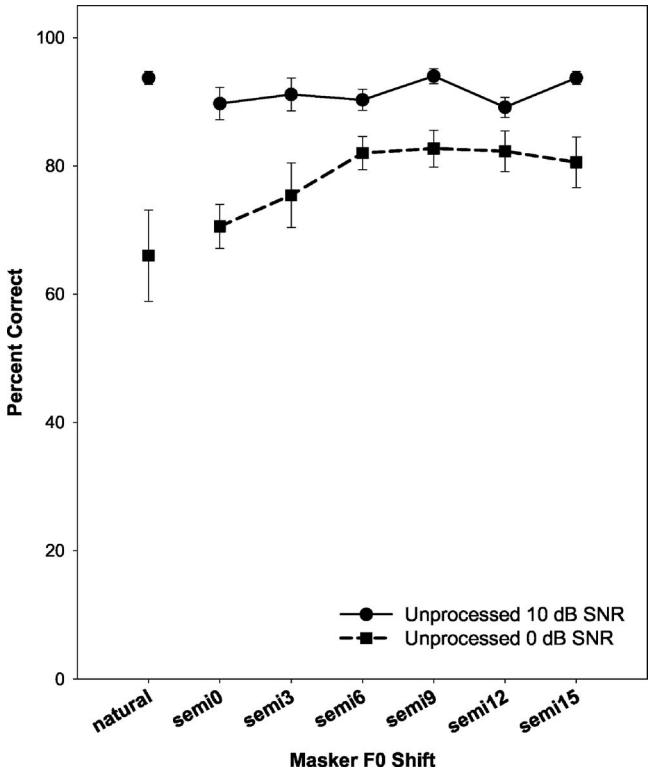


FIG. 3. Results for the normal-hearing subject groups presented with unprocessed speech at either a 0 dB SNR (square with dashed line) or 10 dB SNR (circle with solid line). The *x* axis shows each of the F0 shift conditions. The label "natural" represents the condition where the masker sentence was not processed by the STRAIGHT algorithm. The labels "semi0," "semi3," …, and "semi15" represent the conditions with an F0 shift of 0, 3, …, and 15 semitones, respectively. The error bars represent the standard error of the mean calculated from the scores of the 7 subjects within each group.

included in Fig. 5. What is most interesting about the data shown in Fig. 5 is that even though there is a dramatic difference in speech recognition scores as the number of channels is varied, there is relatively no change in performance as the F0 is shifted. A mixed design ANOVA was performed with the number of channels as the between-subjects factor and the F0 shift conditions as the within-subjects factor. Intelligibility improved dramatically as the number of channels was increased $[F(2,18)=179.3, p<0.001]$. Bonferroni pairwise comparisons showed significant improvements in performance from 4 to 8 and from 8 to 32 channels ($p<0.0125$), but not from 32 channels to the unprocessed speech. In addition, there was no significant difference in performance between the cochlear implant users and the normal-hearing listeners presented with 4-channel AM-only processed speech. Five separate ANOVAs, one for each processing group, were performed to determine whether or not there was a significant effect of F0 shift. The key finding was that for all the normal-hearing groups of subjects presented with AM-only processing and the cochlear implant users, performance as a function of F0 shift did not change. Even with 32 channels of envelope information listeners were not able to take advantage of the F0 difference between the target and masking sentence.

3. AM+FM*-processed speech*

The panels on the right side of Fig. 4 show the F0 contours of the target and masker for the AM+FM-processed speech. In contrast to the results obtained with AM-only processing (shown in the left panels), the AM+FM-processing preserves partial F0 contours. Note that as the F0 of the AM+FM-processed masker is increased, more of the F0 contour of the target sentence is revealed.

The results with AM+FM-processing are shown in Fig. 6. As observed with AM-only processing, speech recognition performance improved with more channels (i.e., the scores were higher with 8 AM+FM channels than with 4 AM +FM channels) $[F(1,12)=50.5, p<0.001]$. A comparison with Fig. 5 also shows that speech recognition performance with AM+FM processing was higher than with AM-only processing with the same number of channels. Last, Fig. 6 shows that speech recognition scores, at least for the 8-channel AM+FM group, tended to improve as the F0 shift was increased.

A mixed design ANOVA was used to compare AM and AM+FM performance. The F0 shift conditions were the within-subjects factor and the type of processing (four groups of subjects: 4-channel AM, 4-channel AM+FM, 8-channel AM, and 8-channel AM+FM) was the between-subjects factor. The results showed a significant effect of processing $[F(3,24)=50.35, p<0.001]$, F0 shift $[F(6,19)=3.12, p<0.05]$, and a significant interaction between the type of processing and the effect of the F0 shift on speech recognition performance $[F(18,54)=4.15, p<0.001]$. A *post-hoc* Scheffé analysis demonstrated significantly higher performance with the 4-channel AM+FM processing compared to the 4-channel AM-only processing, higher performance for the 8-channel AM-only processing compared to the 4-channel AM+FM processing, and the highest performance with 8-channel AM+FM processing ($p<0.05$ for all comparisons). For the 4-channel conditions, speech recognition scores, collapsed across all F0-shift conditions, were 30% for AM+FM and 13% for AM only. Similarly, for the 8-channel conditions, the scores were 57% and 45% for the AM+FM and AM conditions, respectively.

To examine the effect of the F0 shift on speech recognition with AM+FM processing, a separate mixed design ANOVA including only the AM+FM data and both channel conditions (4 and 8 channel) was performed. In contrast to the results obtained with AM-only processing, there was an improvement in performance as the F0 separation was increased $[F(6,7)=6.8, p<0.05]$. The interaction between the two channel conditions and the effect of F0 shift on speech recognition performance approached significance $[F(6,7)=3.1, p=0.08]$. For the 4-channel AM+FM group, there was much variability in the speech recognition scores as the F0 shift increased, obscuring any clear trend. In contrast, speech recognition performance for the 8-channel AM+FM group gradually improved with increases in F0 shift; the amount of improvement relative to the unshifted F0 (i.e., semi0 condition) was as much as 20 percentage points with an F0 shift of 12 semitones.

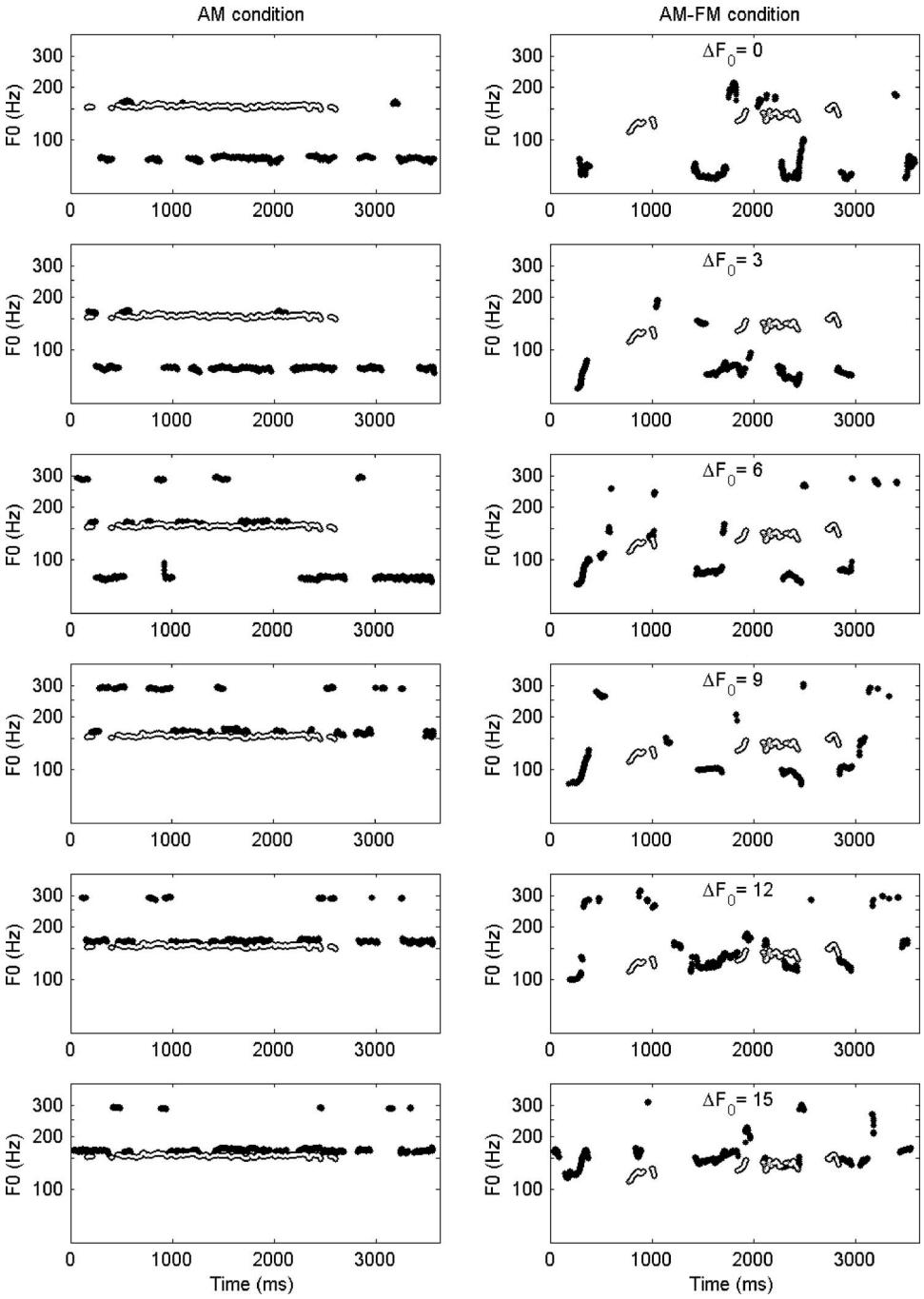


FIG. 4. The 8-channel implant simulation wave forms and F0 contours. The target sentence is "The sheep were led home by a dog" and the masker sentence is "A large size in stockings is hard to sell." The F0 contour of the target sentence is represented as an unfilled line, whereas that of the masker sentence is represented as a solid black line. The F0 contours were extracted from unmixed signals that were scaled to the same rms and superimposed. From the top to the bottom panel the F0 contour for the masker sentence increases.

IV. DISCUSSION

Users of cochlear implants have great difficulty understanding speech in the presence of one or more competing speech sounds. The results of the present study suggest that this may be due, in part, to a lack of F0 information provided by their speech processing algorithm. Several studies have demonstrated that normal-hearing listeners can take advantage of differences in voice characteristics (including the F0) between two competing talkers when presented with natural speech (Bird and Darwin, 1998; Brungart, 2001; Qin and Oxenham, 2003; Stickney *et al.*, 2004). However, when presented with strictly envelope-extracting implant simulations, even with as many as 24 channels, normal-hearing listeners did not benefit from these differences (Qin and Oxenham, 2003). Qin and Oxenham showed that although the 24-channel condition produced similar temporal envelopes as the natural speech, the differences in speech recognition thresholds for a female talker masking a male target sentence in the natural speech condition compared to the 24-channel condition were astounding: a better speech reception threshold of −11.3 dB for natural speech compared to the significantly poorer threshold of 0.6 dB for the 24-channel condition. Furthermore, Stickney *et al.* (2004) demonstrated results from actual cochlear implant users that were quite comparable to the normal-hearing listeners presented with, at most, 8 temporal envelope channels. Statistically there was no significant improvement from using the same male talker as masker and target to using a male talker as target and a female talker as a masker. Together, the results suggest that even with a relatively large number of channels (e.g., 24)

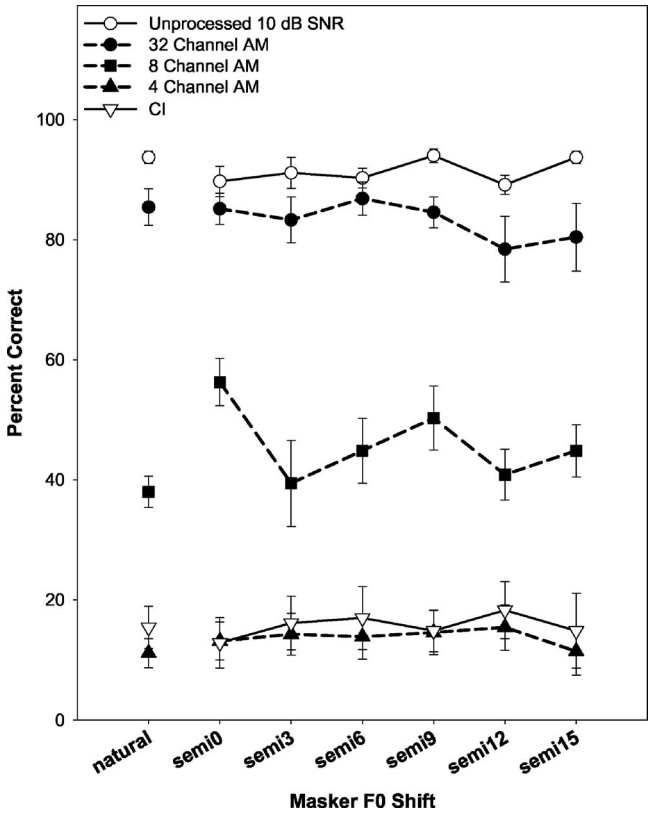


FIG. 5. Results for the normal-hearing subject groups presented with the AM-only processed speech at a 10 dB SNR (closed symbols with dashed lines). Results from the cochlear implant users and normal-hearing listeners presented with unprocessed speech at a 10 dB SNR are included for comparison (open symbols with solid lines). The *x* axis shows each of the F0 shift conditions. The label "natural" represents the condition where the masker sentence was not processed by the STRAIGHT algorithm. The labels "semi0," "semi3," …, and "semi15" represent the conditions with an F0 shift of 0, 3, …, and 15 semitones, respectively. The error bars represent the standard error of the mean calculated from the scores of the 7 subjects within each group.

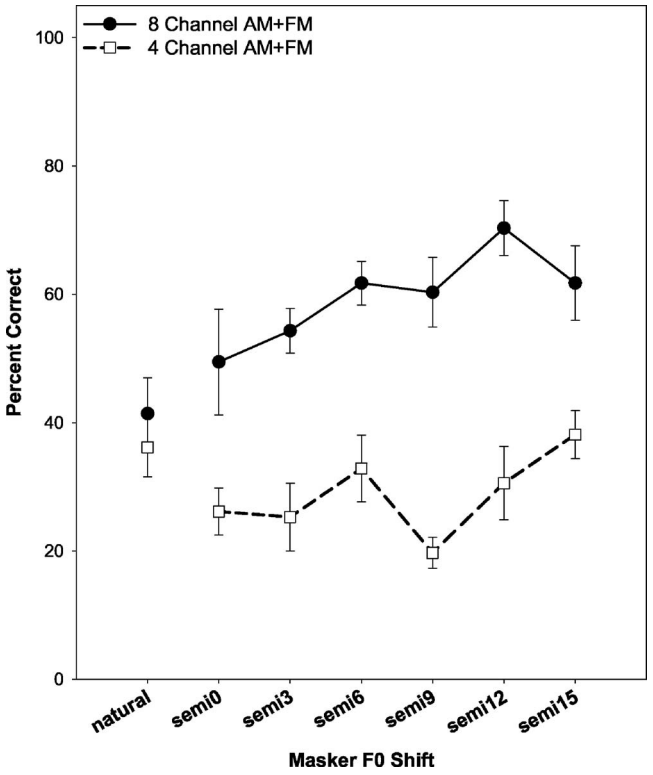


FIG. 6. Results for the normal-hearing subject groups presented with the 8-channel AM+FM-processed speech (closed circles and solid lines) or the 4-channel AM+FM speech (open squares and dashed lines) at a 10 dB SNR. The label "natural" represents the condition where the masker sentence was not processed by the STRAIGHT algorithm. The labels "semi0," "semi3," …, and "semi15" represent the conditions with an F0 shift of 0, 3, …, and 15 semitones, respectively. The error bars represent the standard error of the mean calculated from the scores of the 7 subjects within each group.

cochlear implant listeners provided only with temporal-envelope information may be unable to access the cues that normal-hearing listeners use to successfully segregate competing speech sounds.

How much do differences in F0 contribute to speech recognition under these conditions? The answer to this question can be addressed by comparing the results of the current study with that of Stickney *et al.* (2004). The mean F0 of the female talker from the previous study was 219 Hz. This F0 value is very close to that of the semi12 condition of the present experiment, which is 216 Hz. However, in the earlier study, the masker and target sentence stimuli provided various differences in talker characteristics in addition to F0 differences that could allow listeners to improve their performance. A study by Darwin *et al.* (2003) showed that differences in vocal tract length, in addition to F0, can help to segregate two competing voices, although it contributes much less than a difference in F0. Changing the vocal tract length (simulated by scaling the spectral envelope in a vocoder, which effectively multiplies the frequencies of all formants by a scale factor) produces an audible difference in voice quality that can be used to help segregate the speech of two competing voices. The male-female difference is associated with an upward shift in the frequencies of the formants by about 15%–20% (Peterson and Barney, 1952). However, in the present study, the masker and target sentences were spoken by the same individual, and the STRAIGHT algorithm shifted only the F0, leaving the formants unchanged. If F0 was the primary contributor, then there should be little difference between the two studies when comparing the amount of improvement from the same male masker to female masker condition (previous study) with that from the semi0 to semi12 condition (present study).

Normal-hearing listeners presented with the 4-channel AM-only condition at a 10 dB SNR showed comparable levels of speech recognition performance between the two studies. Stickney *et al.* (2004) found no difference in performance when the target sentences, spoken by a male voice, were masked either by a female voice or the same male voice as the target sentence. Likewise, in the present study, there was very little difference in score when the F0 was shifted from 0 semitones (analogous to the same male talker condition of the previous study) to 12 semitones (analogous to the female talker condition of the previous study); the total percent change with an F0 shift of 0–12 semitones was only 2%. This result suggests that with 4-channel AM-only processing, listeners cannot benefit from differences in F0. Adding FM though led to a 16% improvement in score in the 4-channel condition. However, there was no clear trend (no gradual

increase or decrease in performance) when the F0 was shifted from 0 to 12 semitones, suggesting that the addition of FM in this condition can improve overall intelligibility but may not provide sufficient F0 information. It is therefore possible that with a limited number of channels (such as 4), the additional F0-related information conveyed by the FM was still not sufficient to mediate an increase in performance with increasing F0 separation. This issue was addressed by increasing the number of channels to 8.

The percent change for the 8-channel AM conditions was quite different between the two studies. In Stickney *et al.* (2004), introducing a difference in voice gender between the target and masker, though not showing a statistically significant improvement, did increase the speech recognition score by an average of 28 percentage points. In contrast, in the present study, the percent change in performance when the F0 shift was increased from 0 to 12 semitones was minimal. One interpretation of these findings is that, with the stimuli in the previous study, some listeners might have been able to utilize differences in talker characteristics conveyed by the enhanced representation of temporal envelope cues in the 8-channel condition compared to the 4-channel condition. These listeners might have used cues other than F0 conveyed in the temporal and spectral envelope to improve their score, such as cues associated with differences in speaking rate between the two talkers and relatively coarse spectral differences between the male and female talker. In the present study, when FM information was added to the 8-channel condition, there was an improvement in score (20%) with an F0 shift of 12 semitones, which was not found when only AM information was provided or when listeners were presented with a smaller number of temporal envelope channels, with or without FM. This demonstrates the added benefit of F0 information conveyed by FM for segregating competing speech sounds, but the listener must also have sufficient spectral resolution to make use of the FM cue for segregating competing speech sounds on the basis of pitch (Oxenham *et al.*, 2004). With fewer channels, the analysis filters are broader, and the spectrotemporal resolution may be impaired.

Interestingly, the cochlear implant listeners in Stickney *et al.* (2004) showed a similar pattern of results as the normal-hearing listeners presented with 8-channel AM-only information. Some of the cochlear implant users could benefit from the temporal envelope differences between the two talkers, improving their score by an average of 20 percentage points with the female talker compared to the same male talker. On the other hand, in the present study, there was only a 5% improvement as the F0 was shifted from 0 to 12 semitones. The cochlear implant users evaluated here had poorer levels of performance overall compared to the implant users in the previous study, and their pattern of results were more similar to that of the normal-hearing listeners presented with 4-channel than 8-channel AM-only information. Therefore, some cochlear implant users can benefit from differences in the temporal envelopes of two competing talkers but might not benefit from differences in voice pitch. Likewise, it may also be true that only the better performing cochlear implant users will be able to utilize the additional temporal fine structure information. This is a topic that is itself interesting and will need further study.

Together these findings illustrate the combined usefulness of temporal envelope and temporal fine structure cues for auditory stream segregation (Bregman, 1990). The temporal envelope can provide cues for speaking rate and loudness. The temporal fine structure, on the other hand, can provide cues for voice pitch and formant transitions. With the natural pitch contour, F0 helps, in part, by giving momentary differences in pitch that allow listeners to segregate the two voices (Assmann, 1999). Consistent with this idea, several studies have demonstrated that the natural F0 contour leads to higher performance than a flattened F0 (Assmann, 1999; Binns and Culling, 2005; Watson and Schlauch, 2005). Furthermore, Binns and Culling discovered that the normal F0 contour provided some benefit over a flattened F0 when the target speech was presented in the presence of speech-shaped noise, but this benefit was much more pronounced when there were two competing speech sounds. It is possible that the F0 contour as well as the formant transitions provide a gradual change in frequency that can help the listener better track the target sound. Both the F0 contour and formant transitions would therefore follow the Gestalt principle of good continuation, thereby providing important cues for auditory stream segregation (Bregman, 1990).

Several investigators have identified methods for improving the encoding of the temporal envelope by reducing neural and electrical-field interactions between channels (Wilson *et al.*, 1991), enhancing the modulation depth (Geurts and Wouters, 1999), and increasing the rate of stimulation across channels (Rubinstein *et al.*, 1999). The coding of the additional temporal fine structure cue is also under investigation (Green *et al.*, 2004, 2005; Lan *et al.*, 2004; Nie *et al.*, 2005; Zeng *et al.*, 2005). The parameters used for FM coding of temporal fine structure, described in the present study, can be perceived by users of cochlear implants. This was demonstrated in a study by Chen and Zeng (2004). In their study, cochlear implant users were asked to detect the greatest change in pitch when the target was either a sinusoidal FM or a frequency sweep. Both FM depth and FM rate were varied. For the frequency sweep, the difference limen for FM depth with a 1000 Hz standard (the highest standard frequency tested) was 361 Hz. For the sinusoidal FM also with a 1000 Hz standard, the difference limens for FM depth were 400 and 549.4 Hz for FM rates of 160 and 320 Hz, respectively. At lower standard frequencies, the difference limens were significantly better, indicating an upper limit for FM coding. The FM rate and depth used in the FAME strategy have therefore been limited to 400 and 500 Hz, respectively, to be within the range that is perceivable by cochlear implant users. The FM rate could then be used to vary the interpulse interval of the pulse train carrier of a cochlear implant. Other potential implementations are described in an earlier publication by Nie *et al.*, (2005). Implementation of the FAME strategy for actual cochlear implant users is under way, as is the development of similar speech processing strategies that code temporal fine structure information. While results from actual cochlear implant users

await the implementation of these algorithms, the results from simulation studies indicate that enhancement of existing temporal envelope information and the addition of temporal fine structure have the potential to provide cues that can help cochlear implant listeners in one of the most challenging listening tasks that they are faced with: understanding speech with competing speech sounds.

ACKNOWLEDGMENTS

We are very grateful to Jennifer Lo and Rabia Farooquee who assisted in processing the stimuli and conducting the listening experiments. Dr. KaiBao Nie developed the FAME algorithm and user interface for processing the sentences. The IEEE sentences were created by Dr. Lou Braida and recorded by Dr. Monica Hawley and Dr. Ruth Litovsky. The STRAIGHT algorithm was provided by Dr. Hideki Kawahara. This work was supported by a NIH Grant No. F32 DC05900 awarded to G.S.S. and NIH Grant No. 2R01DC02267 awarded to F.G.Z.

[1]Although different sentence pairs might require greater or smaller F0 shifts for equal intelligibility, the overall pattern of the results for the standard simulation or FAME processing should be similar, specifically little to no effect of F0 shift with the implant simulation and some benefit of F0 shift for FAME processing.

Assmann, P. F. (**1994**). "The role of formant transitions in the perception of concurrent vowels," J. Acoust. Soc. Am. **96**, 1–9.

Assmann, P. F. (**1999**). "Fundamental frequency and the intelligibility of competing voices," Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, August 1–8, pp. 179–182.

Assmann, P. F., and Summerfield, Q. A. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Binns, C., and Culling, J. F. (**2005**). "The role of fundamental frequency (F0) contours in the perception of speech against interfering speech," J. Acoust. Soc. Am. **117**, 2606–2607.

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, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 263–269.

Bregman, A. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perception of simultaneous voices," J. Phonetics **10**, 23–26.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Burns, E. M., and Viemeister, N. F. (**1976**). "Nonspectral pitch," J. Acoust. Soc. Am. **60**, 863–869.

Chen, H., and Zeng, F.-G. (**2004**). "Frequency modulation detection in cochlear implant subjects," J. Acoust. Soc. Am. **116**, 2269–2277.

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

Darwin, C. J., and Hukin, R. W. (**2000**). "Effectiveness of spatial cues, prosody, and talker characteristics in selective attention," J. Acoust. Soc. Am. **107**, 970–976.

Flanagan, J. L., and Golden, R. M. (**1966**). "Phase vocoder," Bell Syst. Tech. J. **45**, 1493–1509.

Geurts, L., and Wouters, J. (**1999**). "Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants," J. Acoust. Soc. Am. **105**, 2476–2484.

Geurts, L., and Wouters, J. (**2001**). "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," J. Acoust. Soc. Am. **109**, 713–726.

Green, T., Faulkner, A., and Rosen, S. (**2004**). "Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants," J. Acoust. Soc. Am. **116**, 2298–2310.

Green, T., Faulkner, A., Rosen, S., and Macherey, O. (**2005**). "Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification," J. Acoust. Soc. Am. **118**, 375–385.

Hawley, M. L., Litovsky, R. Y., and Colburn, S. H. (**1999**). "Speech intelligibility and localization in a multi-source environment," J. Acoust. Soc. Am. **105**, 3436–3448.

Kawahara, H. (**1997**). "Speech representation and transformation using adaptive interpolation of weighted spectrum: VOCODER revisited," Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. **2**, pp. 1303–1306.

Kawahara, H., Masuda-Katsuse, I., and de Cheveigne, A. (**1999**). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," Speech Commun. **27**, 187–207.

Kiang, N. Y. S., Watanabe, T., Tomas, E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT, Cambridge, MA).

Lan, N., Nie, K., Gao, S. K., and Zeng, F.-G., (**2004**). "A novel speech processing strategy incorporating tonal information for cochlear implants," IEEE Trans. Biomed. Eng. **51**, 752–760.

Nie, K., Stickney, G. S., and Zeng, F.-G. (**2005**). "Encoding frequency modulation to improve cochlear implant performance in noise," IEEE Trans. Biomed. Eng. **52**, 64–73.

Oxenham, A., Bernstein, J. G., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of vowels," J. Acoust. Soc. Am. **24**, 175–184.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating masker," J. Acoust. Soc. Am. **114**, 446–454.

Rubinstein, J. T., Wilson, B. S., Finley, C. C., and Abbas, P. J. (**1999**). "Pseudospontaneous activity: Stochastic independence of auditory nerve fibers with electrical stimulation," Hear. Res. **127**, 108–118.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Smith, Z., Delgutte, B., and Oxenham, A. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.

Stickney, G. S., Nie, K., and Zeng, F.-G., (**2005**). "Frequency modulation added to implant simulations improves speech recognition in noise," J. Acoust. Soc. Am. **118**, 2412–2420.

Stickney, G. S., Zeng, F.-G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Watson, P. J., and Schlauch, R. S. (**2005**). "Spectral contributions to intelligibility of sentences with flattened fundamental frequency," J. Acoust. Soc. Am. **117**, 2606.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). "Better speech recognition with cochlear implants," Nature (London) **352**, 236–238.

Zeng, F.-G., Nie, K. B., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C. G., Cao, K. (**2005**). "Speech recognition with amplitude and frequency modulations," *Proceedings of the National Academy of Science*, Vol. 102, pp. 2293–2298.

Companding to improve cochlear-implant speech recognition in speech-shaped noise[a)]

Aparajita Bhattacharya[b)]
Hearing and Speech Research Laboratory, Department of Biomedical Engineering, University of California, Irvine, 316 Med Surge II, Irvine, California 92697

Fan-Gang Zeng[c)]
Hearing and Speech Research Laboratory, Departments of Anatomy and Neurobiology, Biomedical Engineering, Cognitive Sciences, and Otolaryngology—Head and Neck Surgery, University of California, Irvine, 364 Med Surge II, Irvine, California 92697

(Received 12 October 2006; revised 25 May 2007; accepted 25 May 2007)

Nonlinear sensory and neural processing mechanisms have been exploited to enhance spectral contrast for improvement of speech understanding in noise. The "companding" algorithm employs both two-tone suppression and adaptive gain mechanisms to achieve spectral enhancement. This study implemented a 50-channel companding strategy and evaluated its efficiency as a front-end noise suppression technique in cochlear implants. The key parameters were identified and evaluated to optimize the companding performance. Both normal-hearing (NH) listeners and cochlear-implant (CI) users performed phoneme and sentence recognition tests in quiet and in steady-state speech-shaped noise. Data from the NH listeners showed that for noise conditions, the implemented strategy improved vowel perception but not consonant and sentence perception. However, the CI users showed significant improvements in both phoneme and sentence perception in noise. Maximum average improvement for vowel recognition was 21.3 percentage points ($p<0.05$) at 0 dB signal-to-noise ratio (SNR), followed by 17.7 percentage points ($p<0.05$) at 5 dB SNR for sentence recognition and 12.1 percentage points ($p<0.05$) at 5 dB SNR for consonant recognition. While the observed results could be attributed to the enhanced spectral contrast, it is likely that the corresponding temporal changes caused by companding also played a significant role and should be addressed by future studies. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2749710]

PACS number(s): 43.66.Ts, 43.71.Ky, 43.71.Gv, 43.71.Es [AJO] Pages: 1079–1089

I. INTRODUCTION

In realistic listening situations, speech signals are often degraded by noise. While normal-hearing (NH) listeners are remarkably adept at extracting the critical information from noisy speech, this ability decreases rapidly with hearing loss. With the latest generation cochlear implants, the majority of implant users derive substantial benefits in quiet listening conditions, including communication over the telephone. However, their performance in noise is severely impeded, thereby significantly affecting the implant efficiency in real-life situations. In general, cochlear-implant (CI) users require much higher signal-to-noise ratios (SNRs) to match the performance of NH listeners on speech recognition tasks in noise (Fu and Nogaki, 2005; Hochberg *et al.*, 1992; Müller-Deiler *et al.*, 1995; Stickney *et al.*, 2004; Zeng and Galvin, 1999, Zeng *et al.*, 2005). Zeng *et al.* (2005) found that the speech reception threshold (SRT) of NH listeners was approximately 14 dB better than that of the CI users in steady-state noise. The difference was more drastic in fluctuating background noise. With a female talker as the masker, the SRT of NH listeners was approximately 32 dB higher than that of the CI users. SRT was defined as the SNR necessary for a listener to produce 50% correct score.

The auditory system uses nonlinear processing to provide the necessary spectral and temporal resolution. Outer hair cells (OHCs) in the cochlea are responsible for the active mechanisms observed in the peripheral auditory system. These OHCs employ nonlinear adaptive gain processing to encode the large dynamic range by a relatively narrow physiological range of the auditory nerve fibers. The same OHCs are also responsible for the sharp frequency selectivity of the auditory system. Another nonlinear phenomenon arising from complex interactions between OHCs and the basilar membrane is two-tone suppression (Rhode, 1974; Ruggero *et al.*, 1992). It is characterized by a decrease in the evoked response to a tone in the presence of a second tone (Sachs and Kiang, 1968). Two-tone suppression is considered to be the primary mechanism underlying spectral enhancement and is thought to improve the SNR of the stronger components (Rhode *et al.*, 1978; Sachs *et al.*, 1983; Stoop and Kern, 2004). Spectral enhancement is defined as an increase in peak-to-valley difference which is also referred to as "spectral sharpening."

[a)]Portions of this work were presented in "Companding to improve cochlear implants' speech processing in noise," *2005 Conference on Implantable Auditory Prosthesis*, Asilomar Conference Center, Pacific Grove, California, August 2005.

[b)]Author to whom correspondence should be addressed. Electronic mail: abhattac@uci.edu
[c)]Electronic mail: fzeng@uci.edu

Spectral sharpening may also result from competitive interactions between the neurons. Early studies on the Limulus retina demonstrated a neural contrast enhancement phenomenon, termed lateral inhibition, in which the output of a neuron is inhibited by the inputs from the adjacent neurons (Hartline, 1969). Lateral inhibition has also been observed in the auditory system. Blackburn and Sachs (1990) studied the discharge patterns of the cat anteroventral cochlear nucleus neurons in response to steady-state vowels. They found that certain neurons receive strong inhibitory inputs from surrounding neurons so that the overall neural response can maintain the valley between peaks (i.e., formants), particularly at moderate to high sound pressure levels and in the presence of background noise. Along with two-tone suppression, lateral inhibition too is likely responsible for preserving the spectral contrast between peaks and the valleys (Rhode and Greenberg, 1994).

Cochlear damage reduces the dynamic range, broadens the tuning curves (Kiang *et al.*, 1976) and eliminates two-tone suppression (Ruggero *et al.*, 1992). The combined effect is poor spectral resolution and reduced spectral contrast. As a result, the neural representation of formants in response to steady-state vowels is degraded, that results in poor vowel recognition(Loizou *et al.*, 2000; Miller *et al.*, 1997). Degraded spectral resolution also causes abnormal susceptibility to noise in hearing-impaired listeners (Horst, 1987).

In CI users, several factors limit the available spectral resolution and spectral contrast. Current cochlear implants that use amplitude compression to map the large acoustic amplitude range into a narrow electrical dynamic range have a side effect of reducing the spectral contrast. Loizou *et al.* (2000) studied the effect of input envelope amplitude compression on the recognition of vowels and consonants. They used stimuli processed by a six-channel cochlear implant simulator and tested NH listeners. They found that compression degraded the perception of vowels and recognition of place of articulation of the consonants. Studies on intensity discrimination with electric stimulation suggest that sometimes the size of just discriminable steps is smaller than NH listeners (Hochmair-Desoyer *et al.*, 1981; Nelson *et al.*, 1996) but it is not enough to compensate for the degrading effects of reduced dynamic range. This is because the number of discriminable steps within the dynamic range is considerably smaller compared to that of the NH listeners. In addition, nonlinear mechanisms such as two-tone suppression have not been implemented in cochlear implants. As a result, current CI users typically need a spectral contrast of 4–6 dB more than the NH listeners to accurately identify vowels in quiet listening condition (Loizou and Poroy, 2001). Spectral resolution in cochlear implants is limited by the number and location of the electrodes along the cochlea, the number of surviving neurons and the amount of current spreading from the stimulating electrodes. To a large extent, impaired spectral resolution and reduced spectral contrast are responsible for the susceptibility of cochlear implants to noise.

To compensate for the impaired ability to understand speech in noise, many noise suppression algorithms have been proposed for cochlear implants (Hamacher *et al.*, 1997; Toledo *et al.*, 2003; Weiss, 1993; Wouters and Vanden Berghe, 2001; van Hoesel and Clark, 1995). These algorithms use either single microphone or dual microphones and can be adaptive or nonadaptive in nature. Weiss (1993) showed that the INTEL method, which is used to suppress the random wideband noise, reduces the noise-induced deviations in the second formant frequency. Adaptive beamforming is a noise reduction technique, which uses signals from two or more microphones, to attenuate signal coming from directions other than the front. The portable beamformer using two microphones implemented by van Hoesel and Clark (1995) showed considerable improvements in speech intelligibility at 0 dB SNR. Wouters and Vanden Berghe (2001) used a two-channel two-stage adaptive filtering beamformer as a pre-processing stage in LAURA cochlear implants and found an improvement of 10 dB in SNR for the perception of consonant-vowel-consonant words. It is important to note that the speech source was angled at 90° with respect to the noise source in azimuth and the improvement may be smaller for angles less than 90°.

Several spectral enhancement techniques have particularly focused on compensating for the degraded spectral resolution of an impaired ear (Baer *et al.*, 1993; Bunnell, 1990; Clarkson and Bahgat, 1991; Franck *et al.*, 1999; Lyzenga *et al.*, 2002). Bunnell (1990) used a contrast enhancement technique in which the envelope amplitude of each fast Fourier transform bin was enhanced proportionately to the difference in the original envelope amplitude and the average spectrum level. He found a small improvement in the identification of stop consonants in quiet. Baer *et al.* (1993) convolved the spectrum with a difference of Gaussian filter to provide spectral enhancement. They showed that their NH subjects preferred speech in noise with moderate enhancement in terms of quality and intelligibility. This technique, combined with phonemic compression, improved the perception of vowels in hearing-impaired listeners but degraded the understanding of consonants (Franck *et al.*, 1999). To provide noise suppression, Clarkson and Bahgat (1991) expanded the temporal envelopes in different frequency bands. They found a small but significant improvement at 0 dB SNR in NH listeners. Lyzenga *et al.* (2002) employed a similar enhancement technique but followed it by an additional "lift" stage to counteract the effect of upward spread of masking. They applied spectral smearing in the end to simulate loss of frequency selectivity and tested NH listeners. They found that enhancement employed separately did not produce any improvement in SRT but enhancement and lift applied together improved the SRT by approximately 1 dB.

To counterbalance the degraded neural representations of the second ($F2$) and third ($F3$) formant frequencies in the impaired ear (Miller *et al.*, 1999a), Miller *et al.* (1999b) proposed a contrast enhancing frequency shaping algorithm that selectively amplifies $F2$ and $F3$ without modifying the spectral valleys. They found considerable improvements in the neural representation of $F2$ in acoustically traumatized cats.

Recently, Turicchia and Sarpeshkar (2005) have proposed a novel spectral enhancement scheme, companding, which combines two-tone suppression and dynamic gain

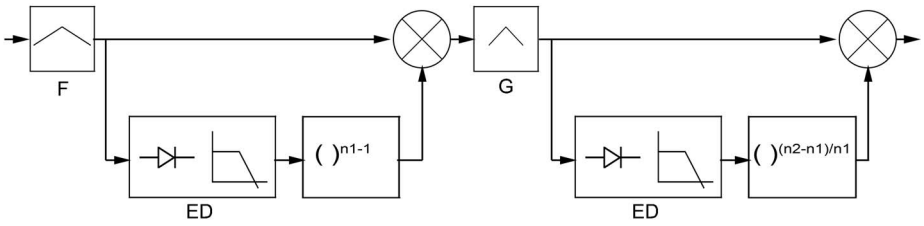


FIG. 1. Detailed architecture of a single channel. ED-envelope detector.

control in order to increase the spectral contrast. One specific goal in their study is to use the scheme to improve speech recognition in noise in cochlear implants. Studies have shown that companding is also present along the auditory pathway. Both cochlea and the cochlear nucleus perform logarithmic compression on the input signals, while the brain performs exponential expansion (Zeng and Shannon, 1994; Zeng and Shannon, 1999).

The overall companding architecture can be found in Turicchia and Sarpeshkar, 2005, and is briefly described here. First, the incoming signal is divided into a number of frequency channels by a bank of relatively broad bandpass filters F. Figure 1 shows the detailed architecture of a single channel companding pathway. The signal within each channel is subjected to amplitude compression. The extent of compression depends on the output of the envelope detector, ED, and the compression index, n_1. The compressed signal is then passed through a relatively narrow bandpass filter G before being expanded. The gain of the expansion block depends on the corresponding ED output and the ratio n_2/n_1. The outputs from all the channels are summed to obtain the processed signal.

The present work implemented and evaluated the companding architecture as a front end for a CI processor. This paper has the following organization. Section II describes implementation of the companding architecture and discusses the rationale behind choosing different parameter values. Section III shows the effect of companding on the acoustic features in both time and frequency domains. Section IV describes the evaluation results of companding in vowel, consonant, and sentence recognition tasks performed by both NH and CI subjects. Section V discusses the importance and possible mechanisms underlying the present findings. Section VI summarizes the present findings and points out future research directions.

II. IMPLEMENTATION

The companding architecture was implemented in MATLAB (The MathWorks, Natick, MA, USA). The precompression filter F and the postcompression filter G are zero-phase, bandpass filters with magnitudes described by the following transfer functions:

$$F'(s) = \left[\frac{2\left(\frac{\tau}{q_1}\right)s}{\tau^2 s^2 + 2\left(\frac{\tau}{q_1}\right)s + 1} \right]^2$$

$$G'(s) = \left[\frac{2\left(\frac{\tau}{q_2}\right)s}{\tau^2 s^2 + 2\left(\frac{\tau}{q_2}\right)s + 1} \right]^2.$$

Turicchia and Sarpeshkar used zero-phase filters for the companding-off cases to avoid interference between the channels. However, they did not use zero-phase filters for the companding-on cases. The time constant is given by $\tau = 1/2\pi f_r$, where f_r is the resonant frequency of the filters in the channels. The envelope detector in each channel consists of an ideal full wave rectifier and a first order low-pass filter. The time constant of the low-pass filter was set as $\tau_{ED} = w\tau$. Turicchia and Sarpeshkar discussed the importance of the relative tuning between pre- and postcompression filters as well as the compression index in the companding strategy.

Here, we found that the degree of spectral enhancement depends on the number of channels. Figure 2 compares the frequency spectra of the steady-state portions of the original vowel /hid/ (lighter trace) and the same vowel processed with companding strategy (darker trace) as a function of the number of channels. The resonant frequencies for each channel were logarithmically spaced between 100 and 8000 Hz. We chose $q_1 = 2$, $q_2 = 12$, $w = 40$, $n_2 = 1$ and $n_1 = 0.3$. The first formant frequency is at 400±50 Hz, the second formant is at 1800±100 Hz and the third formant is at 2570±140 Hz. We compared the peak-to-valley differences of the power spectra with peak corresponding to the first formant and valley corresponding to the dip between the first and second formant frequencies. We found that companding always enhances the highest peak (typically the first formant), but reducing the number of channels to less than 40 produces an undesirable suppression of the relatively weaker second and third formants. On the other hand, increasing the number of channels beyond 50 does not further enhance spectral contrast. Based on these results, the number of channels was chosen to be 50 for all the experiments.

In addition, we found that time constant of the envelope detectors affects the level of spectral enhancement. Figure 3 shows the outputs of a 50 channel companding processing of the vowel /hid/ as a function of the time constant. The lighter traces represent the unprocessed vowel and the darker traces represent the processed vowel. We see that the amount of spectral enhancement increases as w increases from 5 to 30, and plateaus thereafter. For all the experiments from here on, w was chosen as 40. Moreover, values of q_1, q_2, n_1 and n_2 remained the same as in the previous analysis (Fig. 2) for all

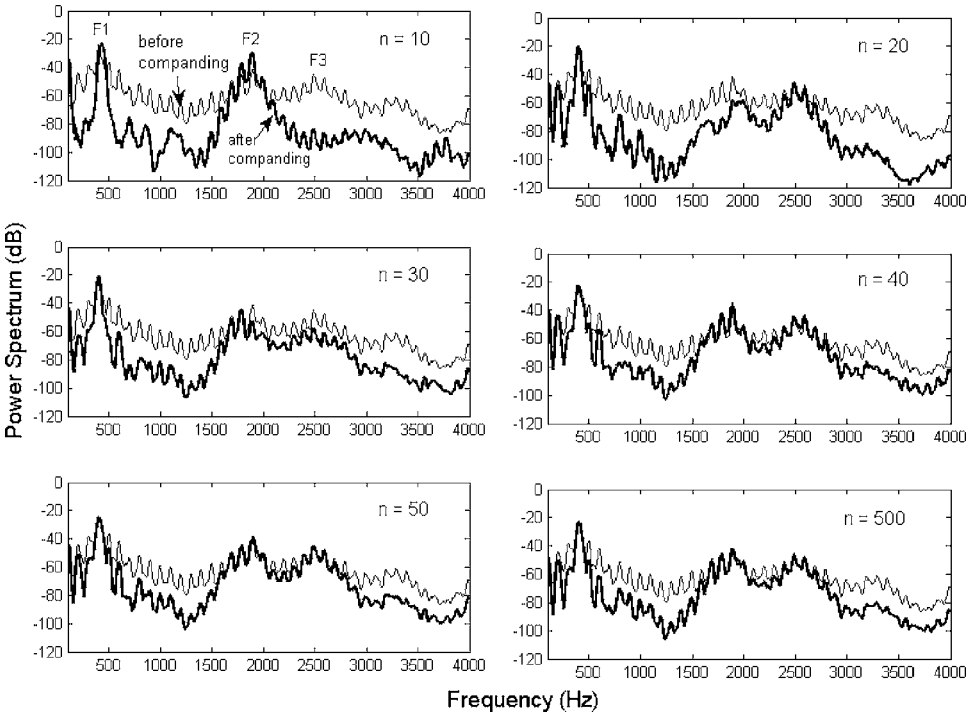


FIG. 2. Spectra of the vowel /hid/ as a function of the number of channels (n). Lighter traces correspond to the original stimuli and the darker traces represent the stimuli after companding.

the experiments reported in this work. All values were chosen based on the simulation results giving the maximum spectral contrast without degrading the signal quality.

Acoustic simulation of the eight-channel cochlear implant consisted of eight fourth-order bandpass Butterworth filters with frequencies between 100 and 5000 Hz (Shannon *et al.*, 1995; Dorman *et al.*, 1997a). Frequency spacing followed the Greenwood model, emulating equal spacing on the basilar membrane. The envelope of the signal in each band was extracted by full-wave rectification and low-pass filtering (eighth-order Butterworth) with a 160 Hz cutoff frequency. The envelope of each band was used to amplitude modulate a sinusoid at the center frequency of the channel. The modulated signals from all the channels were summed to form the acoustic simulation of an eight-channel cochlear implant.

III. ACOUSTIC ANALYSIS

A. Vowels

Figure 4 shows the temporal wave forms of the vowel /hid/ before (panel a) and after (panel b) companding, respectively. It is somewhat surprising to observe that companding also enhances the temporal contrast by sharpening the onset.

Figure 5 shows the spectra of the vowel /hid/ in steady-state speech-shaped noise (SSN) as a function of SNR. The lighter traces represent the inputs (original vowel in noise without companding), whereas the darker traces represent the corresponding outputs after companding. As the SNR decreases, the formant peaks are increasingly lost. We see that

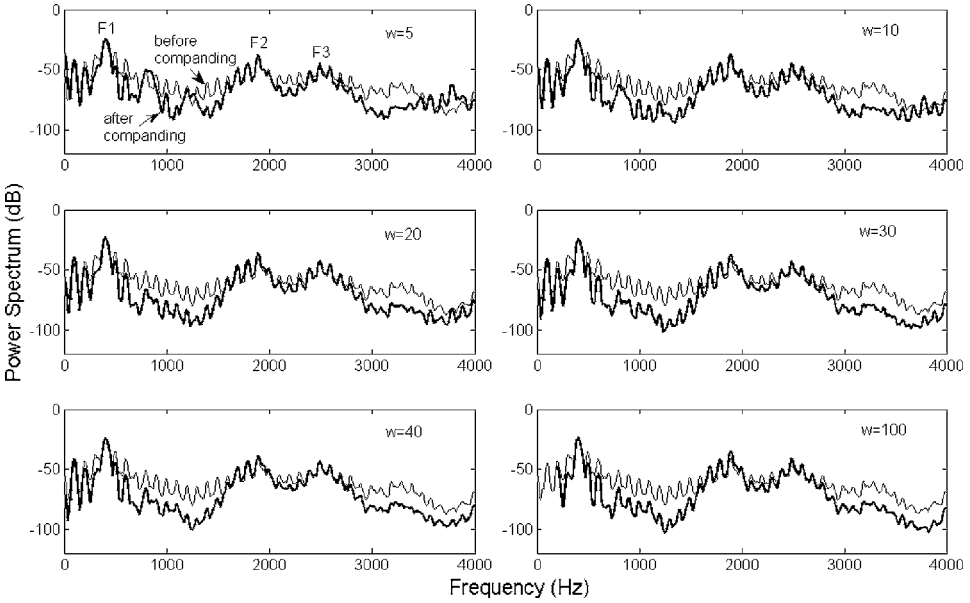


FIG. 3. Spectra of the vowel /hid/ as a function of time constant of the envelope detectors ($w\tau$). Lighter traces correspond to the original stimuli and the darker traces represent the stimuli after companding.

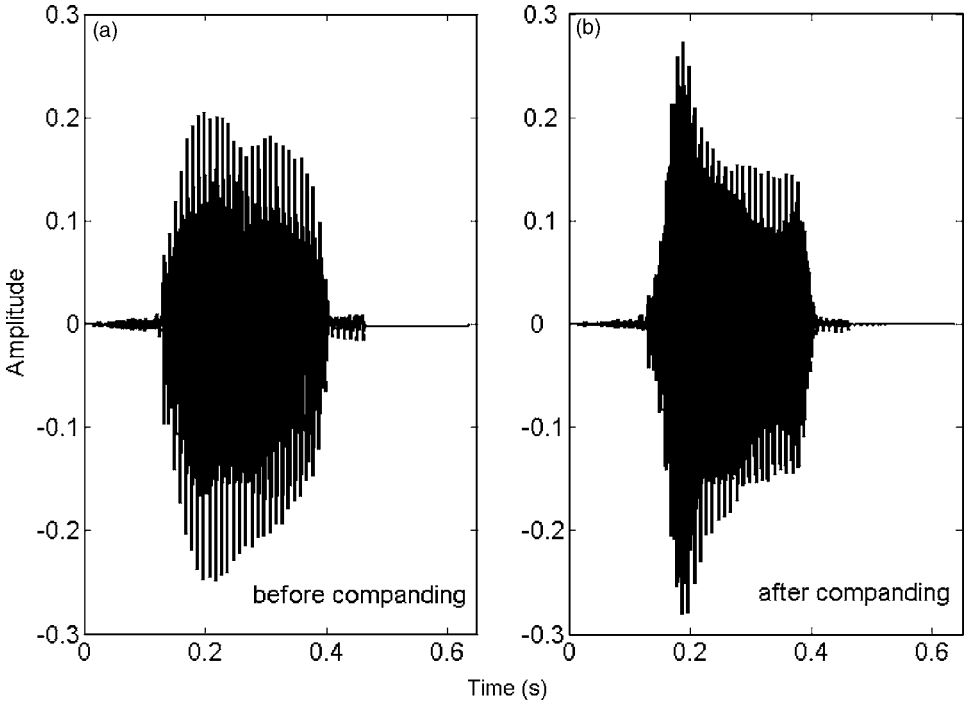


FIG. 4. Temporal wave forms of the vowel /hid/ before and after companding.

companding reduces the background noise while preserving the formant peaks, even at −10 dB SNR, when the decibel difference between the adjacent peaks and valleys is not apparent in the original stimulus.

B. Consonants

Figure 6 shows the spectra and the temporal wave forms of the consonant /aFa/. Panel a shows the spectra of the initial vowel part of /aFa/ and panel b shows the spectra of the consonant part following the initial vowel part. The lighter traces correspond to the input (i.e., the original consonant) and the darker traces correspond to the data after companding. We see that the formant peaks are enhanced during the initial vowel part but the spectral sharpening during the consonant part is relatively weak. This result is understandable, because, unlike vowels, consonants generally have flat spectra and lack prominent spectral peaks. Panels c and d show the temporal wave forms of the same consonant before and after companding, respectively. Similar to the vowel results, companding enhances changes in the temporal wave form envelope of the consonants.

IV. SPEECH RECOGNITION EVALUATION

Speech recognition experiments were conducted in NH and CI subjects. The tests included recognition of vowels,

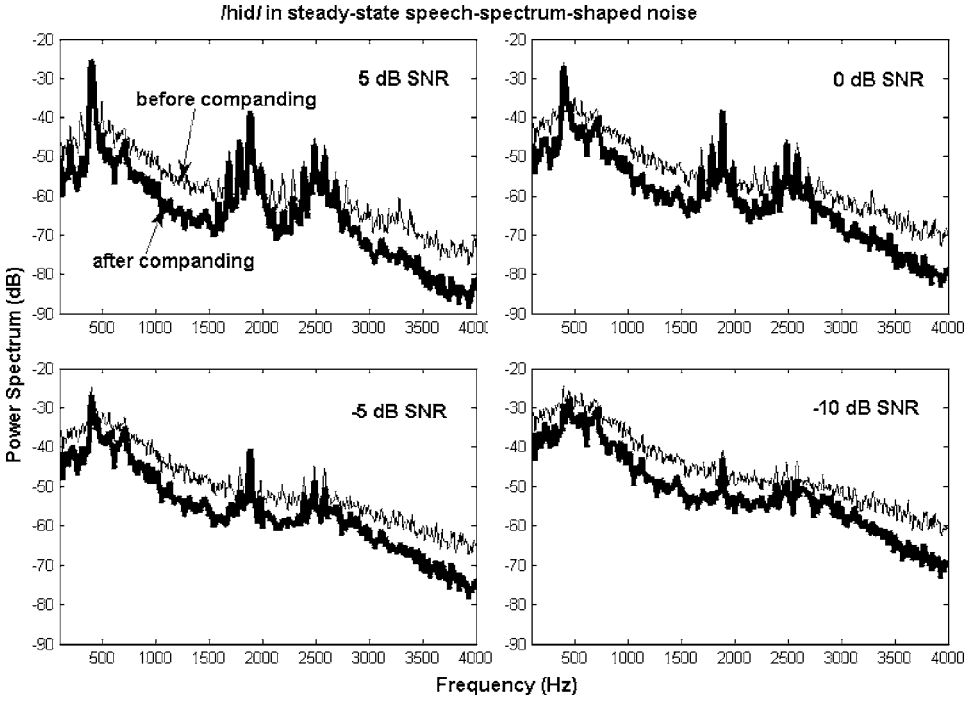


FIG. 5. Spectra of the vowel /hid/ in steady-state speech-spectrum-shaped noise at different SNRs. Lighter traces correspond to the original stimuli and the darker traces represent the stimuli after companding.

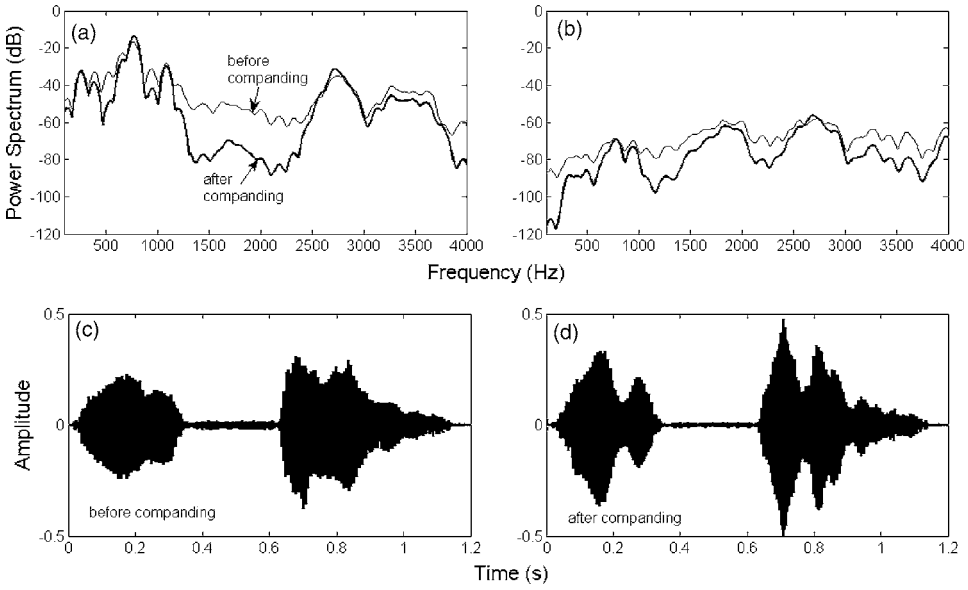


FIG. 6. (a) Spectra of the initial vowel part of the consonant /aFa/. (b) Spectra of the consonant part following the initial vowel part. Lighter traces correspond to the original stimuli and the darker traces represent the stimuli after companding. (c), (d) Temporal wave forms of the same consonant before and after companding, respectively.

consonants and sentences in quiet and in the presence of a steady-state SSN. NH subjects performed these tests using unprocessed and CI simulated stimuli, with and without companding in both cases. The CI subjects used their clinical processors to perform the same tests but with the unprocessed stimuli (with and without companding) only. The root mean square levels of all the stimuli were equalized.

A. Subjects

A total of seven NH subjects participated in the phoneme recognition tests and nine subjects in the sentence recognition tests. Out of these, six subjects participated in both phoneme and sentence recognition tests. The subjects were aged between 19 and 36 years with no reported history of hearing loss. Tests were also conducted on seven implant users between the ages of 56 and 79 years. The implant subjects included 5 Nucleus 24 (C1 to C5) and 2 Clarion II (C6 and C7) users. Subject C6 was pre-lingually deafened. The subjects had at least two years of implant experience at the time of testing. Detailed information on the CI users is presented in Table I. All subjects were native English speakers. They were compensated for their participation.

B. Test material

The phoneme materials included 12 /hvd/ vowels (Hillenbrand *et al.*, 1995) and 20 /aCa/ consonants (Shannon *et al.*, 1999) spoken by a male and a female speaker. The target sentence material consisted of 250 hearing in noise test (HINT) sentences spoken by a male speaker (Nilsson *et al.*, 1994). Both phonemes and the sentences were presented in quiet and in steady-state SSN at different SNRs between −10 dB and +10 dB, spaced at 5 dB intervals. The SSN was constructed by filtering white noise with the talker's long-term speech spectrum envelope derived using a tenth-order LPC analysis. The stimuli were presented via headphones (Sennheiser HDA 200) to the NH subjects while the CI subjects listened to the signal coming out of a speaker (Grason-Stadler 61 Clinical Audiometer). The signals were presented at 70 dB sound pressure level. The noise level was varied to produce different SNRs.

C. Procedure

All experiments were conducted in a double-walled, sound-attenuated booth. For the phoneme recognition tests, a graphical user interface containing 12 vowels or 20 conso-

TABLE I. Detailed information of the cochlear implant users.

Subject	Gender	Age (yrs)	Cause of deafness	Duration of implant use (yrs)	Clinical speech strategy	Vowel	Consonant	HINT
C1	F	77	Blood clot	2	ACE	42%	76%	84%
C2	F	71	Fever	4	ACE	74%	69%	98%
C3	F	62	Unknown	2	ACE	82%	91%	100%
C4	F	70	Virus	7	ACE	64%	84%	94%
C5	M	79	Unknown	2	CIS	58%	49%	92%
C6	F	56	Unknown	4	Clarion CII	21%	25%	15%
C7	F	52	Overnight SNHL	3	Clarion CII	58%	85%	100%

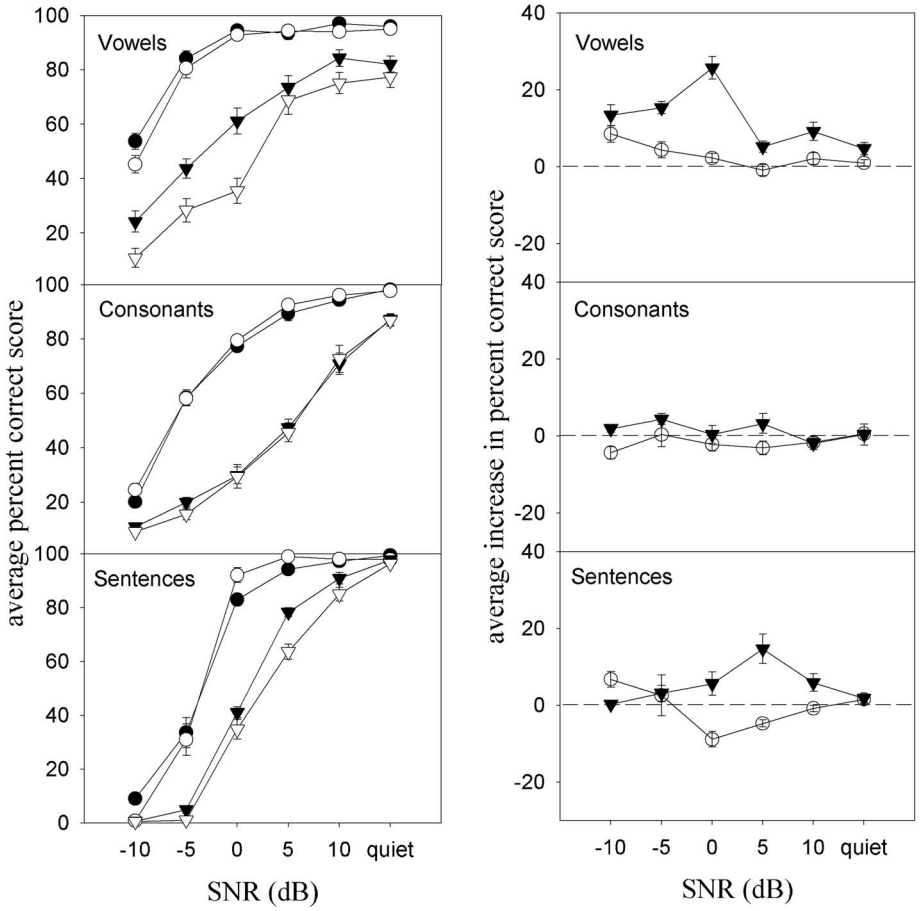


FIG. 7. Phoneme and sentence recognition scores for the NH subjects. Left panels show average percent correct scores and the right panels show average increase in the percent correct scores. Left panel: Circled traces represent the stimuli without CI simulation, whereas the inverted triangular traces correspond to the CI simulated stimuli. Open symbols correspond to the original stimuli and the filled symbols correspond to the stimuli after companding. Right panel: Filled inverted triangular traces correspond to the CI simulated stimuli and the open circular traces correspond to the stimuli without CI simulation. Vertical bars represent the standard errors.

nants displayed as buttons was presented on the computer screen. After a phoneme was presented, the subjects were instructed to click on the button corresponding to the presented phoneme. The subjects were provided with feedback regarding the response after each phoneme presentation and were instructed to guess if they were not sure. All the NH subjects had to take a pretest consisting of unprocessed stimuli in quiet and only those subjects who scored above 90% correct were allowed to participate in the study. The subjects did not receive any training prior to the tests. The noise conditions were presented in the order of increasing level of difficulty, to counterbalance any learning effect. The order of processing conditions for each noise condition was randomized and balanced across subjects. The phonemes were presented randomly for each condition.

In the sentence recognition tests, the subjects were presented with a target sentence. They were asked to type in as many words as possible from the sentence using a computer keyboard. The number of correctly identified words was calculated to give the final percent correct score. No feedback was provided during the test and the subjects were instructed to guess if they were not sure. Similar to the phoneme recognition tests, all NH subjects took a pretest consisting of unprocessed stimuli in quiet and only those subjects who scored above 85% correct were allowed to participate in the study. The noise conditions were presented in the order of increasing level of difficulty, to counterbalance any learning effect. The processed and unprocessed stimuli were presented randomly in each test condition. The sentences were not repeated.

D. Results

Figure 7 shows the phoneme recognition and sentences recognition scores as a function of SNR for the NH subjects. The left panels show the average percent correct score and the right panels show the average increase in percent correct score as a function of SNR. In the left panels, open symbols correspond to stimuli before companding and the filled symbols correspond to stimuli after companding. The circled traces correspond to stimuli without eight-channel CI simulation and the inverted triangular traces correspond to stimuli with eight-channel CI simulation. In the right panels, the open circular traces correspond to stimuli without CI simulation and the filled inverted triangular traces correspond to stimuli with CI simulation.

For the vowel recognition test, analysis of variance (ANOVA) showed that companding improved the performance for both stimuli with $[F(1,6)=98.10, p<0.001]$ and without $[F(1,6)=62.07, p<0.01]$ CI simulation. Increase in percent correct scores varied nonmonotonically with larger improvements seen in the case of stimuli with CI simulation. At −10 dB SNR, difference in the average increases in percent correct scores between stimuli with CI simulation and stimuli without CI simulation were 5%, which increased to 25% at 0 dB SNR and thereafter decreased to 4% in quiet.

For the consonant recognition test, companding did not produce any significant change in the performance for stimuli without CI simulation $[F(1,6)=1.99, p=0.22]$. No significant improvement was seen in the case of consonants with CI simulation either $[F(1,6)=1.51, p=0.27]$. Unlike the

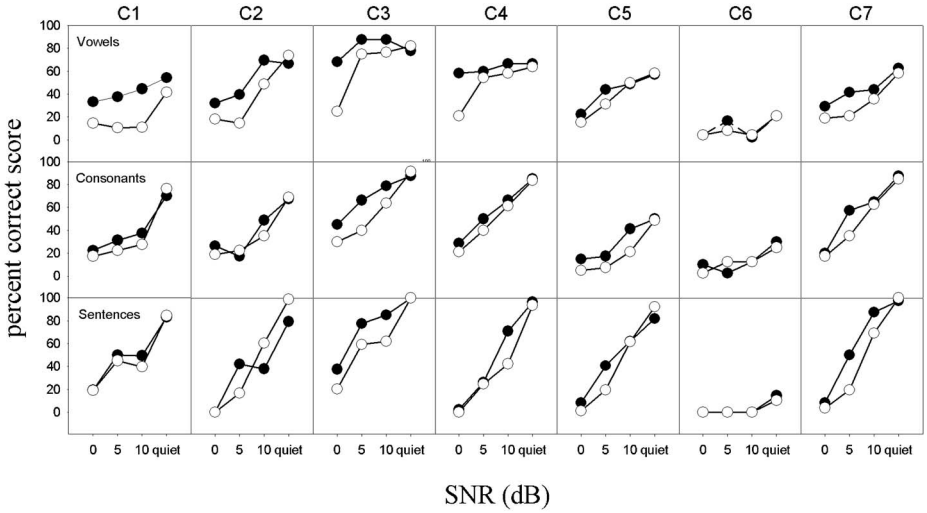


FIG. 8. Phoneme and sentence recognition scores for the CI users. C1 to C5 - Nucleus 24 users, C6 and C7 - Clarion II users. Upper, middle and the bottom panels show scores on vowel, consonant and sentence recognition tests, respectively. Open circles correspond to the original stimuli and the filled circles correspond to stimuli after companding.

vowel recognition results, there was no significant difference in the performances between consonants with CI simulation and the consonants without CI simulation.

For the sentence recognition test, companding seemed to degrade the performance in the case of stimuli without CI simulation at 0, 5, and 10 dB SNRs $[F(1,8)=44.38, p<0.001]$. However, subjects benefited from companding for the stimuli with CI simulation $[F(1,8)=25.53, p<0.05]$. Similar to the vowel recognition results, the improvement varied nonmonotonically with SNR. Larger improvement was seen in the case of stimuli with CI simulation at 0, 5, and 10 dB SNRs ($p<0.05$). At −10 dB SNR, improvement in the case of sentences without CI processing was better than the sentences with CI simulation ($p<0.05$), whereas there was no significant difference in the improvements at −5 dB SNR and in quiet. In general, the performance did not vary significantly with companding in quiet.

Figure 8 shows phoneme recognition and sentence recognition scores as a function of SNR for the CI users. The upper, middle and the bottom panels show the scores on vowel, consonant and sentence recognition tests. Open circles correspond to the original stimuli and the filled circles correspond to stimuli after companding. The improvement in performance for the vowel recognition tests was found to be as high as 43% (C3, 0 dB SNR). Similarly the maximum increases in percent scores were 26% (C3, 5 dB SNR) and 30% (C3, 5 dB SNR) for consonant recognition and sentence recognition tasks, respectively.

Figure 9 shows the average performance of CI users as a function of SNR. For the left panels, open symbols represent stimuli without companding and the filled symbols correspond to stimuli with companding. Because subject C6 showed a floor effect without any notable improvement with companding, her data were excluded from the above average analysis and also from the following statistical analysis.

For the vowel recognition test, subjects performed better with companding $[F(1,5)=35.50, p<0.005]$ at all SNRs except in quiet $[F(1,5)=0.18, p=0.69]$. The maximum average improvement was 21.3% ($p<0.05$) at 0 dB SNR. Similarly, consonant recognition in noise improved with companding $[F(1,5)=26.94, p<0.005]$ with a maximum average improvement of 12.1% ($p<0.05$) at 5 dB SNR. Companding also produced better performance for sentence recognition in noise $[F(1,5)=11.50, p<0.05]$, with a maximum average improvement of 17.7% ($p<0.05$) at 5 dB SNR. No significant improvement was seen in quiet ($p=0.36$). In general, the cochlear implant users benefited from companding in noise conditions.

V. DISCUSSION

A. Comparison with previous studies

Turicchia and Sarpeshkar showed that spectral contrast is an emergent property of the companding strategy and had speculated that the strategy has the potential to improve speech performance in noise. Loizou (2005) implemented a 16 channel companding strategy in actual CI users and found a modest improvement in vowel recognition but no improvement in consonant and sentence recognition. Here we found that a 50 channel companding implementation significantly improved the recognition of both phonemes and sentences in noise. Based on the acoustic analysis presented in Sec. III, we attribute the observed improvement to optimization of the companding parameters in the present implementation (discussed in the next section).

In another study, also investigating the effects of companding, Oxenham *et al.* (2007) tested NH listeners using sentences processed through a noise-excited envelope vocoder. Speech intelligibility was measured in steady-state SSN at 0, 3 and 6 dB SNRs. They varied the number of channels, channel bandwidths and time constants of the envelope detectors. They showed that a 50 channel companding implementation, using parameters similar to that used in the present study, improved the performance by 6 percentage points, averaged across subjects and SNRs. We found an average improvement of 15 and 6 percentage points at 5 and 0 dB SNRs, respectively, which is consistent with their finding. Further, they showed that by reducing the number of analysis channels to 16, the performance was improved by 4 percentage points (2 percentage points less than 50 analysis channels). Thus the degree of benefit from companding drops with reducing the number of channels. They also

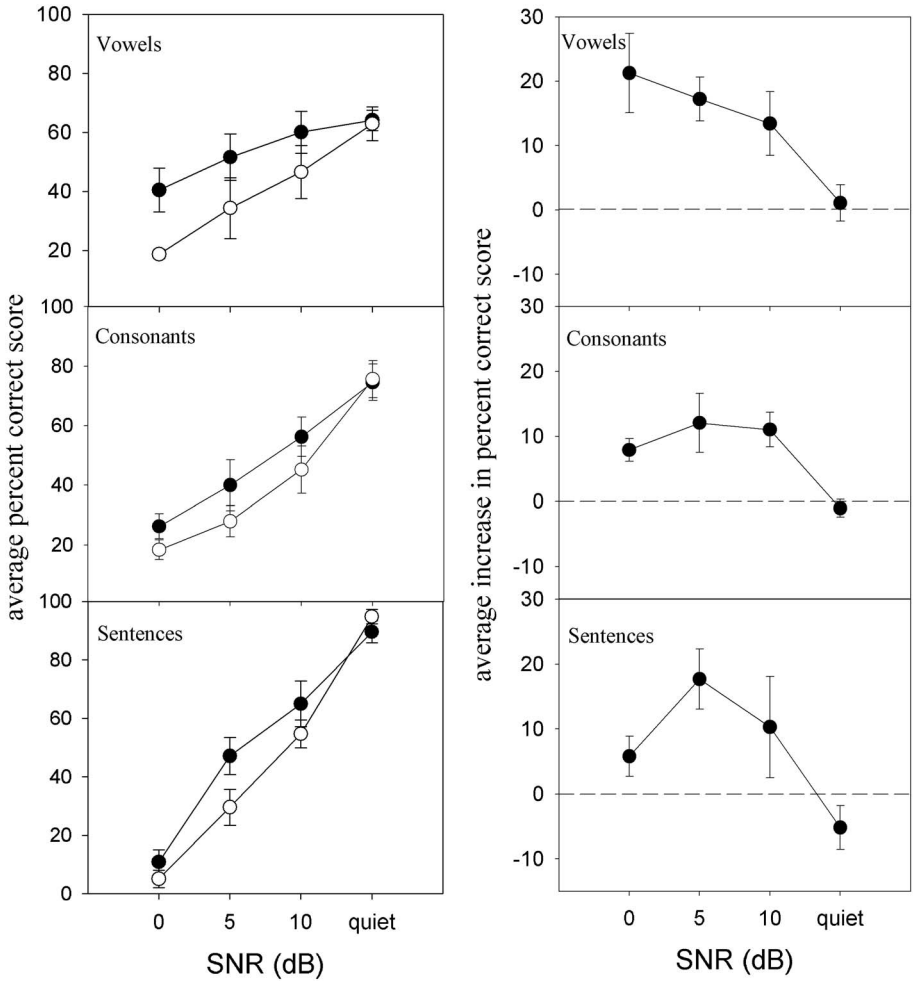


FIG. 9. Average percent correct scores and the average increase in percent correct scores for the CI subjects. Left panel: Open symbols correspond to the original stimuli and the filled symbols correspond to the stimuli after companding. Vertical bars represent the standard errors.

showed that decreasing the sharpness of tuning of the postcompression filter (G) by a factor between 2 and 3 did not decrease the improvement in intelligibility. Finally, as we had predicted from acoustic analysis, they showed that companding with smaller envelope detector time constants produced no benefit.

B. Optimization of the companding parameters

We identified four key parameters that need to be optimized to achieve a desirable spectral enhancement and speech performance in noise. These parameters are (1) number of channels, (2) the relative tuning between the pre- and postcompression filters (F and G), (3) the compression index (n_1), and (4) the time constant of the envelope detector ($w\tau$).

First, the effectiveness of companding depends upon the number of channels. Increasing the number of channels increases the processing time. Too few channels will result in the suppression of local spectral peaks, namely the weaker higher formants (Fig. 2). An optimal number of channels should achieve maximal spectral enhancement with sufficient frequency representation while minimizing the suppression of useful peaks. We found the optimal number of channels to be about 50 for the values tested.

In addition, the relative tuning of the two filters creates a difference in the levels of compression and expansion. A tone at the resonant frequency f_r of a channel is suppressed when the narrow filter G blocks a stronger tone entering the same channel. The tuning of F determines the range of frequencies above and below f_r of the channel that can suppress f_r. The tuning of G determines the range of frequencies around f_r that will not suppress f_r (Turicchia and Sarpeshkar, 2005). Hence the relative tuning of the two filters along with the number of channels determines the localness of spectral enhancement. We used a $q_1=2$ for the pre-compression filter F and $q_2=12$ for the postcompression filter G.

Furthermore, the degree of spectral enhancement depends on the value of the compression index, n_1. The smaller the value of n_1, the greater the difference between the degree of compression and the degree of expansion, resulting in greater enhancement of spectral contrast. We chose n_1 to be 0.3 and n_2 to be 1 in the present implementation.

Finally, the time constant of the envelope detectors can also affect the performance of the strategy. A smaller value of w (higher envelope cutoff frequency) generates a large number of unwanted frequencies within the channel thereby reducing the spectral contrast. We chose w to be 40 in the present implementation.

C. Factors affecting companding performance

The present results show that the effectiveness of companding depends on speech materials, listeners, and listening conditions. We found that companding enhances the spectral

peaks, but mainly for vowels and not consonants. Since the companding strategy in this study has higher channel density at lower frequencies than at higher frequencies, and vowels generally contain stronger spectral peaks at lower frequencies than consonants, greater spectral enhancement is produced for vowels compared to consonants. As a result, more improvement was seen in vowel perception scores for both NH listeners and CI users. This interpretation is consistent with previous studies showing that spectral smearing affected vowel perception more than consonant perception (Boothroyd *et al.*, 1996).

NH subjects listening to CI simulations showed better performance than CI users for both phoneme and sentence recognition tests. The results suggest that implant users are unable to fully utilize the temporal and the spectral cues. Potential reasons behind this limitation are electrode interaction (Fu *et al.*, 1998), mismatch between acoustic frequency and electric pitch (Dorman *et al.*, 1997b) and the patient's surviving neural population. The present eight-channel CI simulation does not simulate these limitations.

Furthermore, the degree of improvement varied greatly between individual CI users. It appears that subjects with better performance in quiet showed larger improvements (e.g., C3 vs. C6). As a matter of fact, subject C6 did not benefit from companding at all. This subject has been deaf for 50 years, with a severe degree of hearing loss compared to the other subjects. The SNR corresponding to maximum improvement was subject dependent (see Fig. 8). Most implant users did not show any improvement in quiet, likely due to the ceiling effect.

Companding improved vowel recognition in both NH listeners and CI users but improved consonant recognition only in CI users. One particular finding of this study was that apart from improving the spectral contrast, companding also enhances the temporal contrast. This along with the fact that CI users can detect smaller modulation amplitudes than the NH listeners (Shannon, 1992), suggests that the improvement was a result of better detection of the enhanced temporal contrast in consonants. Sequential information analysis (Wang and Bilger, 1973) was performed on the pooled confusion matrix obtained from four NH listeners tested using stimuli with CI simulation. Unfortunately, the confusions matrices from CI users and the rest of NH subjects were not saved. Results indicated that transmission of information about plosive and nasal was higher for stimuli with companding compared to stimuli without companding. Similar effects were observed with the temporal enhancement technique known as transient emphasis spectral maxima (Vandali, 2001) which amplifies the low-intensity short-duration rapid changes in temporal envelope. It is also likely that the implant subjects were able to use the temporal envelope information more efficiently when the background noise was suppressed with companding. The effect of enhanced temporal contrast needs to be investigated by future studies.

D. Applications

The present study showed that companding improved CI speech performance in steady-state noise. It remains to be seen whether companding will work in realistic conditions including multiple-talker babbles as noise. Because of the dynamic changes in both signal and noise in these conditions, it is unlikely that companding will improve the actual SNR to the same degree as in the steady-state noise condition. However, the enhanced spectral and temporal contrasts between signal and the noise (another talker or masker) may help in perceptual segregation of the signal from noise.

Companding can be incorporated into cochlear implants by using either the present front-end approach or an integrated signal processing strategy. In order for it to act as the latter, simultaneous control of compression, noise suppression and channel gain should be feasible. The processing architecture and parameter optimization can be significantly different depending on the approach taken. Different designs may have to be considered in real-time implementations of companding with these different approaches.

Finally, companding may be used to enhance hearing aid performance. It has been shown that multi-channel compression in hearing aids causes spectral smearing (Plomp, 1988). Spectral enhancement using companding has the potential to remedy this problem. Future studies are needed to examine the effectiveness of companding in realistic listening conditions and its utility in assistive listening devices.

VI. CONCLUSIONS

This study optimized implementation of the companding strategy and evaluated its effectiveness in enhancing speech performance in noise for CI users. We found that choice of parameters was critical to produce optimal spectral enhancement and adequate noise suppression. We also found that companding enhanced the temporal contrast. Once the parameters were optimized, the implemented strategy significantly improved the CI speech perception in noise. Future studies are needed to evaluate the relative contribution of spectral and temporal contrast enhancement, the effectiveness of companding under realistic listening conditions (e.g., multiple talkers), and implementation and optimization of companding in cochlear implants and hearing aids.

ACKNOWLEDGMENTS

The authors would like to thank all the subjects who participated in the experiments. This work was supported by NIH (2RO1 DC002267).

Baer, T., Moore, B. C., 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.

Blackburn, C. C., and Sachs, M. B. (**1990**). "The representation of the steady-state vowel /eh/ in the discharge patterns of cat anteroventral cochlear nucleus neurons," J. Neurophysiol. **63**, 1191–1212.

Boothroyd, A., Mulhearn, B., Gong, J., and Ostroff, J. (**1996**). "Effects of spectral smearing on phoneme and word recognition," J. Acoust. Soc. Am. **100**, 1807–1818.

Bunnell, H. T. (**1990**). "On enhancement of spectral contrast in speech for hearing-impaired listeners," J. Acoust. Soc. Am. **88**, 2546–2556.

Clarkson, P., and Bahgat, Sayed F. (**1991**). "Envelope expansion methods for speech enhancement," J. Acoust. Soc. Am. **89**, 1378–1382.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997a**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**,

2403–2411.
Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997b**). "Simulating the effect of cochlear-implant electrode insertion depth on speech understanding," J. Acoust. Soc. Am. **102**, 2993–2996.
Franck, B. A. M., van Kreveld-Bos, C. S. 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**, 1452–1468.
Fu, Q. J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," Trans.- Geotherm. Resour. Counc. **6**, 19–27.
Fu, Q. J., Shannon, R. V., and Wang, X. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.
Hamacher, V., Doering, W. H., Mauer, G., Fleischmann, H., and Hennecke, J. (**1997**). "Evaluation of noise reduction systems for cochlear implant users in different acoustic environment," Am. J. Otol. **18**, S46–49.
Hartline, H. K. (**1969**). "Visual receptors and retinal interaction," Science **164**, 270–278.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.
Hochberg, I., Boothroyd, A., Weiss, M., and Hellman, S. (**1992**). "Effects of noise and noise suppression on speech perception by cochlear implant users," Ear Hear. **13**, 263–271.
Hochmair-Desoyer, I. J., Hochmair, E. S., Burian, K., and Fischer, R. E. (**1981**). "Four years of experience with cochlear prosthesis," Med. Prog. Technol. **8**, 107–119.
Horst, J. W. (**1987**). "Frequency discrimination of complex signals, frequency selectivity, and speech perception in hearing-impaired subjects," J. Acoust. Soc. Am. **82**, 874–885.
Kiang, N. Y., 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.
Loizou, P. (**2005**). "Evaluation of the companding and other strategies for noise reduction in cochlear implants," *2005 Conference on Implantable Auditory Prosthesis*, Asilomar, Monterey, California.
Loizou, P. C., Dorman, M., and Fitzke, J. (**2000**). "The effect of reduced dynamic range on speech understanding: Implications for patients with cochlear implants," Ear Hear. **21**, 25–31.
Loizou, P. C., and Poroy, O. (**2001**). "Minimum spectral contrast needed for vowel identification by normal hearing and cochlear implant listeners," J. Acoust. Soc. Am. **110**, 1619–1627.
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. (**1999a**). "Discriminability of vowel representations in cat auditory-nerve fibers after acoustic trauma," J. Acoust. Soc. Am. **105**, 311–325.
Miller, R. L., Calhoun, B. M., and Young, E. D. (**1999b**). "Contrast enhancement improves the representation of /ɛ/ 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 "eh" in cat auditory nerve fibers," J. Acoust. Soc. Am. **101**, 3602–3616.
Müller-Deiler, J., Schmidt, B. J., and Rudert, H. (**1995**). "Effects of noise on speech descrimination in cochlear implant patients," Ann. Otol. Rhinol. Laryngol. **166**, 303–306.
Nelson, D. A., Schmitz, J. L., Donaldson, G. S., Viemeister, N. F., and Javel, E. (**1996**). "Intensity discrimination as a function of stimulus level with electric stimulation," J. Acoust. Soc. Am. **100**, 2393–2414.
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.
Oxenham, A. J., Simonson, A. M., Turicchia, L., and Sarpeshkar, R. (**2007**). "Evaluation of companding-based spectral enhancement using simulated cochlear-implant processing," J. Acoust. Soc. Am. **121**, 1709–1716.
Plomp, R. (**1988**). "The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function," J. Acoust. Soc. Am. **83**, 2322–2327.
Rhode, W. S. (**1974**). "Evidence from Mässbauer experiments for nonlinear vibrations in the cochlea," J. Acoust. Soc. Am. **55**, 588–597.
Rhode, W. S., Geisler, C. D., and Kennedy, D. T. (**1978**). "Auditory nerve fiber responses to wide-band noise and tone combinations," J. Neurophysiol. **41**, 692–704.
Rhode, W. S., and Greenberg, S. (**1994**). "Lateral suppression and inhibition in the cochlear nucleus of the cat," J. Neurophysiol. **71**, 493–514.
Ruggero, M. A., Robles, L., and Rich, N. C. (**1992**). "Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression," J. Neurophysiol. **68**, 1087–1099.
Sachs, M. B., and Kiang, N. Y. (**1968**). "Two-tone inhibition in auditory-nerve fibers," J. Acoust. Soc. Am. **43**, 1120–1128.
Sachs, M. B., Voigt, H. F., and Young, E. D. (**1983**). "Auditory nerve representation of vowels in background noise," J. Neurophysiol. **50**, 27–45.
Shannon, R. V. (**1992**). "Temporal modulation transfer functions in patients with cochlear implants," J. Acoust. Soc. Am. **91**, 2156–2164.
Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., and Wang, X. (**1999**). "Consonant recordings for speech testing," J. Acoust. Soc. Am. **106**, L71–L74.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.
Stoop, R., and Kern, A. (**2004**). "Essential auditory contrast-sharpening is preneuronal," Proc. Natl. Acad. Sci. U.S.A. **101**, 9179–9181.
Toledo, F., Loizou, P., and Lobo, A. (**2003**). "Subspace and envelope subtraction algorithms for noise reduction in cochlear implants," *Proceedings of 25th Annual International Conference of IEEE-EMBC*, Cancun, Mexico, pp. 2002–2005.
Turicchia, L., and Sarpeshkar, R. (**2005**). "A bio-inspired companding strategy for spectral enhancement," IEEE Trans. Acoust., Speech, Signal Process. **13**, 243–253.
Vandali, A. E. (**2001**). "Emphasis of short-duration acoustic speech cues for cochlear implant users," J. Acoust. Soc. Am. **109**, 2049–2061.
van Hoesel, R. J., and Clark, G. M. (**1995**). "Evaluation of a portable two-microphone adaptive beamformer speech processor with cochlear implants patients," J. Acoust. Soc. Am. **97**, 2498–2503.
Wang, M. D., and Bilger, R. C. (**1973**). "Consonant confusions in noise: A study of perceptual features," J. Acoust. Soc. Am. **54**, 1248–1266.
Weiss, M. R. (**1993**). "Effects of noise and noise reduction processing on the operation of the Nucleus-22 cochlear implant processor," J. Rehabil. Res. Dev. **30**, 117–128.
Wouters, J., and Vanden Berghe, J. (**2001**). "Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system," Ear Hear. **22**, 420–430.
Zeng, F. G., and Galvin, J. J. (**1999**). "Amplitude mapping and phoneme recognition in cochlear implant listeners," Ear Hear. **20**, 60–74.
Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (**2005**). "Speech recognition with amplitude and frequency modulations," Proc. Natl. Acad. Sci. U.S.A. **102**, 2293–2298.
Zeng, F. G., and Shannon, R. V. (**1994**). "Loudness-coding mechanisms inferred from electric stimulation of the human auditory system," Science **264**, 564–566.
Zeng, F. G., and Shannon, R. V. (**1999**). "Psychophysical laws revealed by electric hearing," NeuroReport **10**, 193l–1935.

A two-layer composite model of the vocal fold lamina propria for fundamental frequency regulation

Kai Zhang and Thomas Siegmund[a)]
School of Mechanical Engineering, Purdue University, 585 Purdue Mall, West Lafayette, Indiana 47907

Roger W. Chan
Otolaryngology—Head and Neck Surgery, and Biomedical Engineering, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9035

(Received 30 June 2006; revised 21 May 2007; accepted 23 May 2007)

The mechanical properties of the vocal fold lamina propria, including the vocal fold cover and the vocal ligament, play an important role in regulating the fundamental frequency of human phonation. This study examines the equilibrium hyperelastic tensile deformation behavior of cover and ligament specimens isolated from excised human larynges. Ogden's hyperelastic model is used to characterize the tensile stress-stretch behaviors at equilibrium. Several statistically significant differences in the mechanical response differentiating cover and ligament, as well as gender are found. Fundamental frequencies are predicted from a string model and a beam model, both accounting for the cover and the ligament. The beam model predicts nonzero F_0 for the unstretched state of the vocal fold. It is demonstrated that bending stiffness significantly contributes to the predicted F_0, with the ligament contributing to a higher F_0, especially in females. Despite the availability of only a small data set, the model predicts an age dependence of F_0 in males in agreement with experimental findings. Accounting for two mechanisms of fundamental frequency regulation—vocal fold posturing (stretching) and extended clamping—brings predicted F_0 close to the lower bound of the human phonatory range. Advantages and limitations of the current model are discussed. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2749460]

PACS number(s): 43.70.Aj, 43.70.Bk [AL] Pages: 1090–1101

I. INTRODUCTION

The fundamental frequency of vocal fold vibration (F_0) is a central characteristic of human phonation, and is the primary variable affecting vocal pitch. An understanding of the dependence of F_0 on age and gender is thus of significant interest. Furthermore, this issue is of importance for the development of realistic computer models of phonation or the development of biomaterials for surgical applications.[1] Intrinsically, the fundamental frequency of phonation is dependent on the mechanical properties of the vocal fold lamina propria, including the vocal fold cover, i.e., the epithelium and the superficial layer of the lamina propria, and the vocal ligament, i.e., the intermediate layer and the deep layer of the lamina propria.

Constitutive models that can reliably describe the equilibrium stress-strain or stress-stretch response of components of the vocal fold lamina propria are critical to the prediction of the equilibrium F_0. Experimental data on the tensile stress-stretch response of human vocal fold cover and vocal ligament specimens have been obtained, clearly demonstrating a nonlinear relationship between stress and stretch.[1,2] As in our previous study,[3] Ogden's hyperelastic model[4] is applied to characterize the mechanical response of tissue specimens. Having the empirical data for both the cover and the ligament described by the same model allows for an examination of the differences in constitutive parameters between the cover and the ligament. Furthermore, age- and gender-related differences of the tissue response can be investigated.

The ideal string model[3,5] is a commonly used model for the prediction of F_0. This model considers a structural vibration of the vocal fold lamina propria in dependence of vocal fold length and tension. Alternatively, vocal fold vibration has been analyzed by beam models. In a beam model not only the tensile stiffness but also the bending stiffness is accounted for. Thus, unlike for the string model, the spatial direction of vibration relative to the beam axis (the anterior-posterior direction) has to be specified. Beam models[6–8] are of interest since, in general, string models tend to underpredict the fundamental frequency when compared to empirical speaking F_0 data.[9,10] Titze and Hunter[6] developed a beam model accounting for the vocal ligament as the dominant load-carrying component of the lamina propria, and idealized the tissue mechanical response as linear elastic. Descout *et al.*[7] developed a multilayer beam model also with linear elastic properties. Bickley[8] considered a homogeneous beam and developed a model without accounting for the effects of vocal fold stretch. It is to be noted that both the string and the beam models consider structural vibration as opposed to the mucosal wave model that considers phonation as the propagation of shear waves on the vocal fold surface in an inferior-to-superior direction.[5]

The present study describes enhanced string and beam models by accounting for the layered structure of the lamina propria. Both models account for the changes in cross-

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

 0001-4966/2007/122(2)/1090/12/$23.00

TABLE I. Summary of the constitutive model parameters and the geometrical properties of all specimens; μ_j—initial equilibrium shear modulus; α_j—dimensionless power (nonlinearity of the elastic response); $A_{0,j}$—initial cross-section area of the vocal fold cover or the vocal ligament; L—*in situ* length of the vocal fold; (c) —vocal fold cover; (l) —vocal ligament.

Gender	Age	μ_j [kPa]	α_j	$A_{0,j}$ [mm^2]	L [mm]	Gender	Age	μ_j [kPa]	α_j	$A_{0,j}$ [mm^2]	L [mm]
Male	17 (c)	1.37	8.8	14.55	14.1	Female	46 (l)	2.61	16.0	4.29	16.5
	19 (l)	1.37	11.0	10.46	16.3		54 (l)	2.28	15.5	9.14	16.9
	33 (c)	3.50	14.1	11.47	17.9		73 (c)	0.57	16.0	8.94	16.9
	33 (l)	4.40	15.7	8.09	17.9		73 (l)	0.64	14.3	6.07	16.9
	49A (l)	1.41	19.2	13.34	24.9		80 (c)	4.12	14.5	7.83	14.5
	49B (l)	0.80	20.0	13.49	19.8		80 (l)	1.18	15.5	8.77	14.5
	51 (c)	1.77	14.6	13.85	21.4		82 A (c)	3.63	15.4	7.49	14.9
	51 (l)	1.39	16.0	16.46	21.4		82A (l)	10.46	13.0	6.39	14.9
	54 (l)	1.64	14.7	5.25	20.5		82B (c)	0.43	19.2	8.47	14.6
	65 (c)	6.67	16.1	8.20	20.5		82B (l)	0.23	16.9	6.56	14.6
	65 (l)	3.48	24.7	6.08	20.5		83 (c)	1.31	19.5	6.89	15.1
	66 (c)	5.60	14.5	9.80	20.4		83 (l)	0.21	19.0	5.31	15.1
	66 (l)	1.22	24.5	11.90	20.4		85 (c)	1.43	16.8	8.23	12.4
	79 (l)	1.42	20.0	4.23	22.0		97 (c)	1.77	12.5	9.53	13.9
	88 (l)	3.26	19.2	5.18	20.7						
	99 (c)	6.39	14.4	7.31	17.9						
	99 (l)	1.51	22.0	9.43	17.9						

section areas with tensile deformation. Specifically, the beam model is developed for vibrations in the medial-lateral direction, i.e., the predominant direction of vocal fold vibration. While a string model will always predict zero F_0 in an unstretched state, the beam model as developed in this study can be used to predict fundamental frequencies for unstretched vocal fold.[11] Past beam models[6] based on Ref. 12 did not provide a solution for the unstretched state, but rather predicted infinite F_0 at that state. Numerical solutions to the beam vibration equations are compared to two closed form solutions, i.e., the solution for the unstretched state and the solution for the stretched state. In the present study, the fundamental frequency models are combined with the constitutive models characterizing the experimental stress-stretch response of both vocal fold cover and vocal ligament specimens.

The fundamental frequency models allow for investigations into the regulation of F_0. While aspects of neuromuscular control and sensorimotor feedback certainly contribute to the regulation of F_0,[13–15] the present study focuses on the effects of vocal fold length change on fundamental frequency. Two mechanisms are considered. The process of vocal fold posturing, i.e., the length changes due to activities of the cricothyroid muscle and the thyroarytenoid muscle,[5,16] not only modifies the vocal fold length but simultaneously changes the stiffness of the tissue due to the nonlinear stress-stretch response to tension.[1] In addition, F_0 could be changed by a process referred to as extended clamping,[6,17,18] where the posterior portions of the membranous vocal folds at the vocal processes are presumably pressed or "clamped" together by the arytenoid cartilages such that vibration over this portion is inhibited. This mechanism could become effective once vocal folds have been stretched,[19] and may reduce the effective length of the vocal fold in the stretched state.

In summary, we hypothesize that:

(1) The mechanical response of vocal fold tissue can be described by a hyperelastic constitutive model, and that the parameter values of the constitutive model can reveal statistically significant differences in tissue response depending on tissue type, gender and age;
(2) The two-layer composite beam model in conjunction with realistic material parameters can reveal the contributions of the vocal fold cover and the vocal ligament to F_0, with the model quantifying the effect of bending stiffness relative to tensile stiffness; and that fundamental frequencies can be calculated for unstretched vocal folds as well as for vocal folds subjected to posturing (stretching) and extended clamping;
(3) This modeling approach allows for predictions of gender- and age-related differences in F_0, and that it can be established whether these differences result from the geometrical features or the mechanical response of the tissue components in the vocal fold lamina propria.

II. METHODS

A. Measurements of tensile mechanical response of the vocal fold

The passive uniaxial tensile stress-stretch response of the vocal fold cover and the vocal ligament was measured by sinusoidal stretch-release deformation (loading-unloading), with the use of a dual-mode servo control lever system (Aurora Scientific Model 300B-LR, Aurora, ON, Canada).[1,3] Measurements of the displacement and force of the lever arm were made by the servo control lever system with a displacement accuracy of 1.0 μm and a force resolution of 0.3 mN. The servo control lever system possessed a displacement range of up to 8–9 mm in the frequency range of 1–10 Hz. Specimens were stretched to a fixed maximum length ℓ_{rev} at load reversal. The uniaxial stretch λ_u at load reversal is defined as $\lambda_{u,\text{rev}}, = \ell_{\text{rev}}/L$ with L being the *in situ* length, i.e.,

the initial length of the specimen. The experimental protocol was approved by the Institutional Review Board of UT Southwestern Medical Center.

Vocal fold cover and vocal ligament were dissected with instruments for phonomicrosurgery.[1] Specimens were dissected from 21 larynges excised within 24 h postmortem, procured from autopsy from human cadavers free of head and neck disease and laryngeal pathologies. All subjects were nonsmokers, and were Caucasians or Hispanics, although race was not a factor in the procurement. Tissue specimens were obtained from 12 male subjects (Table I), of which vocal fold cover specimens were obtained from six, whereas vocal ligament specimens were obtained from all but the youngest subject (age $Y=17$). Specimens were obtained from nine female subjects (Table I), of which vocal fold cover specimens were obtained from seven. Vocal ligament specimens were also obtained from seven subjects. Before dissection, the *in situ* vocal fold length, defined as the distance from the anterior commissure to the vocal process, i.e., the membranous vocal fold length in a relaxed, cadaveric state, was measured by digital calipers for each specimen. Throughout the dissection, each specimen remained attached to small portions of the thyroid and the arytenoid cartilages, allowing for the attachment of sutures for mechanical testing under natural boundary conditions. Exploiting the displacement range of the lever system, male specimens were loaded to stretches of up to $\lambda_{u,\mathrm{rev}}=1.35$ while female specimens were loaded to stretches of up to $\lambda_{u,\mathrm{rev}}=1.6$.

The experimental equilibrium response is obtained from the measured hysteretic tensile stress-stretch curves as the midpoint values of stresses at equal stretch for the loading and unloading portions of the hysteresis loop in the stabilized (preconditioned) state.[3]

B. Constitutive model for equilibrium response

The equilibrium response is characterized by a nonlinear hyperelastic constitutive model. The vocal fold lamina propria has been recognized as viscoelastic and anisotropic with a primarily parallel arrangement of collagen and elastin fibers, particularly for the vocal ligament.[20,21] In the present study, only the elastic tissue response in the anterior-posterior direction is of concern and effects of anisotropy are thus not considered. A first-order Ogden model[4] together with the assumption of incompressible material response allows for an appropriate description of both the vocal fold cover and the vocal ligament. The first-order Ogden model possesses two parameters, the initial shear modulus μ describing the tissue stiffness, and the power α describing the nonlinearity of the elastic response. It has been shown to be promising in characterizing the tensile behavior of other soft tissues.[22,23] This model is described by a strain energy density function w of the form[4]

$$w=\frac{2\mu}{\alpha^2}(\lambda_1^{\alpha}+\lambda_2^{\alpha}+\lambda_3^{\alpha}-3). \tag{1}$$

Deformation is characterized through the principal stretches λ_1, λ_2, λ_3 which for incompressibility satisfy $\lambda_1\lambda_2\lambda_3=1$. For uniaxial loading of a specimen of *in situ* length L to a current length ℓ the principal stretches are $\lambda_1=\lambda_u$, $\lambda_2=\lambda_3=1/\sqrt{\lambda_u}$ where $\lambda_u=\ell/L$ is the stretch in the anterior-posterior direction. The nominal stress $\sigma_{\mathrm{nominal}}$ in the equilibrium response is obtained as derivative of the strain energy density w with respect to the applied stretch λ_u

$$\sigma_{\mathrm{nominal}}=\frac{\partial w}{\partial \lambda_u}=\frac{2\mu}{\alpha}\left(\lambda_u^{\alpha-1}-\lambda_u^{-\alpha/2-1}\right). \tag{2}$$

Under uniaxial tension, the Cauchy stress (true stress) σ can be expressed as the product of the nominal stress and the stretch in the longitudinal direction

$$\sigma=\lambda_u\cdot\frac{\partial w}{\partial \lambda_u}=\frac{2\mu}{\alpha}\left(\lambda_u^{\alpha}-\lambda_u^{-\alpha/2}\right). \tag{3}$$

The tangent Young's modulus E_t, i.e., the instantaneous stiffness at a given level of stretch, is obtained by differentiating the Cauchy stress σ with respect to the stretch λ_u

$$E_t=\frac{\partial \sigma}{\partial \lambda_u}=\mu\left(2\lambda_u^{\alpha-1}+\lambda_u^{-\frac{\alpha}{2}-1}\right). \tag{4}$$

For the stretch $\lambda_u=1$ and incompressible material response, the tangent Young's modulus $E_t=3\mu$.

The hyperelastic model is applied to characterize the equilibrium uniaxial tensile response of each specimen. A unique set of parameters μ and α can be determined for each specimen through curve fitting of Eq. (2) by achieving a local minimum of the sum of squares of the differences between the experimental data and the simulation results through the Levenberg–Marquardt method,[24,25] under the condition that both μ and α are positive.

C. Models of fundamental frequency regulation

1. Composite string model

Similar to our previous study,[3] the investigation of the fundamental frequency of phonation begins with a string model of phonation.[5] However, here the string is composed of the vocal fold cover and the vocal ligament. Figure 1(a) depicts a cross section of a typical vocal fold with the total initial cross-section area A_0 composed of the cover $A_{0,c}$ and the ligament $A_{0,l}$. The length of vocal fold cover and vocal ligament are assumed to be identical at all times. Thus, the longitudinal stretch of the vocal fold cover $\lambda_{u,c}$ and the vocal ligament $\lambda_{u,l}$ are identical

$$\lambda_u=\lambda_{u,c}=\lambda_{u,l}. \tag{5}$$

The partial differential equation governing the vibration of a flexible string with two fixed ends is[5]

$$T\frac{\partial^2 v}{\partial x^2}=\rho A\frac{\partial^2 v}{\partial t^2}, \tag{6}$$

where ρ is the density, T is the tensile force applied to the composite string and A is the current total cross-section area of the string, and v, the displacement from the equilibrium position (in the medial-lateral direction) of an infinitesimal part of the string with a distance x (anterior-posterior) from a

string end point, is a function of x and time t, $v=v(x,t)$. Equation (6) can be simplified as

$$\bar{\sigma}\frac{\partial^2 v}{\partial x^2}=\rho\frac{\partial^2 v}{\partial t^2}, \tag{7}$$

where $\bar{\sigma}=T/A$ is the average, or homogenized longitudinal Cauchy stress. This quantity is related to the Cauchy stress in the vocal fold cover σ_c and the vocal ligament σ_l through

$$\bar{\sigma}=f_c\cdot\sigma_c+f_l\cdot\sigma_l, \tag{8}$$

where f_c and f_l are the current area ratios of the cover and the ligament, respectively, $f_j=A_j/(A_l+A_c)$, with the subscript j denoting cover ($j=c$) or ligament ($j=l$). The area ratios are assumed to be constant along the longitudinal axis of the vocal fold, as well as during the elongation process, $f_{0,j}=A_{0,j}/(A_{0,l}+A_{0,c})=f_j$. Assuming perfect bonding between the vocal fold cover and the vocal ligament and enforcing equilibrium between applied force and internal force resultants in the cover and the ligament, the Cauchy stresses in the cover and the ligament σ_c and σ_l can be calculated in dependence of the stretch λ_u and the constitutive parameters through Eq. (3).

The lowest allowed frequency, or the fundamental frequency for the string model F_0^{string}, can be expressed as a function of the current vocal fold length ℓ, the tissue density ρ, and the tissue homogenized longitudinal Cauchy stress $\bar{\sigma}$

$$F_0^{\text{string}}=\frac{1}{2\ell}\sqrt{\frac{\bar{\sigma}}{\rho}}. \tag{9}$$

The fundamental frequency is thus dependent on the magnitude of deformation applied as a result of vocal fold length changes due to posturing. Predictions of fundamental frequency can be obtained from Eq. (9) with use of the constitutive model once the material parameters are determined. For the hyperelastic tissue response the combination of Eqs. (3), (8), and (9) leads to a prediction of F_0^{string} in dependence of the constants of the hyperelastic constitutive model and stretch λ_u as

$$F_0^{\text{string}}=\frac{1}{L\sqrt{2\rho}}\sqrt{\frac{\mu_c f_c}{\alpha_c}(\lambda_u^{\alpha_c-2}-\lambda_u^{-\alpha_c/2-2})+\frac{\mu_l f_l}{\alpha_l}(\lambda_u^{\alpha_l-2}-\lambda_u^{-\alpha_l/2-2})}. \tag{10}$$

The fundamental frequency is thus dependence of the hyperelastic response and the area ratios of the cover and the ligament, the *in situ* vocal fold length, tissue density, and stretch. It should be noted here that Eqs. (9) and (10) are developed by use of the Cauchy stress, a factor not considered in our previous study[3] and other past work, e.g., Refs. 5–7.

2. Composite beam model

A beam model is proposed with the beam cross section accounting for the presence of the vocal fold cover and the vocal ligament. Following Fig. 1(b), the cover and the ligament are both approximated by rectangular shapes. The line connecting the geometrical center of the idealized vocal fold cover to that of the idealized vocal ligament is in the medial-lateral direction. The idealized rectangular cross-section dimensions are h_j and b_j as the thickness and transverse depth of the vocal fold cover and the vocal ligament, respectively. The aspect ratio m_j is defined as $m_j=h_j/b_j$ for the two cross-section components, Fig. 1(b). Here the aspect ratios are assumed to be identical for both the cover and the ligament, $m=m_c=m_l$.

When the composite beam is under tension the partial differential equation describing its vibrations becomes[12]

$$T\frac{\partial^2 v}{\partial x^2}-\bar{E}_t I\frac{\partial^4 v}{\partial x^4}=\rho A\frac{\partial^2 v}{\partial t^2}, \tag{11}$$

where $\bar{E}_t$, is the composite tangent Young's modulus and I is the area moment of inertia of the composite cross section. Equation (11) can be written into the following form:

$$\bar{\sigma}\frac{\partial^2 v}{\partial x^2}-\bar{E}_t\kappa^2\frac{\partial^4 v}{\partial x^4}=\rho\frac{\partial^2 v}{\partial t^2}, \tag{12}$$

where $\kappa=\sqrt{I/A}$ is the radius of gyration of the composite cross section. The value of $\bar{E}_t\kappa^2$ is a measure of the bending stiffness for the composite beam. It is determined from the following consideration. Following standard derivations of Euler Bernoulli beam theory for layered beams with rectangular cross-section geometry,[26] the bending moment of the vocal fold can be written as

$$M=\bar{E}_t I\frac{\partial^2 v}{\partial x^2}=\frac{A^2(E_{t,c}^2 f_c^3+E_{t,l}^2 f_l^3+6E_{t,c}E_{t,l}\sqrt{f_c^3 f_l^3}+4E_{t,c}E_{t,l}f_c f_l)}{12m(E_{t,c}f_c+E_{t,l}f_l)}\frac{\partial^2 v}{\partial x^2}, \tag{13}$$

where $E_{t,c}$, and $E_{t,l}$ are the tangent Young's moduli of the cover and the ligament given by Eq. (4), respectively. Under the assumption of incompressibility, the current cross-section area A is related to the initial cross-section area A_0 through $A=A_0/\lambda_u$. $\bar{E}_t\kappa^2$ then can be obtained as

$$\bar{E}_t\kappa^2 = \frac{\bar{E}_t I}{A}$$

$$= \frac{A_0(E_{t,c}^2 f_c^3 + E_{t,l}^2 f_l^3 + 6E_{t,c}E_{t,l}\sqrt{f_c^3 f_l^3} + 4E_{t,c}E_{t,l}f_c f_l)}{12m\lambda_u(E_{t,c}f_c + E_{t,l}f_l)} \tag{14}$$

In Eq. (12), v is the time varying displacement from the equilibrium position in the medial-lateral direction, and is expressed as $v(x,t)=\xi(x)\cdot e^{-2\pi Ft}$ with F being the frequency of vibration, and $i=\sqrt{-1}$. Clamped boundary conditions at the vocal process (the posterior end point, $x=0$) and at the anterior commissure (the anterior end point, $x=\ell$) are assumed. The clamped boundary conditions imply that both the displacement $\xi(x)$ and its derivative $\partial\xi/\partial x$ need to vanish at the boundaries

$$\xi(0) = \xi(\ell) = 0$$

$$\left.\frac{\partial\xi}{\partial x}\right|_{x=0} = \left.\frac{\partial\xi}{\partial x}\right|_{x=\ell} = 0. \tag{15}$$

Substituting the presumed form of solutions of $v(x,t)$ into Eq. (12), one obtains

$$\frac{d^4\xi}{dx^4} - 8\pi^2\beta^2\frac{d^2\xi}{dx^2} - 16\pi^4\gamma^4\xi = 0 \tag{16}$$

with

$$\beta^2 = \frac{\bar{\sigma}}{8\pi^2\bar{E}_t\kappa^2} \quad \gamma^4 = \frac{\rho F^2}{4\pi^2\bar{E}_t\kappa^2}. \tag{17}$$

The parameter β characterizes the ratio between restoring forces due to tension and bending, while γ quantifies the ratio between inertia and bending. The solution to this ordinary differential equation [Eq. (16)] is

$$\xi(x) = D_1\cosh(2\pi\psi x) + D_2\sinh(2\pi\psi x) + D_3\cos(2\pi\zeta x) + D_4\sin(2\pi\zeta x) \tag{18}$$

in which

$$\psi = [(\beta^4+\gamma^4)^{1/2}+\beta^2]^{1/2} \quad \zeta = [(\beta^4+\gamma^4)^{1/2}-\beta^2]^{1/2}. \tag{19}$$

Applying the boundary conditions Eq. (15) to Eq. (18) yields

$$D_1 + D_3 = 0$$

$$D_1\cosh(2\pi\psi\ell) + D_2\sinh(2\pi\psi\ell) + D_3\cos(2\pi\zeta\ell) + D_4\sin(2\pi\zeta\ell) = 0$$

$$\psi D_2 + \zeta D_4 = 0$$

$$\psi D_1\sinh(2\pi\psi\ell) + \psi D_2\cosh(2\pi\psi\ell) - \zeta D_3\sin(2\pi\zeta\ell) + \zeta D_4\cos(2\pi\zeta\ell) = 0. \tag{20}$$

The following characteristic equation is derived by eliminating D_1 through D_4 from Eq. (20) with the help of the relationship $\psi^2-\zeta^2=2\beta^2$

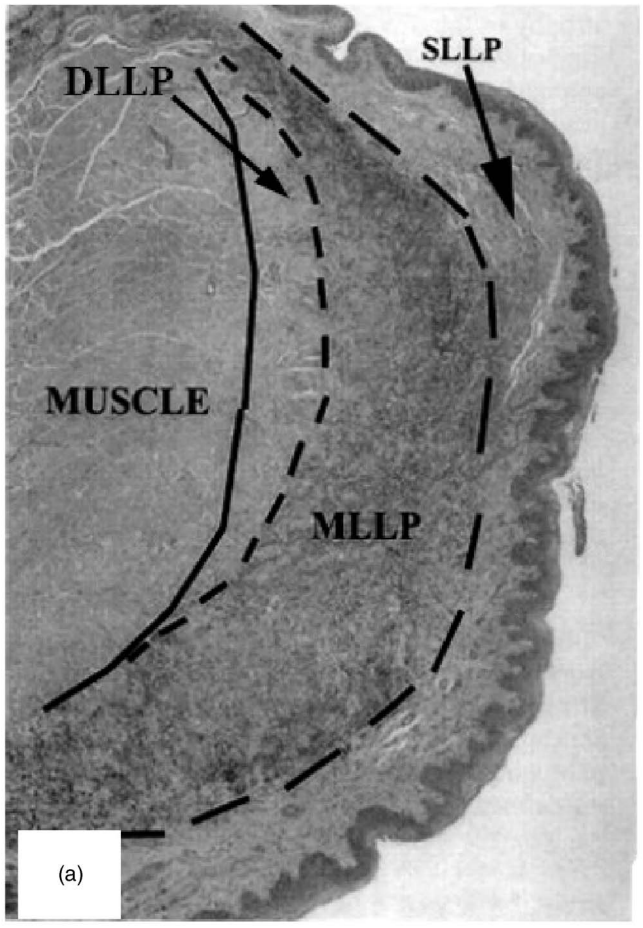


FIG. 1. (a) Layered structure of the human vocal fold illustrating the vocal fold cover, or superficial layer of the lamina propria (SLLP), and the vocal ligament, or middle and deep layers of the lamina propria (MLLP and DLLP) (mid-membranous vocal fold coronal section of a 43-year-old male stained for elastin); from Gray *et al.*[30] (reproduced with permission from Annals of Otology, Rhinology and Laryngology); (b) the geometrical approximation of the cross section of the two-layer composite beam model of the lamina propria with $A_{0,\ell}$ and $A_{0,c}$ indicating the cross-section areas of the ligament and the cover, respectively.

$$\tan(\pi\ell\zeta) = -\sqrt{1+\left(\frac{2\beta^2}{\zeta^2}\right)}\tanh(\pi\ell\sqrt{\zeta^2+2\beta^2}). \tag{21}$$

The fundamental frequency, i.e., the lowest allowed frequency of vibration F, can then be obtained from Eqs. (17) and (19)

$$F_0 = 2\pi\gamma^2\sqrt{\frac{\bar{E}_t\kappa^2}{\rho}} = 2\pi\sqrt{(\zeta^2+\beta^2)^2-\beta^4}\sqrt{\frac{\bar{E}_t\kappa^2}{\rho}}$$
$$= 2\pi\zeta\sqrt{(\zeta^2+2\beta^2)\frac{\bar{E}_t\kappa^2}{\rho}} \quad (22)$$

with ζ being the lowest positive value satisfying Eq. (21).

When β goes to zero, i.e., when the tension is zero and the vocal fold is at its original length L, the characteristic equation [Eq. (21)] reduces to

$$\tan(\pi L\zeta) + \tanh(\pi L\zeta) = 0 \quad (23)$$

with the lowest positive value $\zeta = 0.7528/L$ and the fundamental frequency F_0 is given from Eqs. (4), (14), and (22) as

$$F_0(\lambda_u = 1) = 2\pi\zeta^2\sqrt{\frac{\bar{E}_t\kappa^2}{\rho}}$$
$$= \frac{1.7804}{L^2}\sqrt{\frac{A_0(\mu_c^2 f_c^3 + \mu_l^2 f_l^3 + 6\mu_c\mu_l\sqrt{f_c^3 f_l^3} + 4\mu_c\mu_l f_c f_l)}{m\rho(\mu_c f_c + \mu_l f_l)}}. \quad (24)$$

When β is large but not infinite, i.e., when tension dominates the restoring force of the oscillating beam, an approximate expression for the lowest allowed value of ζ can be obtained through expanding both sides of Eq. (21) and retaining only the first two terms in the series expansions as in[6,12]

$$\zeta \approx \frac{1}{2\ell}\left[1 + \frac{2}{\pi}\sqrt{B} + \left(\frac{4}{\pi^2} + 0.5\right)B\right] \quad \text{with } B = \frac{\pi^2\bar{E}_t\kappa^2}{\bar{\sigma}L^2\lambda_u^2}. \quad (25)$$

Substituting Eq. (25) into Eq. (22), the fundamental frequency F_0 then becomes

$$F_0 \approx 2\pi\zeta\sqrt{\frac{2\beta^2\bar{E}_t\kappa^2}{\rho}} = \frac{1}{2\ell}\sqrt{\frac{\bar{\sigma}}{\rho}}\left[1 + \frac{2}{\pi}\sqrt{B} + \left(\frac{4}{\pi^2} + 0.5\right)B\right]. \quad (26)$$

In Eq. (26) the parameter B is given from Eqs. (8), (14), and (25) as

$$B = \frac{\pi^2\bar{E}_t\kappa^2}{\bar{\sigma}L^2\lambda_u^2}$$
$$= \frac{\pi^2 A_0(E_{t,c}^2 f_c^3 + E_{t,l}^2 f_l^3 + 6E_{t,c}E_{t,l}\sqrt{f_c^3 f_l^3} + 4E_{t,c}E_{t,l}f_c f_l)}{12m\lambda_u^3 L^2(E_{t,c}f_c + E_{t,l}f_l)(\sigma_c f_c + \sigma_l f_l)} \quad (27)$$

with σ_j and $E_{t,j}$ given by Eqs. (3) and (4), respectively.

In order to demonstrate the validity of Eqs. (24) and (26), numerical solutions to Eq. (21) were obtained through an algebraic equation solver and substituted into Eq. (22) to calculate values of F_0.

III. RESULTS

A. Characterization of the vocal fold cover and the vocal ligament

The geometrical features of the cover and the ligament specimens were obtained after dissection. Table I summarizes the *in situ* vocal fold lengths L and initial cross-section areas of the cover and the ligament in the undeformed state, $A_{0,c}$ and $A_{0,l}$. By definition the *in situ* lengths of the cover and the ligament are identical for each individual subject. Results from Mann-Whitney U tests investigating gender differences and differences between cover and ligament are given in Table II. The results indicated that the female specimens possess shorter *in situ* lengths than the male specimens at a level of statistical significance ($p=0.0002$). The *in situ* lengths of the male specimens were found in dependence of age as $L(\text{male})(Y) = 21.02[1.0 - \exp(-0.07331\cdot Y)]$ ($R^2 = 0.6025$). No age dependence could be determined for females due to the small age range of the samples available.

The cross-section areas of the cover and the ligament were found to be not statistically significantly different for either the male or the female specimens ($p=0.500$ and $p=0.063$, respectively). Gender-related differences in the geometrical characteristics of the specimens were also examined. No statistically significant differences were found for vocal fold cover specimens ($p=0.264$) as well as for vocal ligament specimens ($p=0.204$). In the subsequent analysis a tissue density of $\rho = 1040\ \text{kg/m}^3$ for both the vocal fold cover and the vocal ligament is assumed throughout the study.[20]

Figure 2 shows two examples of the tissue equilibrium response together with the description through Ogden's hyperelastic model. The stress-stretch curves for the two specimens shown are very similar qualitatively but differ significantly in stress level. This strong qualitative similarity is also found if all stress-stretch curves are compared, but there is a significant inter-subject variability that exists. A strong nonlinear dependence of stress on the applied stretch is observed, especially at higher stretch ($\lambda_u \geq 1.2$). This finding is similar to experimental findings on vocal fold tissues.[1–3] Ogden's model can characterize both stress-stretch curves well, despite the large inter-subject differences in the specific stress levels associated with the individual tissue response.

The material parameters μ_j, α_j ($j=c,l$) were determined for all specimens and respective values are given in Table I. First, the vocal fold cover and the vocal ligament are compared to each other. In order to conduct paired statistical tests and to make this comparison meaningful, only those subjects with both cover and ligament specimens available are selected for the analysis, including five male ($Y=33$, 51, 65, 66, 99 years) and five female subjects ($Y=73$, 80A, 82A, 82B, 83 years). For the paired Mann-Whitney U tests, only the differences between the cover and the ligament of the same subject are of concern. Table II summarizes the mean values and standard deviations of the parameters, as well as the results of these statistical tests. For males no statistically significant difference was found between μ_c(male) and μ_l(male)($p=0.093$), while α_c(male) and α_l(male) are statistically significantly different ($p=0.031$). For female subjects

TABLE II. Summary of the results of statistical tests conducted on the constitutive and geometrical parameters for the cover-ligament comparison and the gender difference.

		Cover			Ligament					
Gender	Parameter	Mean	SD	n	Mean	SD	n	U-test p value	Paired test?	Significantly different?
Male	μ [kPa]	4.79	2.09		2.40	1.45		0.093	Yes	No
	α	14.7	0.8	5	20.6	4.4	5	0.031	Yes	Yes
	A_0 [mm^2]	10.13	2.62		10.39	4.00		0.500	Yes	No
Female	μ [kPa]	2.01	1.74		2.54	4.44		0.406	Yes	No
	α	16.9	2.3	5	15.7	2.3	5	0.094	Yes	No
	A_0[mm^2]	7.92	0.81		6.62	1.29		0.063	Yes	No
		Male			Female					
Component	Parameter	Mean	SD	n	Mean	SD	n	U-test p value	Paired test?	Significantly different?
Cover	μ [kPa]	5.11	2.27		1.89	1.44		0.021	No	Yes
	α	14.9	0.8		16.3	2.5		0.158	No	No
	A_0 [mm^2]	9.79	2.90	4	8.20	0.89	7	0.264	No	No
	L [mm]	20.91	1.88		15.10	1.50		0.0002	No	Yes
Ligament	μ [kPa]	1.79	0.93		2.52	3.63		0.235	No	No
	α	20.0	3.4		15.7	1.9		0.006	No	Yes
	A_0 [mm^2]	9.48	4.49	9	6.65	1.75	7	0.204	No	No
	L [mm]	20.91	1.88		15.10	1.50		0.0002	No	Yes

no statistically significant differences between the vocal fold cover and the vocal ligament were observed. Paired Mann-Whitney U tests for parameters μ and α resulted in p-values of 0.406 and 0.094, respectively.

The dependence of the hyperelastic model parameters on gender is investigated for the cover and the ligament separately. In order to examine gender-related differences without the confounding effect of age, the statistical analyses with unpaired U tests were conducted for subjects of a similar age range, with the three youngest male subjects (Y=17, 19, 33) excluded. Table II summarizes the results of the statistical tests on gender difference. For the cover specimens a U test indicates the male-to-female difference in the mean values of equilibrium shear modulus to be statistically significant (p =0.021). A similar analysis was also conducted between α_c(male) and α_c(female) but indicated that differences in these parameters are statistically insignificant (p=0.158). For the ligament specimens the gender-related difference in the mean values of equilibrium shear modulus was not significant (p=0.235), while the difference between α_l(male) and α_l(female) was statistically significant with p=0.006.

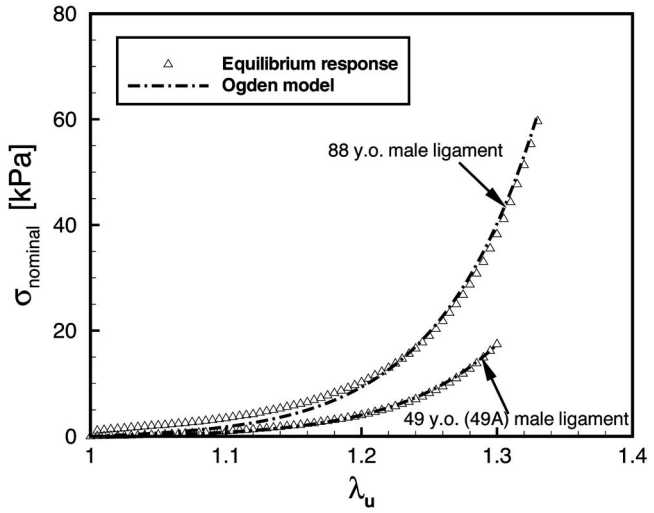


FIG. 2. Comparison between experimental data (the delta-shaped symbols) and simulation results (the dash-dot lines) of tensile equilibrium stress-stretch response at 1 Hz for two ligament specimens (88-year-old male and 49-year-old male).

For male vocal fold cover specimens a relationship between the tissue constitutive parameters and age was established, thereby expanding on previously published data for the male vocal fold cover.[3] The values for μ_c(male) and α_c(male) were expressed as $\mu_c(\text{male})(Y)=7.0[1.0-\exp(-0.01884\cdot Y)]$[kPa] (with the coefficient of determination $R^2=0.62$) and $\alpha_c(\text{male})(Y)=15.35[1.0-\exp(-0.05841\cdot Y)]$ ($R^2=0.87$). The same nonlinear regression analysis was conducted for the constitutive and geometrical parameters of the male ligament specimens, including μ_l(male) and α_l(male). It was found that the nonlinearity of the elastic response α_l(male) depends on age with $\alpha_l(\text{male})(Y)=22.59[1.0-\exp(-0.03521\cdot y)]$ ($R^2=0.58$). No fit describing the age dependence of μ_l(male) could be found, hence in further analysis an average, age-independent value of μ_l(male)=1.99 kPa was employed. Also, no changes in the area ratios f_j with age were observed.

B. Fundamental frequency prediction

Fundamental frequencies are predicted based on the average constitutive and geometrical parameter values for males (with the three youngest subjects excluded for the same reasons stated in Sec. III A) and females. An aspect

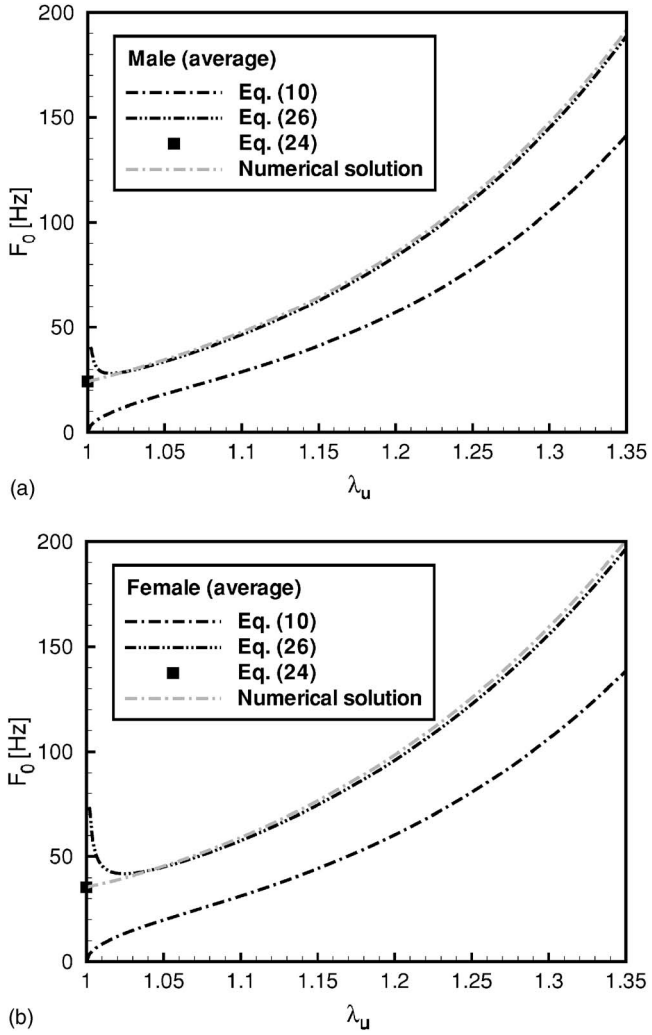


FIG. 3. Dependence of the fundamental frequency F_0 predicted by the string model and the beam model upon the stretch level for (a) the average male vocal fold and (b) the average female vocal fold.

ratio $m=3$ for both the vocal fold cover and the vocal ligament is assumed as an approximation. Following Fig. 1(a) an aspect ratio of $m=3$ is a reasonable geometrical approximation of the layered microstructure of the lamina propria.

Figure 3 summarizes the dependence of the predicted F_0 on vocal fold stretch, based on the following models considered in this study: (i) the composite string model, Eq. (10), (ii) the composite beam model for the unstretched vocal fold, Eq. (24), (iii) the composite beam model for the stretched vocal fold, Eq. (26), and (iv) the numerical solution to the composite beam model. The closed form solution derived from the beam model, Eq. (26), describes the vocal folds in a stretched state. Equation (26) is only valid when the vocal fold is stretched to the extent that tension dominates the restoring force, i.e., when β is much larger than ζ in Eq. (21). Predictions of F_0 through Eq. (26) are thus not valid at small values of stretch. The numerical solution demonstrates the range of validity of Eq. (26). Numerically predicted values of F_0 are found to agree very well with the predicted values of Eq. (24) at $\lambda_u=1.0$, as well as with the values of Eq. (26) for stretch beyond 1.05.

The string model predicts $F_0=0$ for the unstretched state ($\lambda_u=1.0$), whereas the new beam model, Eq. (24), predicts nonzero value of F_0 for the unstretched state: F_0(male) $=24.2$ Hz and F_0(female)$=35.4$ Hz. This unstretched state is, however, not physiologically relevant for phonation, whereas a stretch range of 1.2–1.3 is reasonable in speaking.[27] As the stretch is increased a nonlinear increase in F_0 is predicted. For males, at a stretch of 1.2, a typical magnitude of vocal fold elongation,[16,27] the F_0 predicted by the string model is 57.2 Hz, whereas the composite beam model predicts a fundamental frequency of 83.8 Hz for a stretch of 1.2. For adult males empirical speaking F_0 was found to be between around 90–150 Hz, depending on age, geometrical and anatomical variations, and other factors.[9,10] While the F_0 predicted by the string model is low compared to empirical data, the beam model [Eq. (26)] is capable of predicting F_0 values approaching the phonatory range at a stretch of 1.2. For females, at a stretch of 1.3, a typical magnitude of vocal fold elongation,[16,27] the F_0 predicted by the string model is 106.2 Hz, whereas the composite beam model predicts a fundamental frequency of 159.2 Hz for a stretch of 1.3. For adult females the empirical speaking F_0 ranges from around 150 to 250 Hz.[9,10] While the F_0 predicted by the string model is again low compared to empirical data, the beam model [Eq. (26)] is once again more capable of predicting F_0 values approaching the phonatory range at a stretch of 1.3.

While Fig. 3 considers fundamental frequency regulation through vocal fold stretch, Fig. 4 considers results for F_0 regulation through "extended clamping," i.e., reduction in the effective vocal fold length in the stretched state by a presumed compression of the arytenoid cartilages such that vibration is inhibited over a segment of the membranous vocal fold close to the vocal process. In the present model, the extent of activity of this mechanism is quantified by defining the effective length ratio as the quotient of the length of the freely vibrating vocal fold ℓ_{free} and the current vocal fold length $\ell=L\lambda_u$

$$\Phi=\frac{\ell_{\text{free}}}{\ell}=\frac{\ell_{\text{free}}}{L\lambda_u}. \tag{28}$$

Figure 4 demonstrates the dependence of F_0 on the effective length ratio, for males and females, respectively, at several stretch levels, and compares the predicted F_0 with empirical data. This diagram thus provides maps from which the effectiveness of the combined action of the two mechanisms of fundamental frequency regulation—posturing and extended clamping—can be assessed. For males, the lower bound on empirically obtained fundamental frequencies (F_0 $=90$ Hz) can be reached by $\lambda_u>1.2$ at $\Phi=1.0$ (no extended clamping) or $\lambda_u>1.15$ at $\Phi=0.8$ (as extended clamping is activated). The upper bound on empirically obtained fundamental frequencies ($F_0=150$ Hz), however, would either require considerable stretch $\lambda_u>1.3$ at $\Phi=1.0$, or a combination of $\lambda_u>1.2$ and an activation of extended clamping with $\Phi=0.8$. For females, the lower bound on empirically obtained fundamental frequencies ($F_0=150$ Hz) requires higher stretch than for males, $\lambda_u>1.28$ at $\Phi=1.0$, or $\lambda_u>1.22$ at $\Phi=0.8$. The upper bound on empirically obtained fundamen-

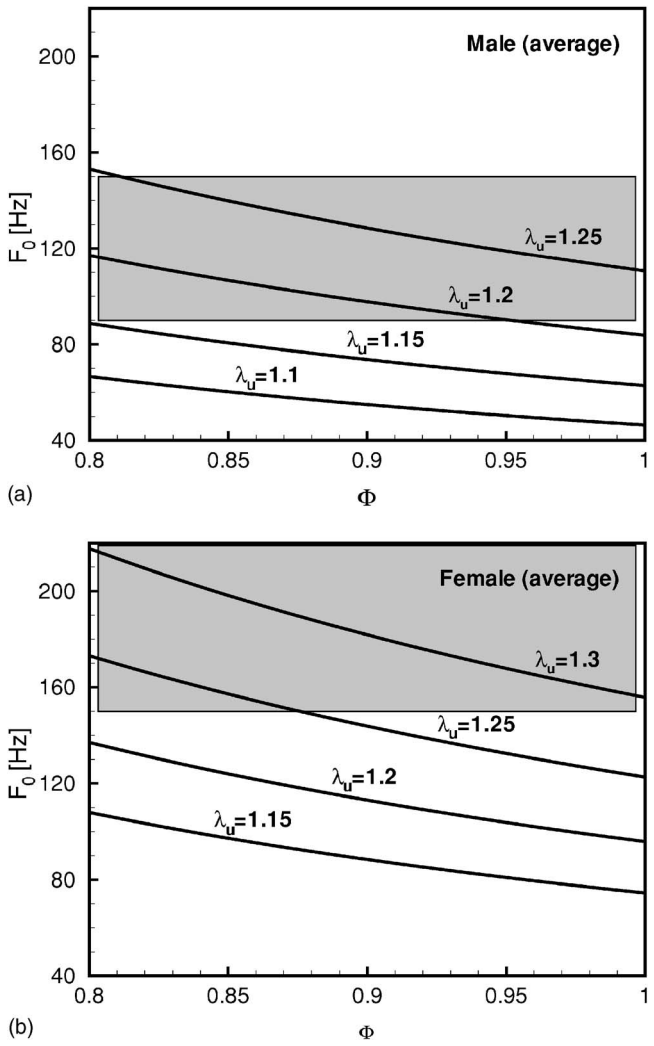


FIG. 4. Dependence of the fundamental frequency F_0 upon the effective length ratio Φ at several stretch levels λ_u for (a) the average male vocal fold and (b) the average female vocal fold. Shaded areas indicate empirical speaking fundamental frequencies.

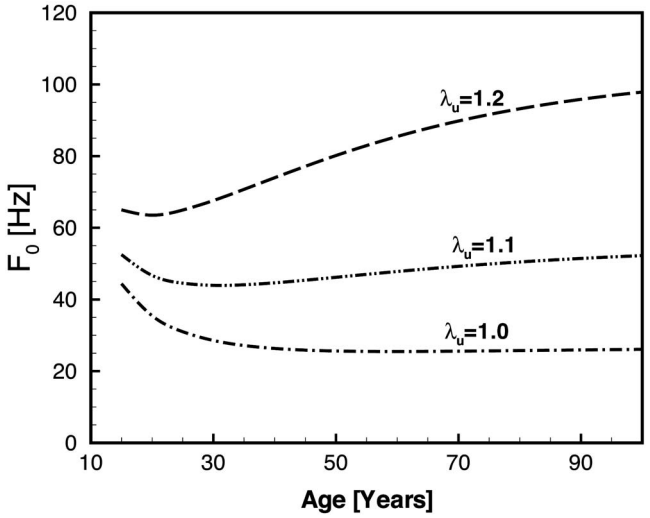


FIG. 5. Fundamental frequency F_0 predicted by the two-layer beam model for males in dependence of age at different stretch levels.

tal frequencies (F_0=220 Hz) would require very high stretch $\lambda_u > 1.5$ at Φ=1.0, or a combination of $\lambda_u > 1.3$ and an activation of extended clamping with Φ=0.8.

For male subjects, Eqs. (24) and (26) can be used to examine the dependence of F_0 on age, based upon the age dependence of the constitutive parameters and *in situ* length values as given in Sec. III A. For the cross-section area of vocal ligament and cover no age dependence could be established, and average values of all male subjects were used for these parameters, $\bar{A}_{0,c}$=10.9 mm^2, $\bar{A}_{0,l}$=9.5 mm^2. Figure 5 shows the age dependence of predicted F_0 at three different stretch levels (λ_u=1.0, 1.1 and 1.2). The predictions are characterized by a decrease of F_0 with age in early years until $Y \approx 20$, which is consistent with developmental lifespan changes and could be attributed to the significant increase of vocal fold length during puberty. Subsequently, for increasing age and λ_u=1.1 and 1.2, a gradual increase of F_0 is predicted, also consistent with empirical lifespan changes.[9,10]

In order to find out whether the constitutive parameters (μ_j, α_j) or the geometrical parameters (A_0, L, f_j) play a more significant role in this gender difference, we conduct a numerical experiment. Thereby, we exchange these two sets of parameters between males and females and calculate fundamental frequency values from the composite beam model [Eq. (26)] for two cases: case 1 with a vocal fold with male geometry and female material properties, and case 2 with female geometry and male material properties. At high stretch levels (λ_u=1.2) the fundamental frequencies for the numerical experiment are F_0(case 1)=63.3 Hz and F_0(case 2)=127.9 Hz compared to F_0(male)=83.8 Hz, F_0(female)=95.9 Hz. At high stretch level the change in geometry affects males more (53% increase for F_0(case 2) relative to F_0(male)) than females (33% decrease for F_0(case 1) relative to F_0(female)), but changes in material properties are more significant for females (33% increase for F_0(case 2) relative to F_0(female)) than for males (25% decrease in F_0(case 1) relative to F_0(male)). At low stretch levels (λ_u=1.05) the predicted frequencies are F_0(case 1)=27.9 Hz and F_0(case 2)=54.1 Hz, compared to F_0(male)=33.7 Hz, F_0(female)=45.0 Hz. Now, the change in geometry affects males even more strongly (63% increase for F_0(case 2) relative to F_0(male)) than females (38% decrease for F_0(case 1) relative to F_0(female)). Changes in material properties are now overall less significant and affect females (20% increase for F_0(case 2) relative to F_0(female)) on a similar level as males (18% decrease in F_0(case 1) relative to F_0(male)).The reduced influence of material properties on fundamental frequency predictions is due to the fact that at low levels of stretch the nonlinearity of the tissue mechanical response only influences results a little.

To investigate the contributions of the vocal ligament to F_0, the beam model is also applied to the vocal fold cover only. Figure 6 shows F_0 predictions for the two-layer beam model and the cover-only beam model for males and females, respectively. Fundamental frequencies predicted for the cover-only model are significantly below those of the composite model, especially for females. For males the present study did not find the difference between the composite model and the cover-only model to be significant, however, the number of samples considered is small. For

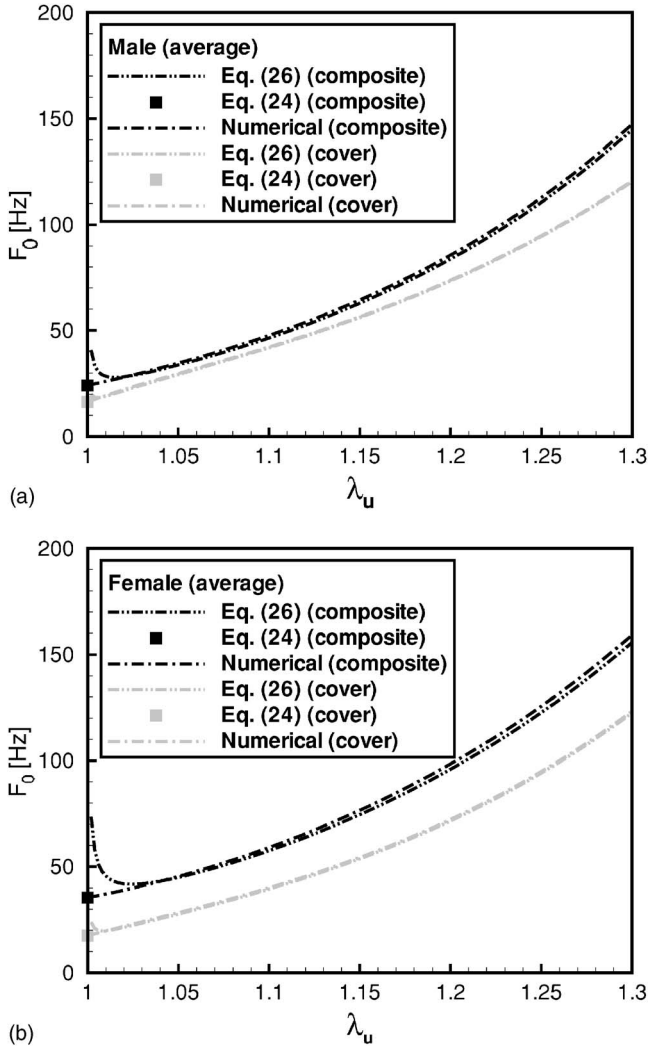


FIG. 6. Dependence of the fundamental frequency F_0 projected on the two-layer composite beam model and the cover-only beam model upon the stretch level for (a) the average male vocal fold and (b) the average female vocal fold.

females the presence of the ligament contributes to a much higher F_0. These predictions could reflect the finding of the initial shear modulus of the male cover being significantly larger than that of the female.

IV. DISCUSSION

The present study is motivated by the premise that a better understanding of fundamental frequency regulation and F_0 predictions can only be possible if an accurate description of the tissue, i.e., its mechanical response and its geometric features, is combined with an appropriate mechanical model of vocal fold vibration.

The present results show that for males the nonlinearity of the elastic response, as characterized by the parameter α, of the ligament is significantly higher than that of the cover, whereas there appeared to be no significant difference in the initial equilibrium shear modulus μ between the cover and the ligament. This combination of parameters implies that when subjected to the same amount of stretch the ligament would carry a larger tensile stress. This may allow the vocal fold cover to remain relatively loose for facilitating the propagation of the mucosal wave, and also prevent it from being damaged from excessive tension by allowing the vocal ligament to assume most of the tensile stress. Interestingly, this may not apply for females as no statistically significant differences in the constitutive parameters were found between the cover and the ligament.

As for the gender dependence of the constitutive parameters, it was found that for vocal fold cover specimens the initial equilibrium shear modulus μ of males is significantly larger than that of females, whereas for vocal ligament specimens the nonlinearity of the elastic response α for males is significantly larger than that for females. These findings may suggest that the homogenized or average stress in the vocal fold of the average male would be larger than that of the average female at equal stretch. In the present study no direct correlation between the model parameters and the underlying histological structure of the tissues is made. However, qualitatively the present results agree with findings in previous studies which have investigated the age and gender dependence of vocal fold histological structure in terms of collagen and elastin content.[1,28–30] Tissues with higher stiffness—that of males relative to females, and those of older males relative to younger males—were shown to possess higher collagen content.[28] The age-related increase in F_0 for males, Fig. 5, could be attributed to the stiffening of vocal fold tissue with increasing age as related to increased collagen content with age.[28] The gradually increasing F_0 values predicted for $\lambda_u = 1.1$ and 1.2 agree well with empirical lifespan changes where F_0 gradually increases with age in males above 40 years old.[9,10] It must be noted that any extrapolation of age dependence to include children cannot be considered at this stage of investigation. For children no age dependence of the constitutive and geometrical parameters has been determined in the present study and it is to be expected that significant changes in vocal fold structure occur during development and puberty.[5,8]

Fundamental frequencies also depend on the geometric features of the vocal fold. Female vocal folds have been found to be generally shorter than male vocal folds.[5] This is again confirmed in the present investigation, with the lengths of female vocal folds only 72% of those of male vocal folds on average. It has been demonstrated that female vocal folds are subjected to a higher stretch level during speaking than male vocal folds.[27] Shorter vocal folds and higher stretch levels may both compensate for the lower elastic modulus or a lower degree of nonlinearity, contributing to a higher overall F_0 for females compared to males. In the beam model predicted F_0 also depends on the aspect ratio of the beam cross section. The layered microstructure of the lamina propria motivates the use of rectangular cross-section geometries in the present study. In comparison, Titze and Hunter[6] employed a square cross section in their beam model. It is noted that all equations for F_0 presented in this study are in a general form such that different values of cross-section aspect ratio m can be employed in future studies.

From the predictions of the beam model it is found that the bending stiffness of the lamina propria alone can account for the restoring force of vibration when the vocal fold is fully relaxed. Thus the beam model can predict nonzero values of fundamental frequency in the unstretched state. For vocal folds in the stretched state, Fig. 3 indicates that the effects of bending stiffness are significant at any level of stretch. While the relative difference between F_0 predicted from the string and the beam model certainly decreases with increased stretch, the absolute difference between predictions of the string and the beam model in fact increases. On average, the contribution of bending stiffness is found to be larger for females than for males. From investigating, based on Fig. 3, an easily computed estimate for the predicted F_0 can be proposed. The sum of F_0 as predicted from the string model [Eq. (10)] and the beam model in the unstretched state [Eq. (24)] provides an approximation of the solution to the beam model [Eq. (26)]:

$$F_0(\lambda_u) \approx F_0(\lambda_u = 1) + F_0^{\text{string}} = \frac{1.7804}{L^2}\sqrt{\frac{A_0(\mu_c^2 f_c^3 + \mu_l^2 f_l^3 + 6\mu_c\mu_l\sqrt{f_c^3 f_l^3} + 4\mu_c\mu_l f_c f_l)}{m\rho(\mu_c f_c + \mu_l f_l)}} + \frac{1}{L\sqrt{2\rho}}\sqrt{\frac{\mu_c f_c}{\alpha_c}\left(\lambda_u^{\alpha_c - 2} - \lambda_u^{-\alpha_c/2-2}\right) + \frac{\mu_l f_l}{\alpha_l}\left(\lambda_u^{\alpha_l - 2} - \lambda_u^{-\alpha_l/2-2}\right)}. \tag{29}$$

It can be seen from Fig. 3 that the fundamental frequency models predict higher F_0 for the average female vocal fold than for the average male vocal fold, consistent with empirical data on human speaking F_0.

V. CONCLUSION

This study combines constitutive models of the vocal fold tissue with structural vibration models in order to investigate the dependence of the fundamental frequency of human phonation upon tissue properties, gender, age and F_0 regulation mechanisms. The mechanical responses of vocal fold cover and vocal ligament specimens are characterized, and it is shown that some statistically significant differences in the mechanical properties of these tissues exist in addition to the well established length difference between males and females. Fundamental frequencies calculated with the proposed two-layer composite beam model with vibration in the medial-lateral direction are significantly higher than those predicted by the string model. The fundamental frequency models presented differ from previous models in that vocal fold tissue is treated as incompressible, such that changes in vocal fold cross-section area during elongation are accounted for while previous models[6] assumed a constant cross-section area. Furthermore, both the vocal fold cover and the vocal ligament are integrated in the prediction of F_0 whereas previous studies projected the model predictions upon either the ligament[6] or the cover.[3] The beam model predicts nonzero fundamental frequency for the unstretched vocal fold [Eq. (24)], whereas the string model always predicts a zero fundamental frequency. It is proposed that—despite the nonlinear tissue response—a simple but reasonable prediction of F_0 can be obtained by summation of the F_0 predicted from the string model and the beam model for the relaxed vocal fold. Accounting for the bending stiffness of the vocal folds leads to significantly higher predicted F_0 than the string model over a wide range of stretch relevant to phonation. At vocal fold stretch levels deemed physiologically relevant the beam model is capable of predicting F_0 values approaching the lower bound of the phonatory range. Accounting for vocal fold stretch and extended clamping, i.e., the effective vocal fold length, the current two-layer beam model is capable of predicting F_0 values consistent with empirical data.[9,10]

Further additions and improvements can be developed based on the material parameters and the vibration model provided. It is believed that the vocalis muscle regulates the vibration of the vocal fold to some extent.[5] The two-layer composite beam model could be improved by placing the beam onto an elastic foundation representing the vocalis muscle. Such an additional constraint will lead to predictions of a higher F_0, and also introduce an additional mechanism of F_0 regulation through control of the muscle stiffness. Analytical solutions to the beam model could also be derived for cases with irregular cross-section geometries, and would then account for the macula flavae.[6] With such additions a more realistic and complete model would arise leading to more accurate predictions of the speaking F_0 at a physiological level.

ACKNOWLEDGMENTS

This work was supported by the National Institute on Deafness and Other Communication Disorders, NIH Grant No. R01 DC006101. The authors thank Neeraj Tirunagari and Min Fu of UT Southwestern Medical Center for their assistance in experimental measurements of vocal fold tissue elasticity.

[1] R. W. Chan, M. Fu, L. Young, and N. Tirunagari, "Relative contributions of collagen and elastin to elasticity of the vocal fold under tension," Ann. Biomed. Eng., 2007. (in press).
[2] Y. B. Min, I. R. Titze, and F. Alipour-Haghighi, "Stress-strain response of the human vocal ligament," Ann. Otol. Rhinol. Laryngol. **104**, pp. 563–569 (1995).
[3] K. Zhang, T. H. Siegmund, and R. W. Chan, "A constitutive model of the human vocal fold cover for fundamental frequency regulation," J. Acoust. Soc. Am. **19**(2), 1050–1062, (2006).
[4] R. W. Ogden, "Larger deformation isotropic elasticity—on the correlation

of theory and experiment for incompressible rubberlike solids," Proc. R. Soc. London, Ser. A **326**, 565–584 (1972).
[5]I. R. Titze, *Principles of Voice Production* (Prentice–Hall, Englewood Cliffs, NJ, 1994).
[6]I. R. Titze and E. J. Hunter, "Normal vibration frequencies of the vocal ligament," J. Acoust. Soc. Am. **115**, 2264–2269 (2004).
[7]R. Descout, J. Y. Auloge, and B. Guerin, "Continuous model of the vocal source," *International Conference on Acoustics, Speech and Signal Processing*, Denver, CO, pp. 61–64 (1980).
[8]C. Bickley, "Acoustic evidence for the development of speech," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA (1989).
[9]W. Brown, R. Morris, Jr., H. Hollien, and E. Howell, "Speaking fundamental frequency characteristics as a function of age and professional singing," J. Voice **5**, 310–315 (1991).
[10]M. I. P. Krook, "Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency analysis," Folia Phoniatr. **40**, 82–90 (1988).
[11]M. Blomgren, Y. Chen, M. L. Ng, and H. R. Gilbert, "Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers," J. Acoust. Soc. Am. **103**, 2649–2658 (1998).
[12]P. M. Morse, *Vibration and Sound*, 2nd ed. (McGraw–Hill, New York, 1948).
[13]I. R. Titze, "Mechanisms underlying the control of fundamental frequency," *Vocal Fold Physiology: Acoustic, Perceptual and Physiological Aspects of Voice Mechanisms*, edited by J. Gauffia and B. Hammberg (Singular, San Diego, 1991), pp. 129–138.
[14]I. R. Titze, J. J. Jiang, and E. Lin, "The dynamics of length change in canine vocal folds," J. Voice **11**, 267–276 (1997).
[15]T. A. Burnett and C. R. Larson, "Early pitch-shift response is active in both steady and dynamic voice pitch control," J. Acoust. Soc. Am. **112**, 1058–1063 (2002).
[16]I. R. Titze, *The Myoelastic Aerodynamic Theory of Phonation* (National Center for Voice and Speech, Iowa City, IA, 2006).
[17]D. W. Farnsworth, "High-speed motion pictures of the vocal cords," Bell Lab. Rec. **18**, 203–208 (1940).
[18]F. S. Brodnitz, *Vocal Rehabilitation* (Whiting Pressing, Rochester, MN, 1959).
[19]W. Zemlin, *Speech and Hearing Science. Anatomy & Physiology* (Prentice–Hall, Englewood Cliffs, NJ, 1997).
[20]R. W. Chan and I. R. Titze, "Viscoelastic shear properties of human vocal fold mucosa: Measurement methodology and empirical results," J. Acoust. Soc. Am. **106**, 2008–2021 (1999).
[21]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).
[22]K. Miller and K. Chinzei, "Mechanical properties of brain tissue in tension," J. Biomech. **35**, 483–490 (2002).
[23]D. F. Meaney, "Relationship between structural modeling and hyperelastic material behavior: Application to CNS white matter," Biomech. Model. Mechanobiol. **1**, 279–293 (2003).
[24]K. Levenberg, "A method for the solution of certain problems in least squares," Q. Appl. Math. **2**, 164–168 (1944).
[25]D. Marquardt, "An algorithm for least squares estimation of nonlinear parameters," SIAM J. Appl. Math. **11**, 431–441 (1963).
[26]S. Timoshenko, *Strength of Materials* (Krieger, Malabar, FL, 1955)
[27]H. Hollien, "Vocal pitch variation related to changes in vocal fold length," J. Speech Hear. Res. **3**, 150–156, (1960).
[28]T. H. Hammond, S. D. Gray, and J. E. Butler, "Age- and gender-related collagen distribution in human vocal folds," Ann. Otol. Rhinol. Laryngol. **109**, 913–920 (2000).
[29]T. H. Hammond, S. D. Gray, J. Butler, R. Zhou, and E. H. Hammond, "A study of age and gender related elastin distribution changes in human vocal folds," Otolaryngol.-Head Neck Surg. **119**, 314–322 (1998).
[30]S. D. Gray, I. R. Titze, F. Alipour, and T. H. Hammond, "Biomechanical and histological observations of vocal fold fibrous proteins," Ann. Otol. Rhinol. Laryngol. **109**, 77–85 (2000).

Listeners' identification and discrimination of digitally manipulated sounds as prolongations[a)]

Norimune Kawai,[b)] E. Charles Healey, and Thomas D. Carrell
Special Education and Communication Disorders, University of Nebraska-Lincoln, 206 Barkley Memorial Center, Lincoln, Nebraska 68583

(Received 19 May 2006; revised 19 May 2007; accepted 25 May 2007)

The present study had two main purposes. One was to examine if listeners perceive gradually increasing durations of a voiceless fricative categorically ("fluent" versus "stuttered") or continuously (gradient perception from fluent to stuttered). The second purpose was to investigate whether there are gender differences in how listeners perceive various duration of sounds as "prolongations." Forty-four listeners were instructed to rate the duration of the /ʃ/ in the word "shape" produced by a normally fluent speaker. The target word was embedded in the middle of an experimental phrase and the initial /ʃ/ sound was digitally manipulated to create a range of fluent to stuttered sounds. This was accomplished by creating 20 ms stepwise increments for sounds ranging from 120 to 500 ms in duration. Listeners were instructed to give a rating of 1 for a fluent word and a rating of 100 for a stuttered word. The results showed listeners perceived the range of sounds continuously. Also, there was a significant gender difference in that males rated fluent sounds higher than females but female listeners rated stuttered sounds higher than males. The implications of these results are discussed. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2750158]

PACS number(s): 43.70.Dn, 43.71.Es, 43.71.An [ARB] Pages: 1102–1110

I. INTRODUCTION

The prolongation of an initial speech sound in a word represents one of the core behaviors of stuttering (Van Riper, 1982; Guitar, 2005) and is a type of speech disruption that differentiates between speech that is fluent and speech that is stuttered. Although a sound prolongation is a core feature of stuttering, little is known regarding the minimal duration that prompts average listeners to identify and discriminate a sound prolongation as "abnormally long" or "stuttered." Van Riper (1982) suggested that prolongations of short duration (e.g., half a second) may be perceived as abnormal, but a sound prolongation that is deemed "stuttered" would last substantially longer.

Conture (2001) pointed out that there is a paucity of studies that have documented the characteristics of stuttering from a perceptual point of view. A review of the literature shows there are only a few studies that have documented the average durations of sound prolongations produced by people who stutter (PWS). Lingwall and Bergstrand (1979) were the first to investigate the duration of voiced continuant prolongations from an adult listener's perspective and found that when sounds were 294 ms in duration and longer, listeners judged them as "abnormal" and any sound duration 913 ms or longer was considered stuttered. In a later study, Zebrowski (1991) examined the duration of sound prolongations of four-year-old stuttering and nonstuttering children. From a 300-word spontaneous speech sample, Zebrowski found that the duration of each word perceived as "disfluent" averaged 435 ms for the children who stuttered and 404 ms for the children who did not stutter. Unfortunately, Zebrowski did not define or explain the criteria that were used in considering a word as "fluent" or stuttered. In another study, Zebrowski (1994) measured the duration of sound prolongations from a conversational sample of school-age children who stuttered and found that the average sound prolongation for this group was 724 ms, with a minimum duration of 403 ms. As in her 1991 study, Zebrowski did not explain or define how she determined a sound was considered fluent or stuttered.

In 2001, Susca and Healey digitally manipulated words beginning with a vowel, voiceless fricative, or semivowel within a reading passage to artificially create words that were perceived by listeners as stuttered words. The durations of the digitally prolonged sounds used in the passage were an average length of approximately 300 ms. These sound prolongations were independently judged as sound prolongations based on a criterion of 90% or more agreement among judges who were certified speech-language pathologists and graduate students majoring in speech-language pathology. More recently, Jones *et al.* (2005) asked listeners who were undergraduate or graduate students to determine the durational difference between normal and abnormally long speech sounds. Jones *et al.* (2005) found that listeners' average threshold for a prolonged vowel sound was 279 ms and 235 ms for a voiceless fricative, regardless of the speech rate used. These durational measures were similar to those found by Susca and Healey (2001a).

The majority of previous research on the perception of sound prolongations has been limited to a perceptual categorical difference between fluent and stuttered sounds. Durational differences in sound prolongations among past stud-

[a)]Portions of these data were presented at the annual convention of the American Speech-Language-Hearing Association, San Diego, November 2005.

[b)]Electronic mail: nkawai@unlserve.unl.edu

ies suggest that the perceptual boundaries are variable across listeners. Perhaps there is a durational threshold that is reached that prompts listeners to judge a prolonged sound as fluent or stuttered. In this sense, the perception of sound prolongations might be categorical in nature, like voice onset time (Lisker and Abramson, 1967; Pisoni *et al.*, 1982; Pisoni *et al.*, 1994). However, it is also possible that sound prolongations are perceived continuously or as a gradient perceptual phenomenon much like what is found for Cantonese lexical tones (Francis *et al.*, 2003) and vowels, fricatives, and nonspeech sounds (Mirman *et al.*, 2004).

The concept that speech production represents a continuum from normally fluent to severely stuttered speech is not new. Adams and Runyan (1981) argued that the distinction between stuttered and fluent speech could be difficult to discern within the continuous stream of speech. They theorized that a disfluent speaker's speech production continually varies from fluency to "tenuous" fluency (i.e., slow, dysrhythmic, weak, tremorous, unstable, or imprecisely articulated fluency) to stuttering throughout an utterance. Adams and Runyan did not explain what prompts listeners to perceive that a sound or word is produced with normal or tenuous fluency or when it is considered stuttering. It is possible that listeners are sensitive to minor variations in the production of an utterance such that a sound or word may be perceived as "slightly fluent" or "slightly disfluent." These two perceptual categories may represent ambiguous perceptual points that exist between a range of durations that represent fluent and stuttered speech sounds. Given this possibility, sound prolongations might be perceived along a continuum rather than as discrete categories of fluent and stuttered sounds.

Whether sound prolongations are perceived categorically or continuously might also relate to a number of factors such as the sound being prolonged, the speech context from which the perception is made, perceived vocal tension or pitch changes during the stuttered moment (Gregory, 2003), or the gender of the listener. Of these factors, the gender of the listener has received the most scientific attention. There is evidence concerning how females demonstrate greater acceptability of disabilities than males (Horne, 1985; Ferguson, 1999). With regard to stuttering, Burley and Rinaldi (1986) and Schroder *et al.* (2002) found that men tend to rate people who stutter more negatively than women. Dietrich *et al.* (2001) also discovered that females rated the character traits of people who stutter more favorably than males. By contrast, Patterson and Pring (1991) failed to replicate Burley and Rinaldi's gender differences while Hulit and Wirtz (1994) found that age and gender of individuals were not good predictors of attitudes toward PWS. Susca and Healey (2001b) also failed to find differences between male and female and old and young listeners when evaluating mild and severe stuttering. These equivocal findings make it difficult to know if listener gender and age influence the perceptions of stuttering and PWS. Additional research on the impact of the gender of the listener relative to perceptions of stuttering seems warranted.

In summary, a few studies have investigated the durations of sound prolongations but have done so from the perspective of listeners having to determine if the sound is fluent or stuttered (Van Riper, 1982; Zebrowski, 1991, 1994). These studies tell us little about the minimum duration of a sound prolongation that could be identified as stuttered by an average listener. Past research is not consistent in specifying the minimal duration required in order for listeners to perceive it as abnormally long or stuttered (Lingwall and Bergstrand, 1979; Susca and Healey, 2001a; Jones *et al.*, 2005). Sound durations perceived as prolongations have ranged from 235 to 403 ms. Perhaps one reason for such a wide range of sound durations is related to listeners being forced to choose between fluent versus stuttered sounds when there might not be distinct perceptual categories. If the perceptual boundaries are not categorical then they might be continuous in nature. Furthermore, it is possible that the perceptual boundaries might be different between men and women listeners considering that a few studies (Dietrich *et al.*, 2001; Schroder *et al.*, 2002) have found that females react more favorably to stuttering than do males.

Given that there are conflicting findings in the extant literature regarding the minimal duration that is required for a sound to be perceived as a sound prolongation, two experimental questions were developed for this study. First, do listeners perceive gradually increasing durations of sounds categorically or continuously? Second, are there differences between male and female listeners in the perceptual identification of a sound that gradually increases in duration?

II. METHODS

A. Participants

A total of 44 listeners (22 males and 22 females) participated in this study. The age range for all listeners was from 19 to 50 years. The mean listener age for males was 28.5 years and 21.8 years for females. Listeners were selected from undergraduate courses in speech-language pathology and special education at the University of Nebraska-Lincoln (UNL) as well as faculty and staff in the departments across UNL. None of the listeners were graduate students in the department of speech-language pathology, certified speech-language pathologists, faculty in speech-language pathology, family members of people who stutter, or anyone with extended contact with people who stutter. All listeners' primary language was English and none had any history of speech, language, or neurological disorders as determined by reports from each listener. To verify that all participants had normal hearing, hearing acuity was tested using standard audiological hearing screening procedures.

B. Stimuli

In order to control that the target speech sound was produced fluently and did not contain any subperceptible or tenuous fluency, it was necessary to obtain a speech sample from a normally fluent speaker rather than a person who stuttered. Thus, a 27-year-old male, who was a doctoral student in speech-language pathology at UNL, served as the speaker. He did not have any present or past speech, lan-

guage, hearing, and/or neurological disorders. His native language was English and he spoke with a standard American dialect.

The speaker was placed in a sound treated room and asked to read the Rainbow Passage (Fairbanks, 1960) several times. The Rainbow Passage is commonly used for studies in speech science. One of the repeated samples was judged by the authors as the most representative of the speaker's natural speech sample. The first paragraph of this sample was digitally recorded using the SONY TCD-D10 PROII digital audio tape recorder. Then, the speech sample was transferred to a computer sound file. The reading rate of the speaker was calculated to be 185.4 syllables/min for the entire paragraph which is within normal limits. From the paragraph, the phrase, "These take the shape of a long round arch," which is part of the sentence, "These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon" was used as the target stimulus for the current study.

The /ʃ/ in the word of "shape" was selected as the digitally manipulated target sound because (1) it was easy to digitally extend the duration of the frication noise associated with a voiceless fricative, (2) most stuttering behaviors occur at the beginning of words, even those embedded in sentences or phrases, (3) listeners would hear several fluent words prior to making judgments about the target sound, and (4) listeners could assess the duration of the target sound relative to the speaking rate established for the first three words of the phrase. All digitization was performed using a speech analysis computer program (SOUND FORGE version 4.0c, 1997) at a sampling rate of 44 kHz. The duration of the speaker's natural fluent production of the /ʃ/ sound was 120 ms, at a speaking rate of 185.4 syllables/min. The duration of the /ʃ/ sound was normal at this speaking rate (Klatt, 1973; Umeda, 1977).

Two methods were used to create the digitally prolonged speech sounds. First, 20 ms of the middle portion of the frication was digitally copied. In order to prevent the occurrence of noise, the 20 ms portion was deleted at a point where the amplitude of the segment began and ended at baseline (0 dB). This 20 ms segment was added to the target /ʃ/ where its wave form crossed the baseline. By adding the 20 ms segment to the middle of the original /ʃ/ sound, a sound duration of 140 ms was created. This process was repeated to create /ʃ/ durations of 160, 180, and 200 ms. However, the manipulated /ʃ/ became unnatural sounding when the frication noise was extended to 220 ms. For this and other sounds exceeding this duration, another method was used to create the digitally prolonged sounds. Instead of copying the middle 20 ms segment of the 200 ms sound, it was necessary to copy a small portion of the beginning, middle, and end of the 200 ms sound and add each segment to the beginning, middle, and end of a new target sound.

Specifically, a 7 ms segment at the beginning, a 6 ms segment in the middle, and a 7 ms segment at the end of /ʃ/ of the 200 ms sound were copied and added to the beginning, middle, and end in order to create a 220 ms sound duration. This same three-segment digital manipulation process was continued for sounds ranging from 240 to 500 ms, in 20 ms durational increments. A total of 20 sound prolongation samples were created for this experiment (ranging from 120 to 500 ms). Figure 1 shows the wave forms for the fluent /ʃ/ duration (120 ms) and the stuttered /ʃ/ duration (500 ms).

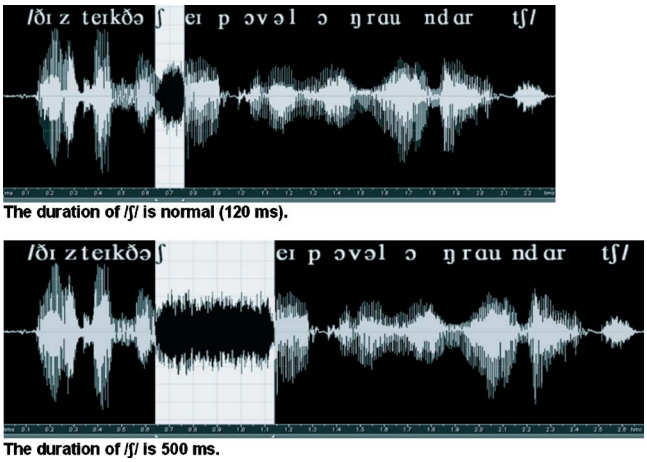


FIG. 1. (Color online) Wave forms of the stimuli of normal duration of /ʃ/ in "shape" (120 ms) (top) and 500 ms (bottom), which is the longest /ʃ/ in the present experiment. The experimental phrase, "These take the shape of a long round arch" was taken from the first paragraph of the Rainbow Passage (Fairbanks, 1960).

C. Procedures

Each listener was taken to a sound treated laboratory with low ambient noise and sound reflection for testing. Before testing began, the first author asked each listener if he/she knew anyone currently or in the past who had a communication disorder. If a listener indicated that he or she had extensive contact or familiarity with a person who stutters, then the listener was not permitted to participate in the study to prevent listener familiarity with stuttering. No one was excluded from this study for this reason. Each participant's hearing was screened at 20 dB at 250, 500, 1000, 2000, 4000, and 8000 Hz using a Beltone Audio Scout portable audiometer. All listeners successfully passed the hearing screening procedure.

Once listeners passed the hearing screening, they were seated at a desk facing a computer screen. Up to two listeners could be tested at one time using two computer screens separated by a divider that did not allow either participant to see either computer screen. On the computer screen, the listeners saw a computer-based magnitude scale (Kawai, 2005) that was developed using a programming language (Microsoft VISUAL BASIC .net, 2003). The program displayed 20 slider bars that appeared on the computer screen (see Fig. 2). Listeners moved the slider bar with a computer mouse up or down to the desired location on the bar based on what they heard. The magnitude scale on the slider bar ranged from 1 (fluent) to 100 (stuttered).[1]

The speech samples were presented to the listeners at a comfortable loudness level (approximately 74 dB SPL) through calibrated Beyerdynamic DT 770 headphones. Calibration of these headphones was conducted using a Brüel & Kjær precision sound level meter (Type 2235) with Microphone Type 4176 connected to the headphones via an artifi-

FIG. 2. (Color online) The screen of the PROSLIDER version 1.0. The arrow was set at middle (50) (see Sentence 1) for each listening trial. When listeners perceived a stimulus as stuttering, they were instructed to move the arrow to the top (100) (see Sentence 2). When they perceived a stimulus as fluent, they were instructed to move the arrow to the bottom (1) (see Sentence 3). Once listeners placed the arrow at the desired rating location and clicked CONFIRM, the number and the arrow disappeared (see Sentence 4).

cial ear. The loudness of the sound stimuli from each computer that the participants listened to was adjusted via an amplifier, Mackie 1202-VLZ PRO and set at approximately 74 dB SPL.

Listeners had to rate each of the 20 segments, 10 different times in random order resulting in 200 ratings from each listener. This was done in order to obtain high intrajudge reliability for the ratings. When listeners completed their rating of the first set of 20 stimuli, they were instructed to click on the check box on the lower-right side of each page, which allowed the data on the page to be automatically recorded into a Microsoft EXCEL spreadsheet. Then they were to click on the command "NEXT PAGE" button to go to the next page. On each slider bar, there was also a small window that showed a number (1–100) so that participants were able to see the number they assigned to each stimulus. Once they clicked on the "CONFIRM" button, the slider and the slider bar disappeared from view. In this way, a listener could not refer to or memorize how they rated any previous stimuli. Listeners were not provided with any definitions of the categories for the level of fluency or stuttering. They were given an oral instruction by the experimenter to place the cursor at the bottom of the slider when they thought a /ʃ/ sound was fluent and place it at the top of the slider when they thought a /ʃ/ sound was stuttered. Listeners were also told to place the arrow between 1 and 100, respectively, if they did not believe the target stimulus was not either completely fluent or stuttered.

In order to play the stimuli, the software MAKETAPE version 2.2 (Srinivasan and Carrell, 2004) was used. This software controls the presentation of auditory stimuli with any inter-stimulus intervals (ISI) and intertrial intervals. In this experiment, the ISI was set at 3.5 s, which most listeners indicated was the most comfortable ISI in our pilot study. For every 20 stimuli, the ISI was set at 10 s rather than 3.5 s because listeners had to click on the check box to save their ratings and then click on the "NEXT PAGE" button to go on to another page.

The arrows in Fig. 2 were set at the point of 50. The arrow on the left at the bottom shows fluent (rating=1), and the one at the top shows stuttered (rating=100). The fourth scale from left on the top row of Fig. 2 shows how the scale looked once the listeners pressed the "CONFIRM" button.

D. Data analysis

Listeners' average ratings for each of the 20 sound durations were computed. The ratings for the 10 repeated presentations of the 20 sound durations were averaged for each participant. Figure 3 shows a plot of those values. In addition, a best-fitting logistic curve (Netter and Wasserman, 1974) was calculated to characterize the fluent-to-stuttered

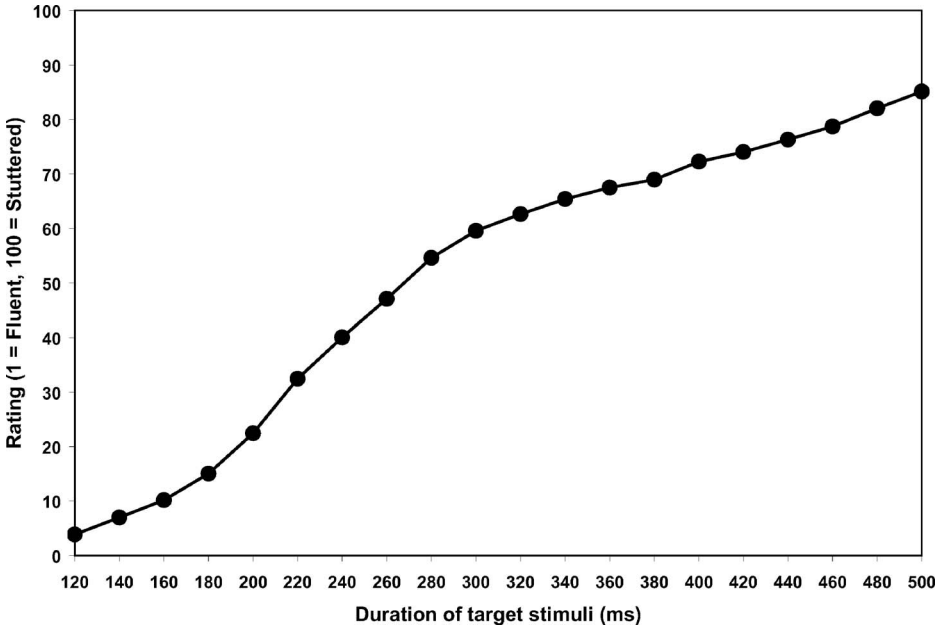


FIG. 3. Mean perceptual ratings for all durations of stimuli.

identification function demonstrated by listeners. The logistic curve models the sigmoidal shape. And in order to determine whether the listeners' ratings showed a nonlinear change, a correlation between the listener's average ratings and the logistic curve values of the 20 data points was calculated.

A one-way analysis of variance (ANOVA) was conducted to determine whether there was a significant difference between males and females' ratings in each duration point. A two-way factorial ANOVA was used to test for an interaction between gender and fluent versus stuttered ratings. In order to test this interaction, the first quartiles and the last quartiles out of 20 data points were used. The first quartiles included the 120, 140, 160, 180, and 200 ms duration stimuli which were more likely to be rated the stimuli fluent, and the last quartiles included the 420, 440, 460, 480, and 500 ms duration stimuli which were more likely to be rated stuttered.

A best-fitting logistic curve also was calculated to characterize the fluent-to-stuttered identification function differences between male and female listeners. The measure of difference between male and female listeners was calculated based on the steepness of the logistic slopes at the 50% crossover point for each gender. In addition, correlations between the actual average ratings and the logistic curve values of the 20 data points were calculated for males and females. The steepness of the slope represents the perceptual speed that listeners identify and discriminate one category from another. Each listener's response pattern was modeled with a best fitting logistic function. The slope of those functions was calculated in order to investigate how quickly the listeners perceived different duration of sounds from one category to another.

III. RESULTS

The first set of results relates to the question of whether variations in sound durations ranging from fluent to stuttered are perceived continuously or categorically. Figure 3 shows a shallow sigmoidal shape and a line that never reaches to either fluent or stuttered extremes. The shape of the line connecting the average ratings across the 20 durational segments shows a continuous perceptual pattern where the duration of the /ʃ/ sound shifts from fluent to stuttered. The predicted-to-measured correlation for the 20 sound durations was $r=0.973$, indicating that the shape of the curve is sigmoidal.

The second set of results relates to differences between males' and females' perceptions of the sound duration categories. A one-way ANOVA was performed on each sound duration point to determine if there was a significant difference in the ratings between male and female listeners. Average ratings of the listeners at each sound duration point are shown in Table I. Because this was a multiple pairwise test, family wise error was accounted for using a Bonferroni correction ($p=0.0025$). The average ratings between males and females did not reach statistical significance for any of the 20 phoneme durations.

In order to examine listener ratings in a different way, an analysis was conducted by taking out the first (120–200 ms) and the last (420–500 ms) quartiles that were least and most likely to be labeled stuttered. A two-way between groups factorial ANOVA was conducted to examine the main effects and interaction relating genders and quartiles. There was a significant interaction of gender and quartiles [$F(1,436)=40.30$, $p<0.01$, $MSe=154.653$, $d=0.607$]. Further analysis based on Fisher's least significant difference (LSD) follow-ups of the cell means revealed that male listeners gave higher ratings than females for the phrases when the durations of the target /ʃ/ sounds ranged from 120 to 200 ms (first quartile), whereas when the /ʃ/ sounds ranged from 420 to 500 ms (fourth quartile), females gave higher ratings than males. As expected, there was a significant main effect for quartiles [$F(1,436)=3244.65$, $p<0.01$, $d=5.443$], which meant all listeners gave higher ratings at the fourth quartile than at the first quartile. Overall, there was no significant effect for genders [$F(1,436)=0.429$, $p=0.51$, $d=0.063$]. However, there was a simple effect between genders at the first and fourth quartiles, which was significant in opposite directions at

TABLE I. Average ratings of the listeners at each duration point.

	Male		Female		Combined	
Durations (ms)	*M*	*SD*	*M*	*SD*	*M*	*SD*
120	5.04	4.48	2.69	2.19	3.87	3.72
140	9.38	9.40	4.62	3.83	7.00	8.07
160	13.90	13.93	6.45	5.30	10.18	12.09
180	19.39	14.64	10.74	8.67	15.06	13.65
200	26.71	16.49	18.19	10.56	22.45	15.03
220	36.18	17.96	28.77	10.76	32.47	15.19
240	41.90	16.80	38.23	11.14	40.06	13.77
260	48.96	16.85	45.20	10.22	47.08	14.19
280	53.20	16.16	56.00	11.25	54.60	14.50
300	58.37	15.27	60.76	10.29	59.57	12.96
320	59.83	14.23	65.41	11.36	62.62	13.80
340	62.96	14.34	67.84	10.45	65.40	12.91
360	64.70	15.68	70.33	12.29	67.52	14.26
380	64.75	15.38	73.14	11.29	68.94	13.43
400	69.01	13.63	75.50	11.62	72.25	12.88
420	70.75	13.09	77.31	11.56	74.03	13.04
440	73.00	12.67	79.63	12.68	76.31	12.84
460	75.70	14.52	81.72	12.51	78.71	13.68
480	78.93	12.91	85.20	12.42	82.06	12.61
500	81.97	13.79	88.30	11.65	85.14	11.93

these quartiles. This meant that at the first quartile, male listeners rated the speech samples significantly higher than females but at the fourth quartile, females rated the samples significantly higher than males (see Fig. 4).

Logistic best-fit approximations to the data were used to assure accurate crossover and slope values. The functions were highly correlated with the measured values (r=0.967 for males and r=0.975 females). The mean slope value for males was 0.05 (SD=0.03) for males and 0.08 (SD=0.10) for females. An independent samples t test was conducted to compute the slope difference between males and females. No significant gender differences were found [$t(42)$=1.39, p =0.17 (two-tailed), d=0.42]. The 50% crossover points for each gender were calculated. The mean 50% crossover points for males was 321.69 ms (SD=66.57) and females was 321.14 ms (SD=37.48). There were no significant gender differences in the 50% crossover points [$t(42)$=0.035, p=0.973 (two-tailed), d=0.001].

IV. DISCUSSION

Recall that one of the main purposes of this study was to determine if listeners perceive sound prolongations categorically or on a continuum. The correlation value associated with the logistic function objectively shows that a sigmodial curve fits the data. This indicates that listeners' perceptual

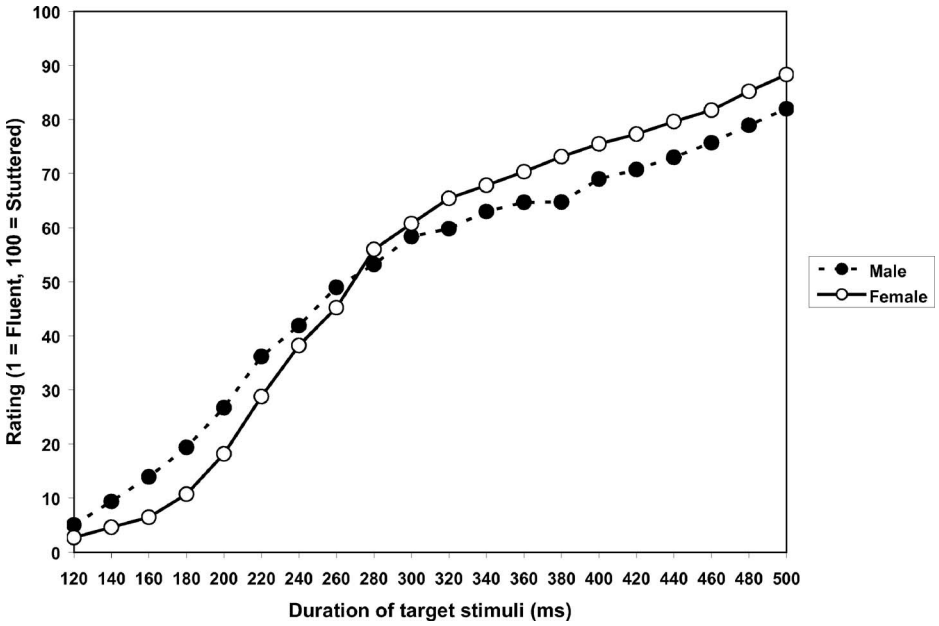


FIG. 4. Perceptual ratings by male and female listeners for each of the 20 sound durations. There was one gender difference in perception of sound prolongations. Males rated the phrases with /ʃ/ sounds higher than females at the first quartile, but females rated the phrases with /ʃ/ sounds higher than males at the fourth quartile. However, the steepness of slopes at the 50% crossover was not significantly different between males and females.

pattern of sound prolongations was not linear. Listeners' ratings did not increase exponentially in the middle and were not flat at the beginning or at the end (no floor and ceiling effects were observed). This suggests that listeners perceived gradual increases in a sound duration along a continuum from fluent to stuttered productions. Therefore, a sound prolongation is not one specific perceptual boundary point but rather a range of durations that creates several levels of perceptual uncertainties about the degree of normalcy of the sound until the sound is clearly abnormally long to be considered stuttered. This might be one of the reasons why previous studies that investigated perception of sound prolongations showed inconsistent results of how long phonemes have to be in order to be perceived as sound prolongations (e.g., Lingwall and Bergstrand, 1979; Susca and Healey, 2001a; Jones *et al.*, 2005; Zebrowski, 1994).

As stated earlier, the perception of a sound prolongation does not appear to be categorical like voice onset time (VOT) (Strange and Jenkins, 1978; Aslin and Pisoni, 1980; Aslin *et al.*, 1981; Pisoni *et al.*, 1982, 1994). Our results showed a gentle sloping pattern (see Fig. 3). This supports the notion that a perceptual category of tenuous fluency appears to exist when listeners evaluate a disfluent person's speech (Adams and Runyan, 1981). Also, the pattern of the curve generated from the present study's data is somewhat similar to the results of classic magnitude estimation curve with a prothetic continuum, which is a continuum that shows a curvilinear relation between magnitude estimates and interval scale values of the same set of stimuli (Schiavetti *et al.*, 1983). Loudness and brightness of lights are the examples of this type of curvilinear perceptual relationship (Stevens, 1957, 1961, 1962, 1974, 1975). Of course, the difference of experimental methodologies between the present study, which used a magnitude estimation type rating system, and classic VOT studies, which use the traditional two-alternative forced-choice task, should be taken into consideration. However, unlike a classic magnitude estimation task, the rating scale of the current study had an element of a two-alternative forced-choice task because the extremes of this rating scale were fluent and stuttered. It might be beneficial to conduct a classic VOT study based on this rating scale so that the results of the current study could confirm that listeners' perceptions of sound prolongations are continuous rather than categorical.

Magnitude estimation has been used to measure nonlinear relationships between perceived magnitude and stimulus intensity. It is also used in the area of speech-language pathology to scale speech intelligibility of people who are hearing impaired (Schiavetti *et al.*, 1981), to assess severity of hypernasality (Whitehill *et al.*, 2002), to evaluate voice quality (Kreiman *et al.*, 1993), and measure stuttering severity (Schiavetti *et al.*, 1983; Schiavetti *et al.*, 1994). For example, Schiavetti *et al.* (1983) investigated how average listeners perceive the severity of stuttering via magnitude estimation. They found as severity increased, listeners ratings also increased in a manner consistent with a prothetic continuum manner. A later psychophysical study by Schiavetti *et al.* (1994) showed that average listeners' perceptions of speech naturalness spoken by PWS and people who are normally fluent also produced a prothetic continuum pattern.

The results of the present study also did not show gender differences for judging sound prolongations at any of the 20 phoneme durations. Previous studies found that females showed more favorable and greater acceptance of stuttering than did males (Dietrich *et al.*, 2001; Schroder *et al.*, 2002). The results of the present study are not consistent with these findings. Perhaps this was due to listeners having only 3.5 s to make a judgment and a rating from auditory information alone. In the studies by Dietrich *et al.* (2001) and Schroder *et al.* (2002), participants were given unlimited time to think about their judgment of stuttering. This could have allowed participants to consider other variables such as speakers' personality, intelligence, and physical or social characteristics in making decisions.

When the average ratings of males and females were calculated at the first and fourth quartiles, significant gender differences were observed at both quartiles. Gender differences are difficult to explain because males had higher ratings in the first quartile and females had higher ratings in the fourth quartile. This suggests that males allow for more variation in ratings for fluent durations while females rate stuttering at shorter durations than males. It might be beneficial to conduct a follow-up experiment using a female speaker to sort out whether the same results would be found. Although there were gender differences in ratings at the first and fourth quartiles, there were no significant gender differences in the identification and discrimination speed from one category to another because there was no difference in the slope steepness and the 50% crossover points of fluency ratings between male and female listeners. This finding suggests that both males and females are similar in the speed with which they identify and discriminate sounds along a continuum of durations.

Although the results of this study clearly show that sound prolongations are perceived on a continuum, the study is not without its limitations. First, the experimental stimuli were limited to manipulations of one phoneme embedded within a short phrase. Perceptual judgments of stuttering might involve more than just the duration of a sound prolongation, such as stuttering severity, the presence or absence of direct information about the person who stutters, the presence of secondary coping behaviors, and/or the context of the spoken message. Second, the stimuli were limited to one sound in a short phrase in order to exercise control over the digitized sound. Third, the speech sound used was a voiceless fricative and was chosen because of the ease and reliability of creating a range of target stimuli. Other types of digitized sounds such as voiced consonants, short vowels, and long vowels might have produced different results had they been used as stimuli, although these sounds would have been more difficult to digitize. Fourth, the target speech was produced at only one speaking rate. Perceptions of stuttering might be related to speaking rate. Last, even though the authors tried to instruct the listeners to give "1" for fluent and "100" for stuttered, it was discovered that seven listeners did not follow those directions and rated each stimulus relative to the shortest target sound or the last sound they heard. For

example, one of the listeners reported that he placed the slider bar at "10" when the duration of /ʃ/ was 180 ms. He thought the target stimulus was fluent, but he still compared it to the /ʃ/ duration that was 120 ms.

In future studies, it might be better to change the criterion from fluent versus stuttered to a percentage of confidence listeners perceive the stimulus resembles a stuttered sound. In this way, listeners will not rate the stimuli relative to the shortest target sound or the last sound they heard. It will also be beneficial to manipulate durations of /ʃ/ to create broader durational range than the current study (i.e., 80–1000 ms) so it might be possible to see clear floor and ceiling effects in the perception curve. In addition, it might be beneficial to conduct the same kind of experiments by changing a gender of a speaker to examine a speaker-listener interaction effect, and by changing a speaker from a PWDNS to a PWS to compare what sources and elements other than durations of sound prolongations contribute to listeners' perceptions of fluent versus stuttered speech. Future studies also might want to include listeners' perception of sound prolongations when different levels of stuttering frequency are included in the speech sample. Moreover, the perceptibility of a sound prolongation might not be just a matter of specific or minimal sound durations but more related to a collection of other speech features that prompt listeners to hear a word or sound as prolonged.

V. CONCLUSIONS

In the current study, the following results were found:

(1) The listeners perceived gradual increases in a sound duration from fluent to stuttered rather than identifying and discriminating fluent or stuttered categories.
(2) There was one gender difference in the perception of sound prolongations. Males rated the phrases with /ʃ/ sounds that are likely to be labeled fluent higher than females, but females rated the phrases with /ʃ/ sounds that are likely to be labeled stuttered higher than males. The gender difference of perceptual speeds (i.e., how quickly the listeners change their perceptions of the target phoneme from fluent to stuttered) was not observed based on the results from a best fitting logistic function (Netter and Wasserman, 1974).

[1]A pilot study by Kawai *et al.* (2005) found no evidence that listeners made judgments differently between abnormal versus normal when listeners were given two different terms to rate. One magnitude scale consisted of two extremes, which were normally long versus abnormally long, and the other scale consisted of fluent and stuttered endings. Listeners judged stimuli as abnormally long and stuttered equally.

Adams, M. R., and Runyan, C. M. (**1981**). "Stuttering and fluency: Exclusive events or points on a continuum?," J. Fluency Disord. **6**, 197–218.

Aslin, R. N., and Pisoni, D. B. (**1980**). "Effects of early linguistic experience on speech discrimination by infants: A critique of Eilers, Gavin, and Wilson (**1979**)," Child Dev. **51**, 107–112.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., and Perey, A. J. (**1981**). "Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience," Child Dev. **52**, 1135–1145.

Burley, P. M., and Rinaldi, W. (**1986**). "Effects of sex of listener and of stutterer on ratings of stuttering speakers," J. Fluency Disord. **11**, 329–333.

Conture, E. G. (**2001**). *Stuttering: Its Nature, Assessment and Treatment* (Allyn & Bacon, Needham Heights, MA).

Dietrich, S., Jensen, K. H., and Williams, D. E. (**2001**). "Effects of the label 'stutterer' on student perceptions," J. Fluency Disord. **26**, 55–66.

Fairbanks, G. (**1960**). *Voice and Articulation Drillbook*, 2nd ed. (Harper & Row, New York).

Ferguson, J. M. (**1999**). "High school students' attitudes toward inclusion of handicapped students in the regular education classroom," Educ. Forum **63**, 173–179.

Francis, A. L., Ciocca, V., and Ng, B. K. C. (**2003**). "On the (non)categorical perception of lexical tones," Percept. Psychophys. **65**, 1029–1044.

Gregory, H. H. (**2003**). *Stuttering Therapy: Rationale and Procedures* (Allyn and Bacon, Boston).

Guitar, B. (**2005**). *Stuttering: An Integrated Approach to its Nature and Treatment*, 3rd ed. (Lippincott Williams & Wilkins, Baltimore).

Horne, M. D. (**1985**). *Attitudes Toward Handicapped Students: Professional, Peer, and Parent Reactions* (Erlbaum, Hillsdale, MI).

Hulit, L. M., and Wirtz, L. (**1994**). "The association of attitudes toward stuttering with selected variables," J. Fluency Disord. **19**, 247–267.

Jones, K., Logan, K. J., and Shrivastav, R. (**2005**). "Duration, rate, and phoneme-type effects on listeners' judgments of prolongations," poster session presented at the Annual Meeting of the American Speech-Language-Hearing Association, San Diego.

Kawai, N. (**2005**). "The Pro Slider (Version 1.0)," University of Nebraska-Lincoln Speech Perception Laboratory, Lincoln, NE.

Kawai, N., Healey, E. C., and Carrell, T. C. (**2005**). "Identification and discrimination of phoneme prolongation," poster session presented at the Annual Meeting of the American Speech-Language-Hearing Association, San Diego.

Klatt, D. H. (**1973**). "Duration characteristics of pre-stressed word-initial consonant clusters in English," Quarterly Progress Report, Research Laboratory of Electronics, MIT **108**, 253–260.

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., and Berke, G. S. (**1993**). "Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research," J. Speech Hear. Res. **36**, 21–40.

Lingwall, J. B., and Bergstrand, G. G. (**1979**). "Perceptual boundaries for judgments of 'normal,' 'abnormal,' and 'stuttered' prolongations," poster session presented at the Annual Meeting of the American Speech-Language-Hearing Association, Atlanta.

Lisker, L., and Abramson, A. S. (**1967**). "Some effects of context on voice onset time in English stops," Lang Speech **10**, 1–28.

"Microsoft Visual Basic .net program (Version 2003)," (**2003**) Microsoft, Redmond, WA.

Mirman, D., Holt, L. L., and McClelland, J. L. (**2004**). "Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly-changing acoustic cues," J. Acoust. Soc. Am. **116**, 1198–1207.

Netter, J., and Wasserman, W. (**1974**). *Applied Linear Statistical Models* (Irwin, Homewood, IL).

Patterson, J., and Pring, T. (**1991**). "Listeners attitudes to stuttering speakers: No evidence for a gender difference," J. Fluency Disord. **16**, 201–205.

Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (**1982**). "Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants," J. Exp. Psychol. Hum. Percept. Perform. **8**, 297–314.

Pisoni, D. B., Logan, J. S., and Lively, S. E. (**1994**). "Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception," in *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Word*, edited by H. C. Nusbaum and J. Goodman (MIT, Cambridge, MA), pp. 121–166.

Schiavetti, N., Martin, R. R., Haroldson, S. K., and Metz, D. E. (**1994**). "Psychophysical analysis of audiovisual judgments of speech naturalness of nonstutterers and stutterers," J. Speech Hear. Res. **37**, 46–52.

Schiavetti, N., Metz, D. E., and Sitler, R. W. (**1981**). "Construct validity of direct magnitude estimation and interval scaling of speech intelligibility: Evidence from a study of the hearing impaired," J. Ship Res. **24**, 441–445.

Schiavetti, N., Sacco, P. R., Metz, D. E., and Sitler, R. W. (**1983**). "Direct magnitude estimation and interval scaling of stuttering severity," J. Speech Hear. Res. **26**, 568–573.

Schroder, L., Melnick, K. S., Koul, R., and Keller, J. P. (**2002**). "Attitudes towards stuttering: Divergence in gender," poster session presented at the Annual Meeting of the American Speech-Language-Hearing Association, Atlanta.

Sound Forge (Version 4.0c) (**1997**). Sonic Foundry, Madison, WI.

Srinivasan, N., and Carrell, T. C. (**2004**). "Maketape (Version 2.2)," Univer-

sity of Nebraska-Lincoln Speech Perception Laboratory, Lincoln, NE. Retrieved January 23, 2005, from http://hush.unl.edu/Lab Resources.html
Stevens, S. S. (**1957**). "On the psychophysical law," Psychol. Rev. **64**, 153–181.
Stevens, S. S. (**1961**). "To honor Fechner and repeal his law," Science **133**, 80–86.
Stevens, S. S. (**1962**). "The surprising simplicity of sensory metrics," Am. Psychol. **17**, 29–39.
Stevens, S. S. (**1974**). "Perceptual magnitude and its measurement," in *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), Vol. **2**, pp. 22–40.
Stevens, S. S. (**1975**). *Psychophysics* (Wiley, New York).
Strange, W., and Jenkins, J. J. (**1978**). "Role of linguistic experience in the perception of speech," in *Perception and Experience*, edited by R. D. Walk and H. L. Pick (Plenum, New York), pp. 125–169.
Susca, M., and Healey, E. C. (**2001a**). "Perceptions of simulated stuttering and fluency," J. Speech Lang. Hear. Res. **44**, 61–72.
Susca, M., and Healey, E. C. (**2001b**). "Effects of age and gender on perceptions of stuttering and fluency," poster session presented at the Annual Meeting of the American Speech-Language-Hearing Association, New Orleans.
Umeda, N. (**1977**). "Consonant duration in American English," J. Acoust. Soc. Am. **61**, 846–858.
Van Riper, C. (**1982**). *The Nature of Stuttering*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ).
Whitehill, T. L., Lee, A. S. Y., and Chun, J. C. (**2002**). "Direct magnitude estimation and interval scaling of hypernasality," J. Speech Lang. Hear. Res. **45**, 80–88.
Zebrowski, P. M. (**1991**). "Duration of the speech disfluencies of beginning stutterers," J. Speech Hear. Res. **34**, 483–491.
Zebrowski, P. M. (**1994**). "Duration of sound prolongation and sound/syllable repetition in children who stutter: Preliminary observations," J. Speech Hear. Res. **37**, 254–263.

Acoustic variability within and across German, French, and American English vowels: Phonetic context effects

Winifred Strange,[a] Andrea Weber,[b] Erika S. Levy,[c] Valeriy Shafiro,[d] Miwako Hisagi,[e] and Kanae Nishi[f]

Ph.D. Program in Speech and Hearing Sciences, Graduate School and University Center, The City University of New York, 365 Fifth Avenue, New York, New York 10016-4309

(Received 30 January 2007; revised 14 May 2007; accepted 25 May 2007)

Cross-language perception studies report influences of speech style and consonantal context on perceived similarity and discrimination of non-native vowels by inexperienced and experienced listeners. Detailed acoustic comparisons of *distributions* of vowels produced by native speakers of North German (NG), Parisian French (PF) and New York English (AE) in citation (di)syllables and in sentences (surrounded by labial and alveolar stops) are reported here. Results of within- and cross-language discriminant analyses reveal striking dissimilarities across languages in the spectral/temporal variation of coarticulated vowels. As expected, vocalic duration was most important in differentiating NG vowels; it did not contribute to PF vowel classification. Spectrally, NG long vowels showed little coarticulatory change, but back/low short vowels were fronted/raised in alveolar context. PF vowels showed greater coarticulatory effects overall; back and front rounded vowels were fronted, low and mid-low vowels were raised in both sentence contexts. AE mid to high back vowels were extremely fronted in alveolar contexts, with little change in mid-low and low long vowels. Cross-language discriminant analyses revealed varying patterns of spectral (dis)similarity across speech styles and consonantal contexts that could, in part, account for AE listeners' perception of German and French front rounded vowels, and "similar" mid-high to mid-low vowels. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2749716]

PACS number(s): 43.70.Kv, 43.70.Fq, 43.71.Hw, 43.71.Es [ARB] Pages: 1111–1129

I. INTRODUCTION

There has been ongoing interest over the last twenty years in investigating the perception of non-native vowels by naïve listeners (listeners with no experience with the target language) and second-language (L2) learners with varying amounts of L2 experience. For instance, several studies have investigated the perception by American English (AE) listeners of front, rounded vowels that contrast with both front, unrounded vowels and back, rounded vowels in German and French (AE contrasts only front, unrounded and back, rounded vowels). An early cross-language categorical perception study with synthetically generated isolated vowels suggested that these non-native vowel contrasts were not problematic perceptually (Stevens *et al.*, 1969). However, more recent studies with talker-generated vowels produced and presented in one or more phonetic contexts have shown significant and persistent difficulties for both naïve listeners and L2 learners in perceptual differentiation of some front, rounded vs back, rounded vowel contrasts, as well as contrasts among front, rounded vowels (Best *et al.*, 2003; Flege, 1987; Gottfried, 1984; Levy and Strange, in press; Polka, 1995). These studies have also reported perceptual difficulties even for L2 vowel contrasts that also occur in the listeners' native language (L1), but are phonetically realized differently across languages. For instance, Gottfried (1984) reported that naïve American listeners had trouble discriminating French [i/e] and [e/ɛ].

In attempting to predict and explain these cross-language perceptual difficulties, it becomes clear that an abstract phonological description of vowels fails to capture significant cross-language differences in these phonetic categories. Rather, to characterize cross-language (dis)similarities as they affect perceptual performance, the vowels of each language must be described with respect to their actual articulatory (and resulting acoustic) realization in various phonetic contexts. This paper reports on a detailed acoustic analysis of vowel corpora produced by native speakers of North German, Parisian French, and New York English. Distributions of vowels produced in a variety of phonetic contexts were analyzed to document similarities and differences in their systematic acoustic variation in the three languages. Using linear discriminant analysis, the contribution of spectral and temporal parameters to the acoustic differentiation of vowels within each language was assessed. In addition, *cross-language* discriminant analyses were used to evaluate cross-language acoustic (dis)similarities and how they

[a]Electronic mail: strangepin@aol.com
[b]Present address: Saarland University, Building C7.1, Rm. 1.1, 66041 Saarbrücken, Germany.
[c]Present address: Department of Biobehavioral Sciences, Program in Speech and Language Pathology, Box 180, Teachers College, Columbia University, 525 W. 120th St., New York, NY 10027.
[d]Present address: Communication Disorders and Sciences, Rush University Medical Center, 203 Senn 1653 W. Congress Parkway, Chicago, Illinois 60612.
[e]Present address: Research Laboratory of Electronics and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Rm. 36–581, 50 Vassar St., Cambridge, MA 02139–4307.
[f]Present address: Boys Town National Research Hospital, 555 N. 30th St., Omaha, NE 68131

changed across different speech styles and phonetic contexts. The productions of male and female speakers were analyzed separately to mitigate issues of vocal tract normalization and to provide a replication of cross-language patterns of acoustic (dis)similarities.

The usefulness to speech perception/production research of such detailed acoustic descriptions covering complete phoneme inventories of languages needs no special advocacy. On the basis of such data, it is possible to provide insights into the patterns of perception, not only by native listeners of the languages but also by non-native listeners. Our present aim was to provide such essential acoustic descriptions for North German, Parisian French, and New York English. Previous research reported difficulties by AE listeners in discriminating German (Polka, 1995) and French (Gottfried, 1984; Levy and Strange, in press) vowels; thus, an acoustic analysis of the vowel inventories allowed us to determine the extent to which perceptual difficulties might be explained by cross-language patterns of acoustic (dis)similarity (Flege, 1995).

In most previous studies of cross-language vowel perception, stimuli were generated by recording (di)syllables spoken in lists (citation materials) in which the surrounding consonants were fixed or absent (e.g., Gottfried, 1984; Polka, 1995). A few studies have used materials in which the target (di)syllables have been produced and presented in short carrier sentences at a rate more closely approximating continuous speech (e.g., Levy and Strange, in press; Strange *et al.*, 2005). Citation productions can be considered a form of "clear speech" in which the vowels may be produced with more acoustic/articulatory distinctiveness than in continuous speech utterances (see Stack *et al.*, 2006). Following the classic study by Peterson and Barney (1952) of American vowels, the hVd context is often used. In English (and German), the initial /h/ can be considered a voiceless onset of the vowel; the tongue/lips/jaw assume the position for the vowel from the beginning of the utterance. (French does not have an initial /h/, but all vowels can occur in word/syllable-initial position.) The following /d/ shows some coarticulatory influence on the spectral characteristics of the vowel toward the end of slowly-produced syllables, but the midpoint formant values are often taken to be the "canonical target" specification of the vowel, independent of any coarticulatory influence from the surrounding consonants.

Studies of AE vowels (Hillenbrand, Clark, and Nearey, 2001; Stevens and House, 1963) have shown that midpoint formant frequencies of vowels produced in CVC syllables vary significantly with the place-of-articulation of the surrounding consonants. Other studies have documented the effects of consonantal context on mid-syllable formant values for vowels of other languages (see Bohn, 2004 for North Frisian; Steinlen, 2005, for North German, Danish, and British English; Strange *et al.*, 2005, for North German). In the present study, we recorded citation (di)syllables of the form (h)Vb(ə) to establish the target values for vowels in each language. The following /b/, because it does not require articulation of the tongue, was thought to be a better choice than the coronal /d/. The same speakers then produced vowels in multisyllabic nonsense words /cvCVC(ə)/ medially within short carrier sentences; the immediately surrounding consonants included both labial and alveolar stops /bVb, bVp, dVd, dVt/ because previous studies had determined that these places of articulation encompassed the greatest variation in vowel formants in AE and North German.

A. Vowel inventories for North German, Parisian French, and American English

1. North German vowels

German is described as having 15 distinctive monophthongal vowels: front, unrounded [iː, ɪ, eː, ɛ, ɛː]; back, rounded [uː, ʊ, oː, ɔ]; front, rounded [yː, ʏ, øː, œ]; and low (back, unrounded) [ɑː, a]. However, in earlier work with North German (NG), it was determined that [ɛː] was only produced in hypercorrect speech; thus, it was excluded from the present analysis (Steinlen, 2005; Strange and Bohn, 1998). The remaining 14 vowels can be grouped into seven pairs of long (tense, close) vs short (lax, open) vowels. However, in NG, these long/short pairs differ significantly in tongue height as well as in vocalic duration, except for [ɑː/a] (Steinlen, 2005; Strange *et al.*, 2004a; 2005). Thus, in terms of phonological descriptions, NG vowels can be said to contrast in tongue height, tongue position, lip rounding, and vowel length/tenseness. Phonetically, tongue height and vowel length redundantly distinguish six pairs: [iː/ɪ, eː/ɛ, uː/ʊ, oː/ɔ, yː/ʏ, øː/œ], while [ɑː/a] are distinguished only by length.

2. Parisian French vowels

French is described as having ten distinctive oral vowels: front, unrounded /i, e, ɛ/; back, rounded /u, o, ɔ/; front, rounded /y, ø, œ/; and the low central vowel /a/.[1] The mid and mid-low, front, rounded vowels [ø, œ] contrast in only a few French words, and have merged in present-day Parisian French (PF), with some allophonic variation as a function of phonetic context. Thus, in this study, a single symbol [ø] is used to indicate the mid, front, rounded vowel in all contexts, in contrast with the high [y]. The three front, unrounded vowels and the three back, rounded vowels differ in tongue height (high, mid, mid-low, respectively). Vowel length is arguably not phonetically functional in PF vowels. Only the pair [oː, ɔ] is said to vary in length systematically, and then only in closed syllables, which make up a small proportion of lexical items in French (Gottfried and Beddor, 1988). Furthermore, Gottfried and Beddor reported that native French listeners did not use vocalic duration to distinguish these vowels perceptually, although they did vary systematically in production (but see Miller and Grosjean, 1997; Miller *et al.*, 2000, for a different perceptual pattern in Swiss French). Thus, unlike German, in which vowel duration appears to have both a phonological and a phonetic function, French vowels appear not to be distinguished phonologically or perceptually by vocalic duration differences.

3. New York English vowels

AE is described as having 11 nonrhotic distinctive vowels: front, unrounded [iː, ɪ, eː, ɛ, æː]; back, rounded [uː, ʊ, oː, ɔː]; and the mid-low and low (back, unrounded) vowels

[ʌ, ɑː]. It does not have distinctive front, rounded vowels, although in coronal consonantal contexts, there is allophonic "fronting" of back, rounded vowels (Hillenbrand *et al.*, 2001; Strange *et al.*, 2005). Front and back vowels vary in tongue height (high, mid-high, mid, mid-low, low) in similar ways to NG. While vowel length is not considered phonologically distinctive, AE vowels vary systematically in intrinsic duration, as indicated by the [ː] above (Peterson and Lehiste, 1960). Note that, while front vowels alternate in duration as a function of tongue height, the back vowels [oː, ɔː, ɑː] do not. In some dialects of AE (including Canadian English), [ɔː, ɑː] have merged into a single mid-low, back, slightly rounded [ɒː]. However, in New York English, these vowels are distinct, with [ɔː] rounded and higher than unrounded, low [ɑː]. The mid, front and back vowels are described as diphthongized [eᴵ, oᵘ] especially in open syllables, and several other vowels have been shown to have perceptually-relevant vowel-intrinsic spectral change (VISC) throughout the vocalic nucleus (Nearey and Assmann, 1986). These changes in formant structure within the vocalic nucleus are most apparent in slowly-articulated citation utterances, and vary across dialects of AE.

B. Language-independent and language-dependent coarticulatory patterns

From these traditional phonological/phonetic descriptions of vowels in the three languages, several speculations can be made about their cross-language (dis)similarities. The front, rounded vowels of German and French can be considered non-native for AE listeners. However, NG front, rounded vowels are sometimes described as more central than PF front, rounded vowels (Steinlen, 2005) and AE back, rounded vowels that are fronted in coronal contexts are phonetically quite similar to NG front, rounded vowels (Strange *et al.*, 2005). To our knowledge, no direct comparisons of the acoustic structure of these American, French, and German vowels, produced in a variety of consonantal contexts, have been reported (but see Hay *et al.*, 2006). Other vowels, which are often transcribed as the "same" in all three languages, may nevertheless differ in their relationships to other vowels in the language and in their dynamic spectral structure. Mid, front and back vowels /e/ and /o/ are monophthongal in German and French, while they tend to be diphthongized in AE. Furthermore, previous comparisons of NG and AE vowels have shown that these mid vowels differ in relative height, with the NG mid vowels higher (i.e., closer to NG high vowels) than for AE vowels. The low vowels in German [ɑː, a] and French [a] may differ in both tongue height and backness from AE low and mid-low [ɑː, ʌ]. Finally, some vowels vary across languages in their temporal structure; [ɔ] is short in German and French, while AE [ɔː] is long. Furthermore, the effect of tongue height on relative durations of vowels is also greater for NG than for AE vowels (Strange and Bohn, 1998; Strange *et al.*, 2004a, 2005). Thus, a comparison of the vowels of these three languages, produced in the same phonetic contexts, was necessary to establish perceptually-relevant phonetic (dis)similarities across languages.

Of specific interest here was how *contextual variability* in the spectral and temporal structure of vowels differed across languages. Classical theories of vowel coarticulation (phonetic vowel reduction) posit the concept of "target undershoot" to predict/explain the variation in mid-syllable formant values in different CVC syllables produced at different speaking rates (Lindblom, 1963). According to this simple dynamic model, as vowels are coarticulated with consonants, the specified target position of the vowel in vowel space remains invariant, but the temporal overlapping of commands to the articulators for consonant and vowel gestures leads to a failure of the articulators to reach the canonical target before the command for the following consonant takes effect. The amount of articulatory (and resulting acoustic) undershoot varies as a function of the difference between consonant loci and vowel target (locus/target distance) and increases at faster speaking rates. If these dynamic consonant/vowel interactions were the only influence on vowel production, variation in the spectral structure of vowels in the same consonantal contexts should be very similar across languages. However, Moon and Lindblom (1994) showed that the amount of target undershoot in AE vowels differed with speech style (clear vs citation) and other studies of AE (Fourakis, 1991; Stack *et al.*, 2006) and Dutch (van Son and Pols, 1992) have failed to show systematic formant changes in vowels over different speaking rates in read continuous speech utterances. With respect to the influence of phonetic context on the spectral structure of vowels, Steinlen (2005) reported that variability of both first (F1) and second formant (F2) values for vowels in the same consonantal contexts differed across languages as a function of the size of the vowel inventory, with Danish (20 vowels; 10 long/short pairs) showing the least variability and British English (11 vowels) showing the most variability. Bohn (2004) has proposed that the amount of contextual shrinking of the vowel space is inversely related to inventory size, although the relationship may not be linear.

Given these differences within and across languages, we hypothesized that contextual variability in the spectral structure of vowels is, to a significant extent, actively controlled by speakers. That is, much of the coarticulatory variation in vowels is due to learned patterns of production (i.e., language-specific coarticulatory constraints) that serve to maintain perceptual distinctiveness even at quite rapid rates of speaking (Diehl and Lindblom, 2004; Lindblom, 1990). We therefore expected to find differences across languages in the distributional characteristics of mid-syllable spectral structure when the vowels were produced in various phonetic contexts. As a consequence, these cross-language differences in variation would result in different patterns of *cross-language* spectral similarity in different phonetic contexts. Because of the different phonological/phonetic functions of vocalic duration across the three languages, we also expected to find language-specific patterns of temporal variation with changes in phonetic context and speech style (Hay *et al.*, 2006).

II. METHOD

A. Speakers

Three female and three male monolingual speakers of each language were selected from a larger set of speakers initially recorded, including 10 German speakers (6 females, 4 males); 17 French speakers (9 females, 8 males), and 11 American speakers (3 females; 8 males). Reasons for removing speakers were based on language background questionnaires and judgments of phonetically trained native speakers of NG, PF, and New York English (hereafter AE), respectively. Reasons included: (a) wrong dialect or too bilingual (8 speakers), (b) dyslexia or trouble reading the orthographic representations of vowels (6 speakers), and (c) "sloppy" speech (1 speaker). An additional 5 speakers were not used because of failure to follow directions or because of difficulties with acoustic analysis (unclear formants). The 18 selected speakers ranged in age from 20 to 48 years old (NG 22–29 years; PF 20–36 years; AE 33–48 years). Although all speakers were recorded in New York City, the German and French speakers were either visitors (9 speakers) or very short-term residents of the US (2 subjects, 1 month; 1 subject, 3.6 months). French and German speakers (and some Americans) had formal foreign language classes in high school and/or college, but reported that they were not able to converse in any non-native language. The language background questionnaires established that the speakers had spoken the appropriate dialect throughout their lives, with little exposure to other dialects or languages in their immediate families.

B. Stimulus materials

1. German

The 14 vowels were represented by standard German orthography, which specifies phonetic vowel identity unambiguously. For citation utterances, vowels were embedded in disyllables hVb(ə), with primary stress on the target vowels: /i/Hieba, /ɪ/Hibba, /e/Hehba, /ɛ/Hebba, /u/Huhba, /ʊ/Hubba, /o/Hohba, /ɔ/Hobba, /y/Hühba, /ʏ/Hübba, /ø/Höhba, /œ/Höbba, /ɑ/Hahba, /a/Habba. Each vowel was assigned a number (1–Hieba, 2–Hibba, etc.) in order to assure that the identity of the intended vowel was retained during acoustic analysis and to facilitate random sequencing of items. For sentence utterances, target nonsense trisyllables /gəCVCə/ were embedded in the sentence, "Ich habe fünf _____ gesagt." (I said five _____).

Protocols consisted of randomized lists of 15 utterances; the first and last utterance contained the same target vowel and the final utterance was discarded to control for list-final intonation effects. Pages of the protocol were arranged such that the first page (two lists) was always the citation utterances, and the next 4 pages were the sentences, with the consonantal context fixed for each page.[2] The second half of the protocol was identical to the first half except that the vowels appeared in different random orders. Thus, speakers recorded two exemplars of each vowel in each context in the first half of the session, and then repeated the whole protocol for a total of four utterances of each vowel in each context.

2. French

Pilot testing indicated that French speakers stressed the target vowel only when it occurred in the final syllable of the nonsense word. However, final consonants are heavily released in French. In addition, the vowel /e/ does not occur in closed syllables. Thus, for citation utterances, the protocols contained words spelled in French orthography as follows: /i/ hibe, /e/ héb'a, /ɛ/ hèbbe, /u /hoube, /o/ haube, /ɔ/ hobe (botte), /y/ hube, /ø/ heube, /a/ habe. These words were produced as Vb(ə) syllables with an audible, schwa-like release. For sentence materials, disyllables /raCVC/ (e.g., "rabibes, rabipes, radides, radites") were embedded in the sentence, "J'ai dit neuf _____ à des amis" (I said nine _____ to some friends). The /r/ is produced as an uvular fricative in French, and so has a similar place of articulation to /g/ in German and English. Since French does not have reduced vowels, the preceding/following /a/ was chosen as comparable in front-back position to schwa in German and English. The organization of the protocols was the same as for German.

3. English

For citation disyllables, words were spelled: /i/ heeba, /ɪ/ hibba, /e/ hayba, /ɛ/ hebba, /æ/ habba (hat), /u/ hooba, /ʊ/ hUba (should), /o/ hoaba (road), /ɔ/ hawba, /ɑ/ hobba, /ʌ/ hubba. As noted, real words were used to aid in pronouncing three vowels with very ambiguous/variable spelling in English. For sentence materials, the trisyllables /gəCVCə/ were embedded in the sentence: "I said five _____ this time." The spelling of the vowels was the same as above and real-word prompts were printed above the nonsense word in the sentences for [æ, ʊ, o]. Pages were arranged the same as for the German and French protocols.

C. Recording procedures

Participants were recorded while seated in an IAC chamber with the experimenter outside in visible contact through the window. An intercom allowed the experimenter to converse with the speaker while monitoring the recording input over headphones (Sennheiser HD565 Ovation). Output from a dynamic microphone (Shure SM48), placed about 15 cm from the speaker's mouth, was fed to a microphone preamplifier (Earthworks Lab 101) and then directly into a Soundblaster Live Wave DF80 sound card of a PC computer (Dell Dimension XPS B800). The stimuli were digitized (22.05 kHz, 16-bit) as monaural computer files, using SoundForge™ 5.0 software. The preamplifier level was set such that input signals varied from the equivalent of −18 to −3 dB of maximum capacity to preclude peak clipping of any part of the signal. Signal/noise-floor ratios of digital recordings were thus about 30–40 dB.

The experimenter spoke only in the speaker's native language throughout the entire session. Speakers were given practice with the protocols with feedback about the style in which they were to speak. For citation utterances, they read through the items and the experimenter answered questions about pronunciation of the nonsense words. The speakers were instructed to read the number preceding each nonsense

word, then pause, and read the nonsense word with falling intonation (demonstrated by the experimenter). The speakers then practiced reading the list of citation utterances while the experimenter set the gain control for the preamplifier. The experimenter corrected their productions if they clearly misread the nonsense items, failed to read the identifying number, or did not pause between the number and the target word. Finally, speakers were instructed to speak a little more loudly than they would normally in a quiet room, as if they were talking through the window to the experimenter. They were also instructed not to speak too carefully, but just to say the words normally. The speakers then read the first column of citation utterances; if any were clearly mispronounced or if they read the wrong number, the experimenter asked them to repeat that utterance (referred to by number). They then progressed to the second column of utterances, with corrections at the end if needed.

After the citation materials were recorded, speakers were familiarized with the sentence materials and were given practice reading them. They were instructed to read the sentences as if they were talking to a friend who was a native speaker of the language and not to emphasize the nonsense word or put any pauses in the sentence. After the speakers had practiced until they read the sentences fluently, recording began. Again, they were reminded to speak a little more loudly, but not too slowly or too over-correctly, and they were told that they could reread a sentence at any time (self correct). Speakers were allowed to take short breaks and drink water any time during the recording session. The entire session, including reading and signing the Human Subjects consent form and filling out the language background questionnaire took approximately 2 hours.[2] Subjects were paid for their participation.

D. Acoustic analysis procedures

Digitized strings of utterances were separated into individual files and after coding, the spoken identifying numbers were removed from the files. Vowel productions were independently verified as "good" exemplars of the intended vowels by a native speaker (not one of the experimenters or speakers) of each language. Acoustic analysis was accomplished in a two-step process, using a specialized Matlab™ 6.0 program (CVCZ by Valeriy Shafiro). First, the waveform and a time-synchronized wideband spectrogram (0–4 kHz) were displayed and target syllable onsets and offsets were determined manually by the experimenter: for citation utterances, onset was defined as the onset of voicing (i.e., not including the voiceless /h/ for NG and AE utterances), and offset was defined as the beginning of /b/ closure, determined from the offset of upper formant energy and decrease in waveform amplitude. For CVC syllables, onset was defined as the release burst of the /b, d/, and offset was defined as the beginning of closure of the final stop (and in the case of flap [ɾ] for /d/ and /t/ in English, the point at which amplitude reached a minimum). Thus, vocalic duration included the entire gesture from release of the preceding consonant to the beginning of full closure for the following consonant.

Second, values for the first three formants (F1, F2, F3) were determined for three 23-ms windows placed at 25%, 50%, and 75% through the vocalic duration of the syllable using FFT and LPC analysis. The three FFT spectra (512 points) were displayed from left to right, with formant center frequencies estimated from LPC algorithms (24 coefficients) superimposed as vertical lines on the spectra. When estimated formants were clearly in error, hand corrections were made; three kinds of errors occurred: (a) formants were merged when two formants were very close together, (b) the third formant was missed when amplitudes were low, and (c) occasionally, spurious formants were identified.[3] Corrections were based on the experimenter's judgments from comparisons of FFT spectra and LPC formant tracks superimposed on the spectrographic display. All hand-corrected formant frequencies were obtained independently by two experimenters and any discrepancies were resolved by a third experimenter (the first author). Approximately 5% of the measurements of F1, 10% of the measurements of F2, and 17% of the measurements of F3 were obtained by hand corrections. For each speaker, mid-syllable (50% point) formant frequencies of the four repetitions of each vowel in each context were plotted in F1/F2 and F1/F3 plots for a final check for measurement errors. For citation utterances, F1 and F2 formant trajectories (25%–50%–75%) were also plotted to note direction and extent of formant movement throughout the middle half of the syllables indicating diphthongization (vowel-inherent spectral change).

III. RESULTS

To provide an initial overview of the variability of vowels in the three languages, Fig. 1 presents scatter plots of mid-syllable formant frequencies (F1/F2 in Bark) for all tokens produced by the three male speakers of each language (female plots look very similar). Thus, there are 60 tokens of each vowel, including the citation utterances and the vowels in sentences in bVb, bVp, dVd, and dVt contexts (5 contexts × 3 speakers × 4 repetitions). Coordinates are arranged so that high vowels appear at the top of the plots, front vowels to the left.

As the plots show, there was extensive contextual variability in the mid-syllable formant frequencies of vowels in all three languages. In general, the back vowels varied more than front vowels in F2 (front-back dimension), while the low vowels varied more than the high vowels in F1 (tongue height dimension). There was considerable overlap in F2 of front, unrounded and rounded vowels for both PF and NG, with generally less overlap of front and back, rounded vowels. There was considerable overlap in F1 for so-called high vowels [i, y, u] and mid vowels [e, ø, o] in all three languages, between so-called mid-high [ɪ, ʏ, ʊ] and mid vowels in NG and AE, and between mid vowels and mid-low vowels [ɛ, ʌ, ɔ] in PF and AE.

In the following sections, comparisons of vowels in citation utterances are presented first (Sec. A). Then analyses of sentence materials (Sec. B) explore the effects of consonantal context on vowels within and across languages. In each section, figures presenting average formant data for male and female utterances are shown for each language;

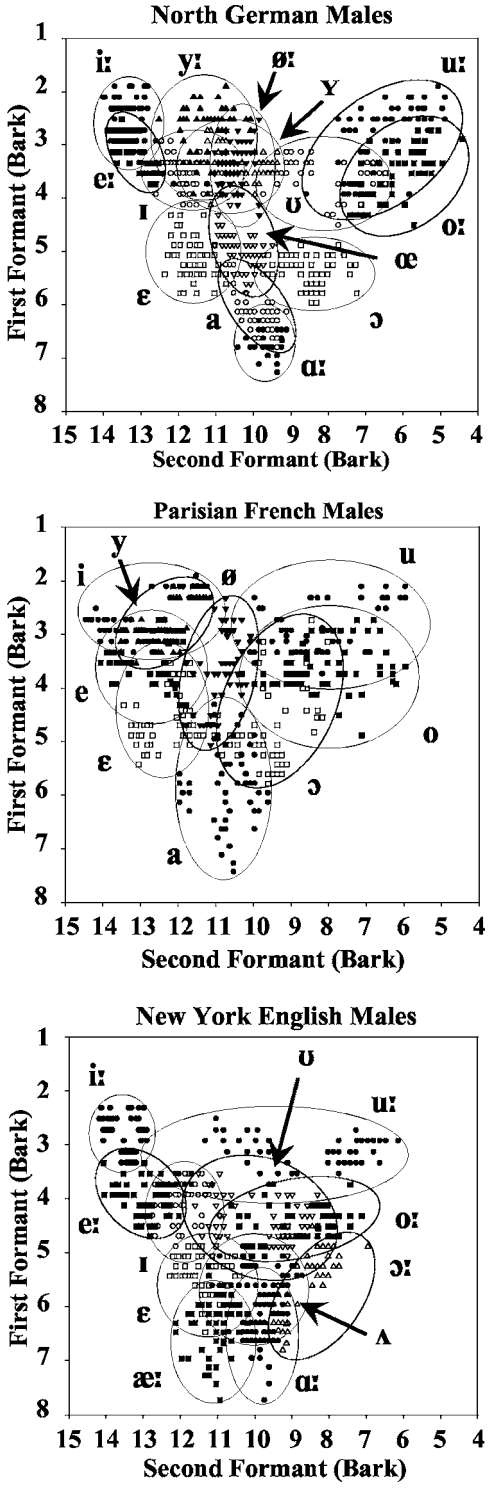


FIG. 1. F1/F2 plots (Bark) of all vowels produced by male speakers of North German (top), Parisian French (middle), and New York English (bottom). For North German and New York English, short vowels are plotted with open symbols; for Parisian French, mid-low vowels are plotted with open symbols. Ellipses were placed by hand to surround all tokens of each vowel.

then within-language discriminant analyses of each language are presented. These analyses quantify the amount of spectral overlap of mid-syllable formant frequencies and the extent to which vowels are differentiated by vocalic duration differences. Finally, cross-language discriminant analysis results

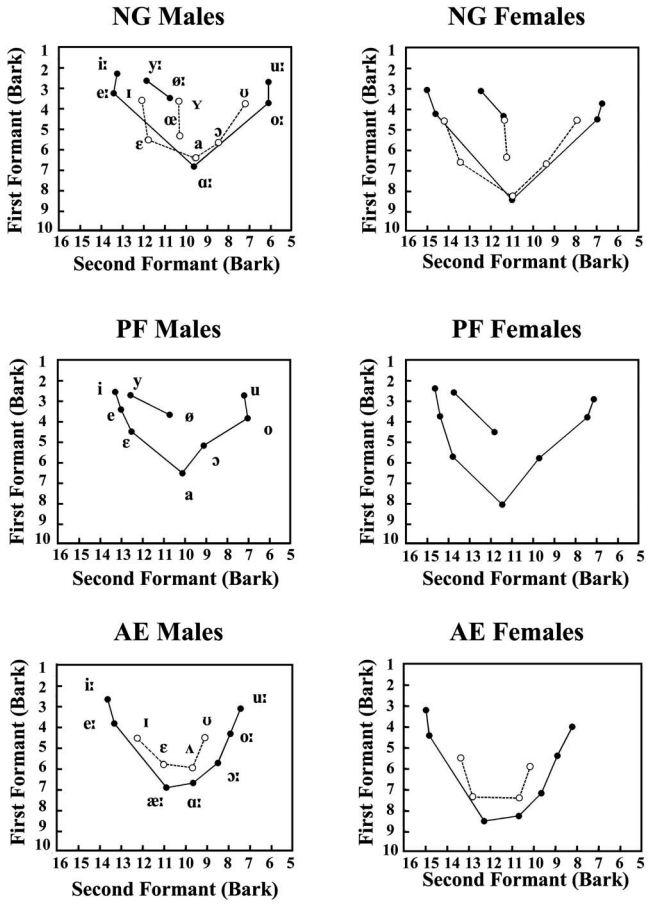


FIG. 2. Average mid-syllable F1/F2 values (Bark) for North German (NG), Parisian French (PF) and New York English (AE) male (left) and female (right) vowels produced in citation utterances. For NG and AE, short vowels are plotted with open circles.

are presented as a quantitative description of the acoustic similarity of NG and PF vowels to AE vowels.

A. Spectral and temporal structure of vowels in citation materials

Inspection of the results of acoustic analysis indicated that, for all three languages, the three speakers within each gender group showed quite similar patterns in the relative location of vowel targets (mid-syllable formant frequencies) in vowel space and similar relative duration differences.[4] Thus, average Bark values for the first three formants at mid-syllable and vocalic durations (ms) were computed over all 12 tokens of each vowel in citation utterances (4 repetitions × 3 speakers). Figure 2 presents the F1/F2 Bark plots for male (left) and female (right) utterances in all three languages. In these figures, the relative spacing of vowels in terms of tongue height (F1) and front-back position (F2) can be compared across languages. Appendices A–C list average Bark values for F1 and average Bark-difference values (F2-F1 and F3-F2) for each vowel for male and female tokens separately. In addition, the average vocalic durations of each vowel are given in the final two columns.

First, a comparison of the point vowels in the three languages indicates that differences in F1 values (tongue height) for high, front [i], and low [ɑ, a] are quite comparable across

languages (F1 Bark differences for Male: NG=4.5, PF=4.1, AE=4.0; Female: NG=5.1, PF=5.7, AE=5.1). Thus, relative distances of other vowels from these vowels can be compared meaningfully. In comparing NG and AE, it is readily apparent that the so-called mid vowels [eː, øː, oː] are as high as or higher (equal or lower F1 values) than the so-called mid-high vowels [ɪ, ʏ, ʊ]; these vowels are distinguished by vocalic duration; long/short (L/S) ratios ranged from 1.6 to 2.3 for NG, 1.4 to 1.7 for AE. In all three languages, high and mid, long vowels differed in F1 by less than 1.5 Bark for front, unrounded; front, rounded; and back, rounded pairs across both gender groups (NG range=0.9 to 1.4 Bark; PF range=0.9 to 1.3 Bark; AE range=1.2 to 1.4 Bark). These vowels were generally less differentiated by vocalic duration; mid/high duration ratios ranged from 1.1 to 1.3 for NG, 0.8 to 1.2 for PF, and 1.2 to 1.4 for AE.

In comparing the relative locations of vowels on the front-back dimension, languages appeared to differ systematically with respect to the high, back point vowel [u]. Thus, while [i] was very similar in F2 Bark values across languages (Male: NG=13.2, PF=13.3, AE=13.6; Female: NG =15.0, PF=14.6, AE=14.9), NG and PF [u] were further back (lower F2) than the AE [u] (Male: NG=6.0, PF=7.1, AE=7.4; Female: NG=6.7, PF=7.1, AE=8.2).

The front, rounded vowels of both NG and PF were closer to front, unrounded vowels than back, rounded vowels, with the high, front [y] having relatively higher F2 Bark values (Male: NG=11.8, PF=12.5; Female: NG=12.4, PF =13.7) than the other front, rounded vowels [ø, (ʏ, œ)] (Male: NG=10.7, 10.3, 10.2; PF=10.7; Female NG=11.3, 11.3, 11.2; PF=11.8). These comparisons also show that PF [y] had a higher F2 value on average than the "same" vowel in NG, with the PF [y] less than 1 Bark lower in F2 than PF front, rounded [i].

One final observation from these vowel spaces regards the "centralization" of short (lax, open) vowels, relative to long (tense, close) vowels in NG and AE. In general, the short vowels of AE were more centralized on both F2 and F1 dimensions than the short vowels of NG in these citation utterances. However, NG vowels tended to be more differentiated temporally than AE vowels in this context (see Appendices). Notice that for both languages, mid-high and mid-low short vowels are considerably lower in vowel space (higher F1) than their high and mid, long counterparts, respectively.

1. Within-language discriminant analyses

While these comparisons of central tendencies of vowel categories across languages begin to capture important (dis)similarities across languages, a more quantitative comparison of *distributions* of vowels is needed to assess the contributions of mid-syllable formant frequencies and vocalic duration to the acoustic differentiation of vowel categories within each language. Linear discriminant analyses (Klecka, 1980) were performed on male and female vowel sets separately (12 tokens/vowel), using Bark values for F1/F2/F3 as input parameters in one set of analyses, and F1/F2/F3 plus vocalic duration in a second set of analyses of the three languages.[5] This statistical technique optimizes the weighting of input parameters for maximum separation of categories, establishes a center-of-gravity in parameter space for each category, and then yields a classification matrix of all tokens. "Correct" classification indicates that a token of an intended vowel was classified as closest to the center-of-gravity for that category.

TABLE I. Overall results of within-language discriminant analyses (percent tokens correctly classified): Citation disyllables.

Language		Input Parameters F1/F2/F3	Input Parameters Fs + Duration
NG	M	85	95
	F	80	90
PF	M	89	92
	F	93	92
NYE	M	83	93
	F	87	93

Table I presents the overall rates of correct classification for each gender group within each language; the first column gives the results of the analyses with spectral parameters only, while the second column shows the analyses when duration was added as an input parameter. Looking first at the analyses with formants only, overall correct classification rates were best for PF vowels, and somewhat lower for NG and AE vowels. For PF vowels, correct classification of individual vowel categories ranged from 58% to 100%. Confusions were mostly between front, unrounded and rounded vowels [e/y] and [i/y], between mid and mid-low, front [e/ɛ]; and between high and mid, back vowels [u/o]. For NG vowels, correct classification of individual vowel categories ranged from 58% to 100%. Errors were mostly confusions between long and short vowels [iː/ɪ, eː/ɪ, uː/ʊ, oː/ʊ, øː/ʏ, ɑː/a], but high and mid, long vowels [eː/iː, uː/oː, yː/øː] were also sometimes confused, as were front, rounded vowels with front, unrounded or back, rounded vowels [yː/ɪ, ʏ/ʊ]. For AE vowels, correct classification of individual vowel categories ranged from 50% to 100%. As with NG, the majority of misclassifications were confusions between long and short vowels [eː/ɪ, æː/ɛ, æː/ʌ, ɑː/ʌ, oː/ʊ, ɔː/ʊ, uː/ʊ], but confusions also occurred between long vowels [eː/iː, ɑː/ɔː, uː/oː].

Correct classification improved for NG and AE vowels when vocalic duration was included as an input parameter, while classification of PF vowels remained almost unchanged, and misclassification patterns were very similar. For NG vowels, remaining misclassifications were primarily among back vowels [uː, ʊ, oː]; for AE vowels, remaining misclassifications were of [eː/iː], [ɑː/ʌ/ɔː], and [uː/oː].

In summary, distributions of vowels produced in citation utterances were acoustically differentiated quite well overall in all three languages on the basis of mid-syllable formant values and vocalic duration. As expected, vocalic duration was most important in differentiating vowels in NG and least important in PF. In all three languages, the high and mid, back vowels [u/o] were sometimes confused. In PF, front, unrounded and rounded vowels [e/y, i/y] were sometimes confused, while in NG, these confusions did not occur. In AE, confusions occurred between front vowels [iː/eː], while these vowels were not confused in either NG or PF. In PF, the front vowels [e/ɛ] were sometimes confused (cf.

TABLE II. Cross-language spectral similarity of NG, PF, and NYE point vowels: Discriminant analyses (F1/F2/F3 Bark) for citation disyllables. Male and Female results are combined.

NG vowels	Modal AE V	% tokens classified	Other Vs >10%
iː	iː	96	
ɑː	ɑː	96	
uː	uː	96	
a	ɑː	79	ɔː
PF vowels			
i	i	100	
a	æː	41	ɔː, ɑː
u	u	100	

TABLE III. Cross-language spectral similarity of NG and PF front rounded vowels to AE front and back vowels: Discriminant analyses (F1/F2/F3 Bark) for citation disyllables. Male and Female results are combined.

NG vowels	AE front vowels	% test tokens	AE back vowels	% test tokens
yː	ɪ, eː, iː	100		
øː	ɪ	71	ʊ, o	29
ʏ	ɪ	63	ʊ	37
œ	ɛ	54	ʊ, ʌ	46
PF vowels				
y	eː, iː	100		
ø	ɪ, ɛ	88	ʊ	12

Gottfried, 1984), whereas in NG and AE, these vowels were never confused. These differences in patterns of spectral overlap reflect language-specific differences in the relative locations of vowels in vowel space.

2. Cross-language discriminant analyses

For comparisons across languages, we were most interested in how NG and PF vowels would be classified with respect to their acoustic similarity to AE vowels, as an acoustic basis of comparison with perceptual data from English listeners tested on French and German vowels (Levy and Strange, in press; Strange *et al.*, 2005; Strange, Levy, and Lehnhoff Jr., 2004). Thus, the AE vowel corpora served as the training set (upon which parameter weightings and centers-of-gravity were established), then NG and PF vowels served as the test sets. Male and female corpora were analyzed separately. Analyses using F1/F2/F3 Bark values and vocalic duration as input parameters produced poorer matches of "similar" vowels across languages than analyses using only formants. Thus, only the latter analyses that establish the *spectral* similarity of NG and PF vowel categories to AE vowel categories are presented here and in subsequent cross-language analyses.

The two point vowels, NG [iː,uː] and PF [i,u], were consistently classified as most similar to AE [iː,uː], respectively (shown in Table II for male and female data combined; 24 tokens/vowel). Thus, even though average F2 values for [u] differed across languages, NG and PF [u] were not more similar to any other AE vowel. While the NG low vowel [ɑː] was a good fit to AE [ɑː], the distribution of the PF low vowel [a] straddled the front and back AE low vowels [æː, ɑː], and nine tokens were more similar to AE [ɔː]. In contrast, most tokens of NG [a] were classified as AE [ɑː], with only five tokens classified as [ʌ, ɔː].

Having established that [i,u] were acoustically similar across languages, we next asked how NG and PF front, rounded vowels compared acoustically to front and back AE vowels. Table III presents the results of cross-language discriminant analyses for these vowels, combined over analyses of male and female corpora. Overall percentages of tokens classified as AE front vowels, summed over front categories (listed in column 2) and back vowels, summed over back categories (listed in column 4) are given (columns 3 and 5). Results indicated that the NG front, rounded vowels were acoustically intermediate between front and back AE vowels except for [y], which was always classified as more similar to front AE vowels [ɪ,eː,i]. For PF front, rounded vowels, all tokens of [y] were classified as similar to AE front vowels, and only three tokens of [ø] were classified as more similar to AE back vowels. Thus, in general, PF front, rounded vowels were more consistently classified as falling within front AE vowel categories than were NG front, rounded vowels.

Third, it was of interest to look at classification patterns for the remaining NG and PF vowels, often transcribed with the same symbols as for AE counterparts. Table IV presents the results of cross-language discriminant analyses for these vowels, again collapsed over male and female analyses. In this table, the AE vowel category to which each NG and PF vowel was most often classified is given in column 2 with percentages of tokens so classified in column 3. Other AE categories to which at least 10% of the tokens were similar are presented in column 4. As these results indicate, three NG vowels [ɪ,eː,ʊ] and three PF vowels [ɛ,o,ɔ] were not

TABLE IV. Cross-language spectral similarity of NG and PF mid-high, mid, and mid-low vowels to AE vowels: Discriminant analysis (F1/F2/F3 Bark) for citation disyllables. Male and Female results are combined.

NG vowels	AE modal vowel	% test tokens	Other AE vs >10%
ɪ	eː	63	ɪ
eː	iː	54	eː
ɛ	ɛ	54	ɪ
ʊ	oː	54	uː
oː	oː	92	
ɔ	ɔː	96	
PF vowels			
e	eː	63	iː
ɛ	ɪ	84	ɛ
o	u:	50	o
ɔ	ʊ	46	oː, ʌ

classified as most similar to their AE transcriptional counterparts a majority of the time. In the case of NG [ʊ] and PF [ɔ], *none* of the tokens were classified as most similar to AE [ʊ, ɔː], respectively. In contrast, NG [oː, ɔ] were relatively good spectral matches to their AE counterparts, while NG [ɛ] and PF [e] distributions straddled AE [ɛ, ɪ] and [eː, iː] categories, respectively.

In summary, the majority of mid-high, mid, and mid-low NG and PF vowels that are transcribed as the "same" as AE vowels were found not to be spectrally similar to their transcriptional counterparts in AE. In general, NG and PF vowels were acoustically more similar to higher AE vowels, or were distributed such that they overlapped with two spectrally adjacent AE categories. The "non-native" PF front, rounded vowels were spectrally more similar to front than to back AE vowels, while NG front, rounded vowels, except for [yː], fell in between front and back AE vowels. Finally, NG [iː, ɑː, ɔ, oː, uː] and PF [i, u] were most similar spectrally to their transcriptional counterparts in AE.

B. Spectral and temporal structure of vowels in sentence materials

In this section, the acoustic characteristics of coarticulated vowels produced in sentences are reported. Of specific interest was a comparison of (dis)similarities across languages in the contextual variability of spectral and temporal parameters as a function of the consonantal context in which they were produced. Several comparisons were made: (1) Average mid-syllable formant frequencies of vowels produced in labial and alveolar consonant contexts were evaluated with respect to how they differed from canonical values derived from citation utterances, (2) effects of consonantal context on relative duration differences were compared within and across languages, (3) context-dependent and context-independent discriminant analyses established the spectral overlap of coarticulated vowel distributions and evaluated those distributions relative to canonical values, and (4) cross-language discriminant analyses evaluated differences in spectral similarity of distributions of NG and PF vowels to AE vowel categories across contexts.

1. Contextual variation in mid-syllable formant frequencies

Average F1/F2/F3 Bark values for vowels produced in each consonantal context (bVb, bVp, dVd, dVt) were computed for individual speakers of each gender/language group. Within groups, patterns of variation in formant structure across contexts were very similar for the three speakers; thus, average values were computed for comparison with canonical values obtained from citation materials. Inspection of the F1/F2 and F1/F3 plots for vowels produced in the four contexts revealed that patterns for bVb and bVp contexts were almost identical, as were patterns for dVd and dVt contexts. Thus, Figs. 3 and 4 present average F1/F2 values for vowels in labial and alveolar contexts, collapsed over the two labial and two alveolar final consonants, respectively (24 tokens/vowel: 2 contexts × 3 speakers × 4 repetitions). As a refer-

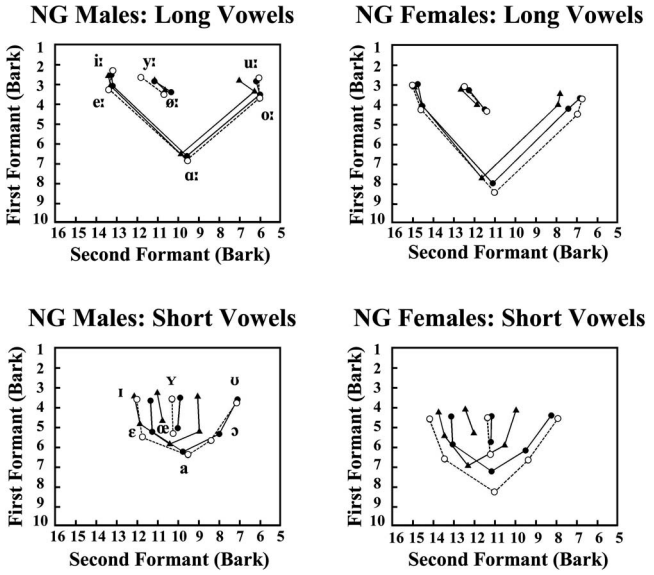


FIG. 3. Average mid-syllable F1/F2 values (Bark) for North German (NG) vowels produced in labial (closed circles) and alveolar (closed triangles) consonant contexts in sentence materials. Average values for long (top) and short (bottom) vowels for male (left) and female (right) speakers are plotted separately. For comparison, average values for citation utterances (open circles) are also shown (same as Fig. 2).

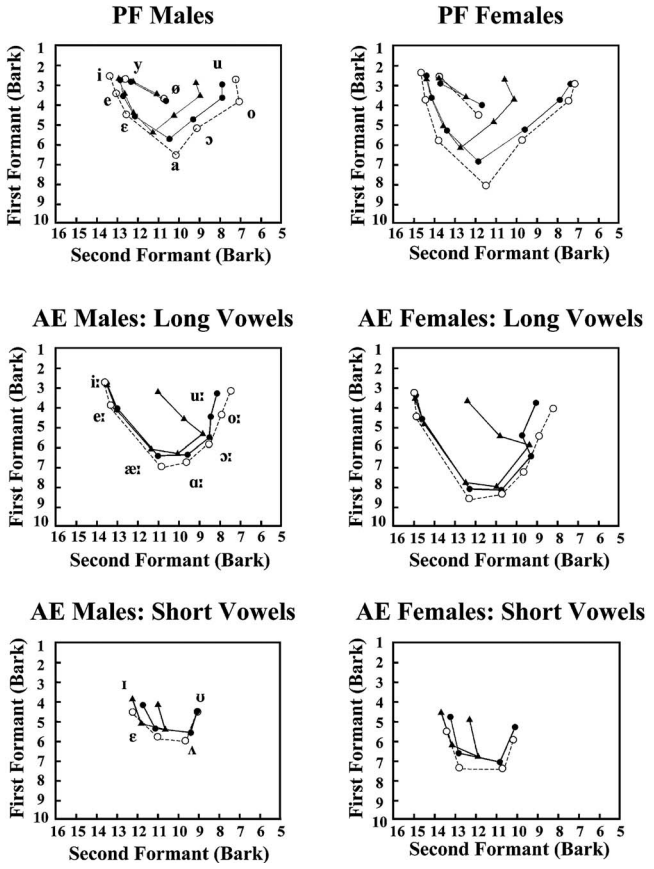


FIG. 4. Average F1/F2 values (Bark) for Parisian French (PF) vowels (upper) and New York English (AE) long vowels (middle) and short vowels (lower) produced in labial (closed circles) and alveolar (closed triangles) consonant contexts in sentence materials. For comparison, average values for citation utterances (open circles) are also shown (same as Fig. 2).

ence, the average values for citation utterances, shown in Fig. 2, are reproduced in these figures, indicated by open symbols and dashed lines.

Figure 3 presents the F1/F2 Bark plots for the NG corpora, with long vowels and short vowels plotted separately for clarity. It is immediately clear that variation in target formant frequencies as a function of coarticulation with labial and alveolar consonants was considerably greater for short vowels than for long vowels for both gender groups, and in general, greater in alveolar than in labial contexts. A second obvious trend is that contextual variation of low and back vowels was greater than for front vowels. Figure 4 presents the F1/F2 Bark plots for the PF corpora (top row), and the AE long vowels (middle row) and short vowels (bottom row), separately. As with the NG vowels, the PF low and back vowels showed greater contextual variation than the front vowels. Second, PF vowels varied considerably more than NG long vowels and perhaps even somewhat more than NG short vowels, especially in alveolar contexts. In contrast, AE long vowels, like NG long vowels, varied relatively little, except for [uː,oː], which showed extreme fronting (higher F2) in alveolar contexts, while AE short vowels showed somewhat larger changes from canonical values.

To quantify these shifts from canonical values more precisely, and to incorporate changes in F3 (not shown in Figs. 3 and 4), average F1/F2/F3 Bark values for each vowel in labial and alveolar contexts for each gender group (24 tokens) were rendered into Bark-difference values (F2-F1 and F3-F2). Then, the Euclidian distance between these values and the average Bark-difference values for citation utterances was computed for each vowel (see Appendices). These Euclidian distances thus represent shifts of vowels in a formant-ratio vowel space as a function of coarticulatory undershoot of canonical acoustic targets (see Fourakis 1991, for a similar analysis).

As Figs. 3 and 4 suggest, shifts from canonical values for coarticulated vowels in labial contexts were, on average, relatively small in all three languages. Euclidian distances for NG vowels across all vowels and both gender groups averaged 0.5 Bark (range=0.1 to 1.3 Bark). For PF vowels the Euclidian distances were also quite small (mean=0.6 Bark, range=0.2 to 1.7 Barks), as they were for AE vowels (mean=0.6, range=0.1 to 1.4 Barks). The largest shifts were for NG [ɪ,a], PF [a], and AE [oː,uː].

In contrast to the relatively small coarticulatory undershoot of vowels in labial contexts, all three languages showed consistently larger contextual effects for vowels in alveolar contexts. Table V presents mean Euclidian distances of vowels in alveolar contexts from canonical values for all three languages (male/female data shown separately). Front, unrounded vowels (rows 1–4) shifted relatively little from canonical values; only NG and AE [ɛ] shifted more than 1 Bark. NG front, rounded short vowels [ʏ, œ] and female PF [ø] (rows 6–8) showed slightly greater fronting in this context. In contrast, most of the low and back vowels in all three languages showed much greater shifts from canonical values, and striking differences in coarticulatory patterns across languages were revealed.

The different patterns of coarticulatory change for back and low vowels in alveolar contexts, illustrated in Figs. 3 and 4, are borne out in this analysis. While NG long vowels [ɑː,oː,uː] varied relatively little from canonical values, the short vowels [a,ɔ,ʊ] shifted more. Thus, there was shrinking of the NG short vowel space, with low and back vowels raised and fronted, while front vowels remained relatively stable. This resulted in an increase in the relative distances between long and short, back and low vowels. For PF, all four low and back vowels differed from canonical values by 2 or more Barks. The shrinking of the vowel space was greater than for NG short vowels; however, the shape of the vowel space (i.e., relative locations of vowels) remained similar across these contextual variations. For AE vowels, the Euclidian distances from canonical values differed strikingly across vowels, illustrating that the *shape of* the vowel space changed for vowels produced in alveolar contexts. Whereas low and mid-low vowels [æː,ɑː,ʌ,ɔː] shifted up and forward relatively little, the mid to high back vowels [oː,ʊ,uː] shifted drastically forward, becoming very similar to front, rounded vowels in NG.

TABLE V. Euclidian distances (Bark) from canonical values for NG, PF, and NYE vowels in alveolar context: Male/female data are presented separately (M/F).

Vowel front	NG	PF	NYE
iː	0.3/0.2	0.6/0.5	0.4/0.4
eː	0.2/0.4	0.5/0.2	0.5/0.7
ɪ	0.3/0.3	—	0.8/0.2
ɛ	0.8/1.2	0.5/0.6	1.6/1.5
yː	0.9/0.1	0.5/0.4	—
øː	0.2/0.9	0.6/1.6	—
ʏ	1.2/1.7	—	—
œ	1.2/**2.0**	—	—
Low/back			
æː	—	—	1.4/1.1
ɑː	0.8/1.6	—	1.0/0.9
a	1.7/**2.9**	**2.3/3.3**	—
ʌ	—	—	1.9/**2.2**
ɔ/ɔː	1.4/**2.3**	**2.0/2.6**	1.0/1.1
ʊ	**2.9/3.2**	—	**3.2/3.8**
oː	0.7/1.7	**2.9/3.6**	**2.7/2.9**
uː	1.3/1.7	**2.5/5.0**	**5.0/6.3**

2. Effects of speech style and phonetic context on vocalic durations

Mean durations of vocalic nuclei for male and female tokens were computed for vowels produced in bVb, bVp, dVd, and dVt context in sentences and were compared with durations of vowels produced in citation utterances. In addition, long/short (L/S) duration ratios for NG vowels (7 L/S pairs) and AE vowels (4 L/S pairs) were compared in citation and sentence utterances. These descriptive data are presented in Table VI. Regarding NG vowels, several generalizations can be noted. First, long vowels in sentences in bVb, dVd, and dVt contexts were only slightly shorter, on average, than these vowels in citation utterances, while short vowels were generally slightly longer in sentences than in citation utter-

TABLE VI. Vocalic durations (ms) and Long/Short ratios in citation and sentence utterances. Male and female data are listed separately (M/F).

Language/ context	Long (ms)	Short (ms)	Long/short ratio
NG citation	145/129	75/87	1.9/1.5
bVb	135/135	84/95	1.6/1.4
dVd	144/144	96/104	1.5/1.4
bVp	118/120	77/84	1.5/1.4
dVt	133/134	85/94	1.6/1.4
PF citation	113/159		
bVb		91/97	
dVd		102/107	
bVp		84/93	
dVt		96/110	
NYE citation	121/150	83/100	1.5/1.5
bVb	154/147	104/104	1.5/1.4
dVd	162/170	122/127	1.3/1.3
bVp	131/129	89/91	1.5/1.4
dVt	163/168	125/128	1.3/1.3

TABLE VII. Within-language discriminant analyses (% tokens correctly classified) for sentence materials. Male and female data are listed separately (M/F).

Contexts	F1/F2/F3 + duration	Re: citation
NG labial	93/93	83/86
NG alveolar	90/88	72/57
NG both	87/88	77/71
PF labial	81/94	63/87
PF alveolar	82/85	55/47
PF both	79/81	59/67
AE labial	91/96	82/87
AE alveolar	92/98	30/51
AE both	89/89	56/69

ances. Long and short vowels were somewhat shorter in bVp context. When L/S ratios were compared, there were only small differences across speech styles and phonetic contexts within gender groups, except for male citation utterances; on average, long/short differences were greater for male utterances.

Some of the same trends are seen in AE vowels, where long vowels were actually longer, on average, in bVb, dVd, and dVt utterances than in citation utterances. (For AE utterances, the final /d/ and /t/ were both produced as voiced flap [ɾ] and vowel duration differences were totally neutralized for all six speakers.) L/S ratios for vowels in sentence materials tended to be smaller than in citation materials. Comparisons of NG and AE sentence materials reveal that NG L/S ratios tended to be larger than AE ratios. In addition, there were differences across NG and AE for the low vowel pairs. NG [ɑː,a] were relatively more differentiated temporally in all four contexts (L/S ratios from 1.6 to 1.8) than were AE [ɑː,ʌ] (L/S ratios from 1.3 to 1.5).

PF vowels were, on average, considerably shorter in sentence materials than in citation utterances, especially in labial contexts. In sentence materials, PF vowels were closer in duration to NG and AE short vowels than to long vowels. This could account, in part, for the relatively large shifts in mid-syllable formant values reported in the previous section for PF vowels [o,u], in comparison with the "same" NG vowels [oː,uː].

3. Within-language discriminant analyses

Given the (dis)similarities across languages in the extent of contextual variation in average mid-syllable formant frequencies (Sec. B 1) and relative vocalic durations (Sec. B 2), the next step was to examine the acoustic differentiation of distributions of coarticulated vowels across languages and contexts. In all, 18 different discriminant analyses were computed for the sentence materials, combining various consonantal contexts: Corpora of vowels produced in labial (bVb, bVp) and alveolar (dVd, dVt) contexts were examined separately in (place) context-dependent analyses, and then vowels produced in all four contexts were entered into a single context-independent analysis. Again, corpora for male and female speakers were analyzed separately. Correct classification rates for each of these input sets were compared to determine the extent of acoustic overlap of vowel categories in each context and when labial and alveolar contexts were combined. Finally, to examine the extent of spectral and temporal *change* from canonical target values across gender/language groups, 18 additional analyses were computed in which the citation materials served as the training sets, while the sentence materials served as the test sets. These analyses allowed us to examine the extent to which distributions of particular vowels produced in sentence materials shifted so they were actually closer to canonical values for a *different* vowel category.

Table VII presents the overall rates of correct classification for the male and female sentence materials for the labial, alveolar, and combined corpora when the parameter weightings and centers-of-gravity were optimized for the sentence materials for each gender/language group (left). The right-hand columns give the overall classification rates for sentence materials (same test sets) when the citation materials (from the same language/gender group) served as the training sets. In all these analyses, F1/F2/F3 Bark values plus vocalic duration were included as input parameters. As in the analyses of citation materials reported above, the inclusion of duration as a parameter improved classification rates for NG and AE vowels, with almost no change for PF vowels.

When parameter weightings and category centers-of-gravity were established on the coarticulated vowel corpora themselves, correct classification rates, overall, were not drastically reduced from those reported for citation materials. For NG vowels, classification rates for vowels in labial context were comparable to those reported for citation materials (see Table I). There were more misclassifications of vowels in alveolar context, and when labial and alveolar contexts were combined (context-independent analyses). For the latter analyses, misclassifications were not evenly distributed across the 14 vowel categories (range=75% to 100% correct). There were confusions between high and mid vowels [iː/eː,uː/oː,yː/øː] and other height confusions [ʏ/œ,

ɛ/a], as well as a few confusions between front, unrounded and rounded vowels [ɪ/ʏ, ɛ/œ]. Front, rounded vowels were almost never misclassified as back vowels; back vowels were only rarely misclassified as front, rounded or unrounded vowels.

The analyses of PF vowels using parameter weightings and centers-of-gravity established for the coarticulated corpora themselves yielded generally lower correct classification rates than for NG (and AE) vowels and relative to PF citation materials. This reflects the greater contextual variation of PF vowels (see Fig. 4) and the greater shrinking of the vowel space. In looking at the context-independent analyses, the misclassifications were not evenly distributed across all nine PF vowels (range=58% to 100% correct). There were confusions between front, unrounded and rounded vowels [ɛ/y, e/y, ɛ/ø]; between back and front, rounded vowels [ɔ/ø, u/ø], as well as misclassifications in height [a/ɔ, y/ø, u/o].

The analyses of AE vowels using parameter weightings and centers-of-gravity established on the coarticulated vowels themselves yielded results more like those of NG than of PF. Again, overall rates of correct classification were about the same for vowels in both labial and alveolar context-dependent analyses as they were for citation utterances (Table I), and only slightly lower for context-independent analyses. For the latter, however, there were large ranges in classification rates across individual vowels (range=50% to 100% correct). Almost all misclassifications were height confusions among front vowels and among back vowels [iː/ɪ, ɛ/ɪ, ɑː/ɔː, ɔː/oː, ʌ/ʊ, oː/ʊ, uː/ʊ] with only a few confusions of back and front short vowels [ʊ/ɪ, ʌ/ɛ]. Thus, despite the striking change in location of back, rounded vowels in alveolar contexts, these vowel distributions did not overlap with front vowel categories. The pattern of height errors indicated that long/short duration differences differentiated front vowels well (except for [iː/ɪ]), but spectrally adjacent back vowel distributions overlapped both spectrally and temporally.

Turning to the question of the extent to which distributions of coarticulated vowels shifted from canonical values in the three languages, the discriminant analyses in which citation utterances were the training sets and sentence utterances were the test sets revealed striking differences across contexts and across languages. For all three analyses within each language, correct classification rates decreased; that is, for all languages, the acoustic structure of vowels differed as a function of speech style and phonetic context sufficiently that some tokens of particular vowels became closer to canonical centers-of-gravity for different vowels. However, distributions of vowels produced in labial contexts were, in general, more similar to canonical values, as would be expected given the relative independence of the tongue and lip gestures.

The overall classification rates of coarticulated vowels with canonical centers-of-gravity as the reference reflect cross-language differences in coarticulatory influences. Overall, the distributions of NG vowels showed the least variation from canonical values in both labial and alveolar contexts. In labial contexts, the decrease in correct classifications was relatively small overall and equally distributed across long and short vowels. Vowels in alveolar contexts were misclassified more often, with more short vowels misclassified (Long=77% correct, Short=48% correct). Classification errors included many misclassifications of back, rounded as front, rounded vowels [ɔ→ œ, ʊ→ ʏ]; front, rounded as front, unrounded vowels [ʏ→ ɪ, œ→ ɛ]; and the low vowel as a higher vowel [a → œ], reflecting the general shifts of back and low, short vowels toward more front and higher positions when they were coarticulated with alveolar consonants.

When PF coarticulated vowels were evaluated against canonical values, correct classification rates decreased relatively sharply, and showed a somewhat different pattern across gender groups. Misclassifications of male utterances were concentrated on three vowels [o, a, ø]: [o, a] were misclassified as higher vowels, [ø] as front, unrounded [ɛ], in both labial and alveolar contexts. Female utterances appeared not to vary as much from canonical values in labial contexts, but in alveolar contexts, misclassifications of [ɔ, a, u] reflected both fronting and raising patterns and [y] was also often misclassified as [e].

The cross-context analyses of AE vowels revealed yet another pattern. Distributions of vowels coarticulated in labial contexts did not vary much from canonical values, and misclassifications were distributed across vowel categories. However, in alveolar contexts, misclassification rates were very high, and were concentrated on the back vowels [uː, ʊ, oː, ʌ], which were classified as spectrally closer to front vowels [ɪ, ɛ] in these contexts. Thus, in alveolar contexts, these back vowels were fronted so drastically that the majority of tokens were classified as more similar to canonical front vowels. In addition, there were some misclassifications for [eː, ɪ, ɛ, ɔː] as higher vowels.

In summary, for all three languages, within-language discriminant analyses indicated that distributions of vowels produced in sentence materials overlapped in vowel space when vowels produced in both labial and alveolar contexts were included in the analyses. However, *context-dependent* analyses showed that, in each context, vowel distributions remained fairly well separated on the front-back dimension in all three languages even though context-specific fronting occurred. There was somewhat more overlap in the height dimension, especially for high and mid (long) vowels in all three languages. When distributions of coarticulated vowels were evaluated against canonical values, significant cross-language differences were revealed, as was expected from comparisons of mean data (Sec. B 1). NG vowels showed smaller shifts overall than the other two languages, except for short vowels in alveolar context. PF vowels in both contexts shifted more from canonical values than for NG vowels. Misclassification patterns indicated significant raising and fronting of [a, o] and fronting of [ø, u], especially in alveolar contexts. Finally, coarticulated AE vowels shifted relatively little in labial contexts, while in alveolar contexts [uː, ʊ, oː] showed extreme fronting, and [ʌ] both raising and fronting.

TABLE VIII. Cross-language spectral similarity of NG, PF, and NYE point vowels: Discriminant analyses (F1/F2/F3 Bark) for Sentence Materials. Male and female data are listed separately (M/F).

	Labial and alveolar		Labial		Alveolar	
NG vowels	Modal NYE V	% tokens classified	Modal NYE V	% tokens classified	Modal NYE V	% tokens classified
iː	iː	98/97	iː	92/94	iː	100/92
ɑː	ɑː	100/77	ɑː	100/92	ɑː	100/71
uː	uː	81/100	uː	79/100	ɔː oː	50/100
PF vowels						
i	iː	100/100	iː	96/100	iː	96/100
a	ʌ, ʊ	44/35	ʌ, ʊ	29/42	ʌ, ʊ	54/21
	ɪ, ɛ, æː	50/65	ɪ, ɛ, æː	71/58	ɪ, ɛ, æː	46/79
u	uː	96/98	uː	100/100	uː, ʊ, oː	33/79

TABLE IX. Cross-language spectral similarity (% tokens classified) of NG and PF front rounded vowels to AE back vowels: Discriminant analyses for Sentence Materials. Male and female data are listed separately (M/F).

NG vowels	NYE vowels	Labial and alveolar	Labial only	Alveolar only
yː	uː, ʊ	96/94	0/17	100/100
øː	uː, ʊ, oː	96/98	8/75	100/100
ʏ	uː, ʊ, oː	94/71	75/54	96/100
œ	oː, ʊ, ʌ, ɔː	83/94	92/88	92/100
PF vowels				
y	uː	25/21	0/0	33/17
ø	uː, ʊ	50/77	0/21	71/96

4. Cross-language discriminant analyses

Given the marked differences across languages in the pattern of contextual variation reported above, we expected that *cross-language* similarity patterns would differ from those for citation utterances, especially for the back and low vowels, and for the front, rounded vowels. For the context-independent comparisons, the male and female AE corpora that included all four contexts (48 tokens/vowels: 4 contexts $\times$ 3 speakers $\times$ 4 repetitions) served as the training sets, with the analogous NG and PF corpora as the test sets. For labial and alveolar (context-dependent) analyses, the two AE corpora for each gender group served as the training sets (24 tokens/vowel: 2 contexts $\times$ 3 speakers $\times$ 4 repetitions). For all analyses, input parameters included only F1/F2/F3 Bark values (as for citation materials); thus, these analyses establish the cross-language *spectral* similarity of distributions of coarticulated vowels produced in sentence materials.

Table VIII summarizes the cross-language classification patterns for NG and PF point vowels (column 1) for the context-independent analyses (columns 2–3), and the two context-dependent analyses (labial, columns 4–5; alveolar, columns 6–7). As these data show, almost all tokens of NG and PF [i] were classified as most similar to AE [i] in both context-independent and context-dependent analyses, reflecting the fact that this vowel was very similar across all three languages and varied little across contexts. Most of the NG and PF [u] tokens were classified as similar to AE [u] in the context-independent analyses, for which centers-of-gravity were computed over both back (labial) and fronted (alveolar) AE allophones. However, when only vowels produced in alveolar contexts were compared across languages, all of the NG [uː] tokens were more similar to AE [ɔː,oː], because of their lower F2 values relative to the fronted AE [uː]. The majority of male PF [u] tokens were also more similar to other AE vowels [ʊ,oː]. NG [ɑː] was classified predominantly as most similar to its AE counterpart, whereas PF [a] in all three comparisons was almost never classified as AE [ɑː]. Rather, both male and female tokens were classified as a variety of higher front and back AE vowels, reflecting the large contextual variation of this PF vowel in sentence materials (see Fig. 1).

Table IX presents the results of cross-language discriminant analyses for the coarticulated NG and PF front, rounded vowels. Classification rates for each of these vowels to AE back vowels, collapsed over height categories, are presented, with overall percentages of classifications as back vowels for the context-independent analyses (column 3) and the two context-dependent analyses (columns 4–5). Again, results for comparisons of male and female utterances are presented separately because of differences in outcomes across genders for some vowels. Looking first at the NG vowels, cross-language similarity patterns established by the context-independent analyses showed a markedly different pattern from citation utterances (Table III, columns 4–5), with nearly all tokens of all four vowels being classified as more similar to back than to front AE vowels. This was due to the shifts in centers-of-gravity for AE back vowel distributions that included the fronted allophones produced in alveolar contexts. The context-dependent analyses corroborated this pattern, with NG and AE vowels in labial contexts showing cross-language similarity patterns resembling those in citation utterances (see Table III). In contrast, almost all tokens of the four NG vowels produced in alveolar contexts were classified as spectrally similar to the (fronted) AE back vowels produced in this context.

Cross-language spectral similarity patterns for the coarticulated PF front, rounded vowels also changed relative to citation utterances. In the context-independent analysis, the majority of tokens of [y] were still classified as more similar to front than to back AE vowels; however, [ø] tokens straddled front and back AE vowel categories. Again, the context-dependent analyses showed that this PF vowel was similar to the (fronted) AE back vowels produced in alveolar context.

There were also changes in cross-language spectral similarity patterns for some of the mid-high, mid, and mid-low vowels, as shown in Table X. NG [ɪ,ɛ] were classified as their AE counterparts more consistently in context-independent analyses (shown in the table) and in both context-dependent analyses (not included in the table) than in citation materials (Table IV). However, a majority of tokens of NG [eː] and PF [ɛ] were still not classified as most similar to their counterparts in AE. NG back vowels [ʊ,oː], and PF back vowels [o,ɔ] were also not classified as most similar to their AE counterparts.

To summarize, cross-language similarity of distributions

TABLE X. Cross-language spectral similarity (% classified) of NG and PF mid-high, mid, and mid-low vowels to NYE vowels: Context-independent discriminant analyses (F1/F2/F3) for sentence materials. Male and Female results are listed separately (M/F).

NG vowels	Modal NYE vowel	% tokens classified	Other NYE vowels (>10%)
ɪ	ɪ	73/91	uː/
eː	iː	69/54	eː/eː
ɛ	ɛ	81/65	ɪ/ɪ
ʊ	uː	67/75	oː/oː
oː	uː	56/88	oː/oː
ɔ	ɔː	94/73	/oː
PF vowels			
e	iː	48/67	ɪ, eː/eː
ɛ	ɪ	48/52	eː, ɛ/eː, ɛ
o	ʊ/uː	44/92	ʊ, oː/
ɔ	ʌ/ʊ	27/33	ɛ, uː, ʊ, oː /oː, uː

of NG and PF coarticulated vowels to category centers-of-gravity for AE coarticulated vowels differed considerably from patterns found for citation materials. All four NG front, rounded vowels and PF [ø] were more similar to (fronted) back AE vowels, while most tokens of PF [y] were still more spectrally similar to front AE vowels. Front, unrounded NG [iː, ɪ, ɛ] and PF [i] were spectrally quite similar to their AE counterparts in context-independent analyses, and in some cases, better matches than in citation materials. In contrast, the majority of tokens of low and back NG and PF vowels, except for NG [ɑː, ɔ], were more similar to centers-of-gravity for other AE vowels than their transcriptional counterparts, due to marked differences in the patterns of fronting and raising of these vowels across languages.

IV. DISCUSSION

This study compared within- and cross-language (dis)-similarities in the spectral and temporal structure of vowels in three languages as a function of speech style (citation vs sentence utterances) and phonetic context (labial vs alveolar preceding/following consonants) in sentence utterances. Mid-syllable formant frequencies in citation utterances established canonical acoustic targets for vowels in the three languages, and could provide an acoustic basis for interpreting results of previous perceptual studies using these sorts of materials. Then spectral relationships among distributions of vowels produced in sentence materials were compared with canonical values to document (dis)similarities in patterns of coarticulatory variation across phonetic contexts within and across languages. The contribution of vocalic duration to the acoustic differentiation of vowels in the three languages was also assessed. The results revealed that contextual variation differed markedly across languages, and therefore, cross-language spectral similarities of vowel categories varied as a function of speech style and consonantal context. In this section, we review the results for citation and sentence materials and discuss how they address hypotheses about the role of vowel inventory make-up and size in determining the nature of the language-specific constraints on coarticulatory undershoot. We point out some limitations of the present study and how they may be addressed in future research. Lastly, we explore the implications of these findings for explaining/predicting patterns of perceived cross-language similarity and discrimination of NG and PF vowels by AE listeners.

A. Within- and cross-language variation of vowels in citation utterances

The results of within-language discriminant analyses showed that vowels in all three languages were well differentiated by mid-syllable formant frequencies and vocalic duration in citation utterances, despite (minor) speaker differences within gender groups (Table I). As expected, vocalic duration was most important in the differentiation of spectrally-similar NG vowels, while duration did not contribute at all to the differentiation of PF vowels. There were only minor differences across gender groups in the relative locations of vowels in a vowel space defined by F1/F2/F3 Bark values (see Fig. 2 and Appendices). For both gender groups, F2-F1 and F3-F2 (Bark) ratios differentiated vowels along the front-back dimension in all three languages: front vowels had large F2-F1 Bark differences (>5 Barks) and small F3-F2 Bark differences (<5 Barks) (except for AE [æː]), while back vowels had the reverse relationship, small F2-F1 differences, and large F3-F2 differences (cf., Syrdal and Gopal, 1986). Front, rounded vowels in both NG and PF tended to have somewhat smaller F2-F1 and larger F3-F2 Bark difference values relative to front, unrounded vowels, but were still more "front" than "back" in terms of the 5-Bark difference criteria given above (except for NG [œ]). F1 values were related to vowel height, although so-called mid-high, short vowels in NG and AE were acoustically equal to or a bit lower (higher F1 values) than so-called mid, long vowels (see Appendices). Thus, phonetically, all three languages contrasted four levels of height spectrally, with further differentiation by duration of AE and NG vowels in the mid to mid-high region.

For cross-language comparisons, only mid-syllable formant values were used to assess similarity relationships by linear discriminant analysis, since absolute and relative durations of "similar" vowels differed across languages and previous research had shown that AE listeners perceptually assimilated NG vowels more on the basis of spectral similarity than temporal similarity (Strange *et al.*, 2005). Results indicated that the point vowels, which defined the overall dimensions of the vowel spaces for all three languages, were roughly equivalent in these citation utterances (Table II): high, front [i] was almost identical in location and distribution across languages; NG and AE [ɑː] were also very similar, while PF [a] straddled AE [æː, ɑː]; NG and PF [u], while most similar to AE [uː], were clearly more back in citation context. PF front, rounded vowels were more similar to front than to back AE vowels, while NG front, rounded vowels straddled front and back AE categories, except for [yː], which was closer to front AE vowels in this context (Table III). Distributions of other NG and PF vowels, usually transcribed as the "same" or "similar" to AE categories, were

found to vary greatly in their spectral similarity to their AE counterparts (Table IV). Most, but not all, NG and PF mid-high, mid, and mid-low vowels appeared to be located more toward the high extreme in vowel space relative to AE vowels.

The acoustic dissimilarities for "similar" AE and PF vowels might account for discrimination difficulties by AE listeners of French vowel contrasts that also occur in their native language (Gottfried, 1984). However, the cross-language spectral similarities of front, rounded vowels to front AE vowels do not account well for why AE listeners perceptually assimilate NG front, rounded vowels to back native categories even in citation utterances (Strange *et al.*, 2004a, 2004b), nor why AE listeners tend to confuse front vs back, rounded French vowels more than they do front, unrounded *vs* rounded vowels (Gottfried, 1984; Polka, 1995). However, cross-language spectral similarities of these vowels in sentence materials may account for these perceptual assimilation and discrimination patterns.

B. Within- and cross-language variation of vowels in sentence materials

Distributions of coarticulated vowels, produced in multisyllabic nonsense words at a rate more closely approximating continuous speech, were evaluated with respect to how spectral and temporal parameters varied from canonical target values. Duration differences for NG and AE long/short vowels were somewhat reduced in sentence materials, relative to citation materials, especially for NG male utterances (Table VI). This replicated earlier studies of NG and AE male productions (Strange *et al.*, 2004a, 2005) and extended the findings to female speakers and to a more homogeneous dialect group of AE speakers. In general, L/S duration ratios were larger for NG than for AE speakers, as was expected, although average differences between language groups varied with consonantal context and with gender. Discriminant analyses of the sentence materials, with and without duration as an input parameter, indicated that for both NG and AE, duration contributed substantially to the acoustic differentiation of spectrally-adjacent vowels. In contrast, PF vowels were not differentiated by relative duration differences; absolute durations were substantially reduced in sentence materials relative to citation utterances, especially for female utterances. In sentence materials, PF vowels were more similar in duration to NG and AE short vowels than to long vowels. Thus, in interpreting cross-language differences in spectral target undershoot, PF vowels were compared with NG and AE short vowels.

In all three languages, the front vowels [i,e] (and [y] in NG and PF) varied least across speech styles and phonetic contexts, providing a reference for comparisons of other vowels within and across languages. It is hypothesized that the relative stability in the acoustic structure of these vowels may be due to the fact that the sides of the tongue are in contact with the teeth and alveolar ridge for their production. The patterns of coarticulatory variation of mid-syllable formant frequencies for other vowels differed markedly across languages (Figs. 3 and 4). In general, there was greater coarticulatory undershoot in alveolar contexts than in labial contexts for other vowels, presumably due to the influence of tongue-tip gestures for the alveolar consonants on tongue body gestures for the vowels. In labial contexts, there was some raising of the low, short vowel [a] in NG, with even greater raising of this vowel in PF. These shifts might be attributable to less lowering of the jaw in the more rapidly produced sentence utterances because of the necessity for the lips to fully close for labial consonants.

In alveolar contexts, NG and PF [a] were even further raised and also fronted in both NG and PF utterances; other short NG vowels [ɛ, ʏ, œ, ɔ, ʊ] and PF [ɔ, o, u] also shifted up and/or forward (Table V), while long NG low and back vowels shifted relatively little. Thus, in alveolar context, the PF vowel space decreased markedly in size but relative positions of vowels remained similar. In NG, the relative positions of long and short NG vowels changed due to greater shifts for the latter (Figs. 3 and 4). The smaller shifts of long vowels in NG would be predicted if undershoot were a function of syllable duration (i.e., the extent of temporal overlap of consonant and vowel commands at syllable midpoint). However, the pattern of coarticulatory change for AE long vowels was very different from the other two languages, with allophonic fronting of both long and short, back vowels [uː, ʊ, oː], and relatively small shifts for mid-low and low, short and long, back vowels. This changed the shape of the vowel space, with the back, high to mid, rounded vowels becoming more like front vowels (large F2-F1; small F3-F2).

These different patterns of contextual change for the same vowels in the same contexts cannot be accounted for by a simple model of target undershoot, but rather support the claim that there are language-specific constraints on coarticulation that serve to maintain acoustic (and perceptual) differentiation of vowels in each language (Diehl and Lindblom, 2004; Lindblom, 1990). Because AE does not contrast front and back, rounded vowels, the back vowels are "free" to vary more in alveolar context. In NG and PF, where back and front, rounded vowels are contrasted, the back vowels move forward relatively less, and the front, rounded vowels (except for the very front [y]) move forward in conjunction with the back vowels to maintain distinctiveness. In PF, with fewer (and more fronted) front, rounded vowels than in NG, the back vowels are fronted more in alveolar contexts (Table V). Thus, context-*dependent* discriminant analyses showed that, even in alveolar contexts, vowel distributions remained fairly well separated on the front-back dimension in all three languages (Table VII).

Context-dependent discriminant analyses suggested that there was greater spectral overlap on the height dimension for coarticulated vowels, especially in NG and AE. In NG, relative duration differences differentiated so-called mid-high, (short) vowels and mid, (long) vowels, despite the somewhat smaller L/S ratios than in citation utterances. However, high and mid (long) vowels were sometimes misclassified in both contexts. In AE, relative duration differences helped to differentiate spectrally similar front vowels, but not back vowels. Coarticulated PF vowels (as well as citation utterances) were not acoustically differentiated by

duration. Thus, in both contexts, there was some acoustic ambiguity of adjacent height categories across speakers within gender groups.

These language-specific patterns of coarticulatory change in mid-syllable formant frequencies cannot be fully accounted for by the hypothesized correlation between vowel inventory size and amount of articulatory/acoustic undershoot of vowel targets (Bohn, 2004; Steinlen, 2005). In general, there were greater coarticulatory shifts (relative to canonical targets) in PF (9 vowels) than in NG (14 vowels), supporting the hypothesis; however, considerations of vocalic duration must be included in this analysis. Differences in the extent of target undershoot between NG short vowels and PF vowels (which were closer to NG short vowels in vocalic duration) were much less striking than between NG long and PF short versions of the "same" vowels [o,u,ø] (Table V). In comparing AE (11 vowels) with the other two languages, the allophonic fronting of high to mid, back vowels should not be considered in correlations of inventory size with overall spectral change across languages. We suggest that the *phonological function* of vowel length (and other features such as lip rounding), as well as the overall size of the vowel inventory, must be considered in attempts to predict/explain differences across languages in the contextual variation of vowels. Further studies comparing vowels in Russian, Japanese, and Spanish are underway in our laboratory to extend this investigation to languages with relatively small vowel inventories (5–6 spectral categories) that differ in the phonological function of vowel length (e.g., Law II *et al.*, 2006). Research on South German, which is reported to distinguish tense/lax vowel pairs almost exclusively by vocalic duration (i.e., contrasts only three heights) would also provide valuable insights about language-specific coarticulatory constraints, as they relate to inventory size and makeup.

C. Limitations of the study

Several limitations of the study should be addressed in future research. While the results for male and female corpora showed fairly similar patterns of within- and cross-language (dis)similarities, some gender differences were noted in the sentence materials. However, the number of speakers of each gender within each language was small; thus, differences may not be reliable. No inferential statistical comparisons of gender differences in shifts in mid-syllable formant frequencies or durations were computed, given the low power of such tests with these small sample sizes.[6] To follow up on whether there are significant differences in the phonetic realization of vowels across genders, more speakers of each gender must be tested.

Conclusions about within- and cross-language (dis)similarities were based on quantitative analyses of distributions of utterances, summing over multiple tokens contributed by a small number of speakers selected for their dialect homogeneity (see Hillenbrand *et al.*, 2001). Generalization to speakers of other dialects of these languages should not be made. For example, conclusions about the relative distinctiveness of spectral and temporal parameters in Parisian French and North German do not apply to Swiss French (Miller *et al.*, 2000) and Southern dialects of German. While this narrows the scope of the conclusions, it is our opinion that dialect-specificity is necessary when investigating the phonetic realization of vowel categories unless it can be established, *a priori*, that vowels do not differ phonetically across the dialects represented in the sample.

In these analyses, only mid-syllable formant frequencies and vocalic durations were considered. Thus, other important acoustic information for specifying vowel gestures was not included (Strange, 1989). Future studies in which formant trajectories throughout the entire vocalic nuclei are tracked may uncover important differences in coarticulatory patterns across languages. For instance, NG speakers might back the place-of-constriction of alveolar consonants in the context of back vowels more than PF speakers do; these contextual variations in consonant production would be indicated by F2/F3 formant onset and offset frequencies.[7] Second, in addition to vocalic duration differences in short (lax) and long (tense) vowels in NG and AE, these vowels are also distinguished by differences in the temporal location of F1 maximum (Stack *et al.*, 2006), and the shape of F1 temporal trajectories (Huang, 1985); tense vowels have relative symmetrical (and rapid) onglides and offglides, while lax vowels show asymmetrical temporal trajectories. For AE lax vowels (including [æː]), F1 onglides are shorter than offglides (Peterson and Lehiste, 1960); for NG lax vowels, F1 onglides are longer than offglides (Strange and Bohn, 1998).[8] To our knowledge, there are no data on the temporal properties of F1 trajectories for PF vowels. The perceptual relevance of these (dis)similarities across languages in the dynamic spectral information should be explored.

D. Implications for studies of perceived phonetic similarity and discrimination of vowels

A major motivation for comparing the acoustic structure of NG, PF, and AE vowels in this study arose from previous research investigating the difficulties that native speakers of AE have in discriminating French (Gottfried, 1984; Levy, 2004; Levy and Strange, in press) and German vowels (Polka, 1995) and the patterns of perceived similarity of NG and PF vowels to AE categories (Strange, 2007; Strange *et al.*, 2004a, 2004b; 2005). The comparisons of the patterns of acoustic variation of vowels in the three languages reported here provide some insights into the patterns of perception by naïve listeners and AE speaking L2 learners of German and French reported in the literature.

First, let us consider front, rounded vowels as they relate to AE categories. While both PF and (long) NG vowels in citation utterances are more similar acoustically to front than to back AE vowel categories, AE listeners routinely categorize them as more perceptually similar to back native categories and have more difficulty discriminating them from back, rounded vowels than from front, unrounded vowels. That is, acoustic similarity of citation utterances does *not* predict perceived similarity nor discrimination difficulty in categorial (name identity) tasks. This can be accounted for by considering the distributional characteristics of native vowel categories in continuous speech contexts with which the AE

listeners have had experience. As the data reported here show, AE front vowels vary little across prosodic and phonetic contexts, whereas AE back vowels vary extensively on the front-back dimension. Thus, NG and PF front, rounded vowels are quite deviant from any exemplars of front vowels in AE, while they are very similar to some allophones of back AE vowels. Thus, in judging perceived similarity and in discriminating non-native contrasts, front and back, rounded vowels are phonetic variants of the *same* phonological category in AE, whereas front, unrounded and rounded vowels are phonetic variants of *different* phonological categories in AE. The results of perceptual assimilation and discrimination studies support the conclusion that this experience with phonetic variation in phonological categories determines AE listeners' performance in perceptual assimilation and categorial discrimination tasks (cf., Best, 1995; Best and Tyler, 2007), and predicts L2 learners' perception (and production) problems in differentiating these categories in German and French.

Cross-language differences in how distributions of NG, PF, and AE vowels are arrayed in the height dimension, and how duration is or is not used to differentiate spectrally-adjacent categories may also account for why some vowel pairs in German and French that are also contrasted in AE are nevertheless problematic. Confusions between high and mid NG and PF vowels (e.g., [i/e, y/ø, u/o]) that do not differ in duration can be expected. Other height contrasts (e.g., [e/ɛ] problems reported by Gottfried 1984) may be more or less problematic depending upon the context in which they are produced and presented.

In a forthcoming paper, results of perceptual assimilation tests of PF and NG vowels by inexperienced AE listeners are presented. In two separate studies, tests of vowels produced/presented in citation disyllables and in sentences with varying consonant contexts revealed differences across languages and speaking styles in the native AE categories to which the PF and NG vowels were assimilated and in the judged category goodness of the vowels. Some, but not all, of these variations in perceived cross-language similarity could be accounted for by contextual differences in spectral similarity reported in the present study.

In general, the results reported here indicate that predictions about perceptual difficulties with vowels by L2 learners, based on acoustic comparisons and perceptual studies using citation utterances may not be generalizable to continuous speech contexts. If our goal is to develop theories of non-native and L2 perception and production in "real world" communication situations, future research on vowels must employ speech materials that more closely resemble the speech input of the language learning environment. The sentence materials used here are still instances of "lab speech," and no doubt differ markedly from truly spontaneous speech in a conversational setting. However, these read sentence materials may approach the speech style that is used in a formal language-learning environment. The phonetic variation of vowels in sentence materials reported here help to explain patterns of difficulty found in the laboratory, and it is hoped, in the language learning environment of adult learners of non-native languages.

ACKNOWLEDGMENTS

The research reported here was funded by a grant to the first author (NIH NIDCD-RO1-00323). Thanks to the following people for help on the data analysis and the manuscript: James J. Jenkins, Franzo F. Law II; Yana Gilichinskaya. Portions of this research were reported in a poster at ASA (Strange *et al.*, 2002) and in a chapter by Strange (2007).

APPENDIX A: NORTH GERMAN CITATION MATERIALS

Mean F3-F2 and F2-F1 Bark differences, F1 (Barks), and syllable durations (ms); male (M) and female (F) data are listed separately.

NG	F3-F2 (Barks)		F2-F1 (Barks)		F1 (Barks)		Duration (ms)	
vowels	M	F	M	F	M	F	M	F
iː	2.3	1.6	10.9	11.9	2.3	3.1	118	112
eː	2.0	0.9	10.1	10.3	3.2	4.5	150	137
ɪ	2.4	1.4	8.4	9.6	3.6	4.6	64	80
ɛ	2.7	2.0	6.2	6.8	5.5	6.6	79	92
yː	1.4	2.0	9.2	9.3	2.6	3.1	132	122
øː	2.6	3.1	7.2	6.9	3.5	4.4	151	140
ʏ	3.2	3.2	6.7	6.8	3.6	4.5	69	85
œ	3.6	3.7	4.9	4.8	5.3	6.4	84	91
uː	8.0	8.0	3.3	3.0	2.7	3.7	136	114
oː	8.6	8.1	2.3	2.4	3.7	4.5	152	128
ʊ	7.1	7.1	3.3	3.3	3.8	4.6	70	80
ɔ	6.1	5.6	2.7	2.6	5.7	6.8	82	92
ɑː	4.9	4.1	2.8	2.5	6.8	8.5	173	147
a	4.9	4.2	3.1	2.7	6.4	8.3	7.3	90

APPENDIX B: PARISIAN FRENCH CITATION MATERIALS

Mean F3-F2 and F2-F1 Bark differences, F1 (Barks), and syllable durations (ms); male (M) and female (F) data are listed separately.

PF	F3-F2 (Barks)		F2-F1 (Barks)		F1 (Barks)		Duration (ms)	
vowels	M	F	M	F	M	F	M	F
i	2.7	2.3	10.7	12.2	2.6	2.4	108	143
e	2.1	1.5	9.6	10.7	3.4	3.7	87	107
ɛ	2.3	2.1	8.0	7.9	4.5	5.8	106	172
y	2.0	0.9	9.8	11.1	2.7	2.6	113	155
ø	3.3	3.3	7.0	7.3	3.7	4.5	136	173
u	6.6	8.0	4.5	4.2	2.7	2.9	106	167
o	7.2	8.1	3.2	3.6	3.8	3.8	126	183
ɔ	5.0	5.7	3.9	3.9	5.2	5.8	113	163
a	4.1	3.8	3.6	3.3	6.7	8.1	125	168

APPENDIX C: AMERICAN ENGLISH CITATION MATERIALS

Mean F3-F2 and F2-F1 Bark differences, F1 (Barks), and syllable durations (ms); male (M) and female (F) data are listed separately.

AE	F3-F2 (Barks)		F2-F1 (Barks)		F1 (Barks)		Duration (ms)	
vowels	M	F	M	F	M	F	M	F
iː	2.0	1.0	10.9	11.7	2.7	3.2	99	118
eː	1.6	0.9	9.4	10.3	3.9	4.5	125	163
ɪ	2.1	1.9	7.6	7.8	4.6	5.5	76	98
ɛ	3.0	2.5	5.2	5.4	5.8	7.3	85	101
æː	3.4	3.0	3.8	3.8	7.0	8.5	126	153
uː	6.3	6.9	4.2	4.2	3.2	4.0	98	131
oː	6.1	6.3	3.5	3.5	4.4	5.4	120	152
ʊ	5.2	5.1	4.5	4.1	4.5	6.0	87	102
ːɔ	5.9	5.5	2.7	2.4	5.8	7.2	153	172
ɑː	4.8	4.6	2.9	2.4	6.7.	8.3	128	163
ɑː	4.8	4.6	3.6	3.2	6.0	7.4	82	102

[1]French nasal vowels /ɛ̃,œ̃,ɑ̃,ɔ̃/ were not considered in this study because their acoustic structure varies as a function of coupling with the nasal tract. Nasal vowels are perceived as quite distinct from oral vowels because of these nasal resonances.

[2]In addition to these conditions, speakers also recorded sentence materials (dVt context) in which sentence prosody (target nonsense word in narrow focus vs post-focus position) and the speaking rate (normal vs rapid rate) varied. Results of acoustic analyses of these materials will be presented in a separate paper.

[3]These were not distributed evenly over vowels: the largest problems were for F3 in NG [uː, ʊ] and PF [u], which had very low amplitude, and F2 and F3 for PF [y].

[4]For AE vowels, formant trajectories throughout the middle half of the syllables (25%–50%–75% Bark values) for individual speakers were also plotted to determine whether there were systematic patterns of VISC or diphthongization common to all speakers. Results yielded large inter-speaker differences in both the direction and extent of formant movement for several vowels, including the so-called diphthongized [eː,oː]. Therefore, in this and all subsequent analyses, trajectory information was not included in within-language or cross-language comparisons.

[5]Discriminant analyses using F1, F2-F1, F3-F2 Bark values were also computed. However, since the pattern of results was very similar to analyses with F1/F2/F3 Bark values as independent parameters, only the latter are reported here.

[6]With respect to tests of mean differences, it is our opinion that comparison of average values is less informative than are discriminative analyses, regardless of sample sizes. Significant differences in central tendencies only establish the reliability of those differences, not their size or perceptual relevance. The discriminative analyses establish relationships among distributions of tokens, and are thus more akin to estimates of d prime (d') or effect size.

[7]The first author wishes to thank Kenneth N. Stevens for suggesting that this may be a factor in cross-language differences in vowel coarticulation.

[8]These spectro-temporal patterns of formant change are not the same as the VISC reported by Nearey and colleagues which occur in Canadian and American CVC syllables and isolated vowels (Nearey and Assmann, 1986). For the AE vowels analyzed here, formant change over the middle half of the syllables, even in citation utterances, was not systematic across speakers.

Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Timonium, MD), pp. 171–204.

Best, C. T., and Tyler, M. D. (**2007**). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language experience in second language speech learning: In honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro (John Benjamins, Amsterdam), pp. 13–34.

Best, C. T., Hallé, P. A., Bohn, O.-S., and Faber, A. (**2003**). "Cross-language perception of nonnative vowels: Phonological and phonetic effects of listeners' native languages," in *Proceedings of the 15th International Congress of Phonetic Sciences*, edited by M. J. Sale, D. Rescasens, and J. Romero (Causal Productions, Barcelona), pp. 2889–2892.

Bohn, O.- S. (**2004**). "How to organize a fairly large vowel inventory: the vowels of Fering (North Frisian)," J. Int. Phonetic Assoc. **34**, 161–173.

Diehl, R. L., and Lindblom, B. (**2004**). "Explaining the structure of feature and phoneme inventories: The role of auditory distinctiveness," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, New York), pp. 101–l62.

Flege, J. E. (**1987**). "The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification," J. Phonetics **15**, 47–65.

Flege, J. E. (**1995**). "Second language speech learning: theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Timonium, MD), pp. 233–277.

Fourakis, M. (**1991**). "Tempo, stress, and vowel reduction in American English," J. Acoust. Soc. Am. **90**, 1816–1827.

Gottfried, T. L. (**1984**). "Effects of consonant context on the perception of French vowels," J. Phonetics **12**, 91–114.

Gottfried, T. L., and Beddor, P. S. (**1988**). "Perception of temporal and spectral information in French vowels," Lang Speech **32**, 57–75.

Hay, J. F., Sato, M., Coren, A. E., Moran, C. L., and Diehl, R. L. (**2006**). "Enhanced contrast for vowels in utterance focus: A cross-language study," J. Acoust. Soc. Am. **119**, 3022–3033.

Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (**2001**). "Effects of consonant environment on vowel formant patterns," J. Acoust. Soc. Am. **109**, 748–763.

Huang, C. (**1985**). "Perceptual correlates of the tense/lax distinction in General American English," Master's thesis, MIT Cambridge, MA.

Klecka, W. R. (**1980**). *Discriminant Analysis* (Sage, Newbury Park, CA).

Law II, F. F., Gilichinskaya, Y. D., Ito, K., Hisagi, M., Berkowitz, S., Sperbeck, M. N., Monteleone, M., and Strange, W. (**2006**). "Temporal and spectral variability of vowels within and across languages with small vowel inventories: Russian, Japanese, and Spanish," J. Acoust. Soc. Am. **20**, Pt. 2, 3296.

Levy, E. S. (**2004**). "Effects of language experience and consonantal context on perception of French front rounded vowels by adult American English learners of French." Ph.D. dissertation, City University of New York-Graduate School and University Center.

Levy, E. S., and Strange, W., "Perception of French vowels by American English adults with and without French language experience," J. Phonetics.

Lindblom, B. (**1963**). "Spectrographic study of vowel reduction," J. Acoust. Soc. Am. **35**, 1773–1781.

Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 403–439.

Miller, J. L., and Grosjean, F. (**1997**). "Dialect effects in vowel perception: the role of temporal information in French," Lang Speech **40**, 277–288.

Miller, J. L., Mondini, M., Grosjean, F., and Dommergues, J.-Y. (**2000**). "Dialect differences in the temporal characteristics of vowels: a comparison of standard (Parisian) and Swiss French," J. Acoust. Soc. Am. **108**, 2507.

Moon, S.-J., and Lindblom, B. (**1994**). "Interaction between duration, context, and speaking style in English stressed vowels," J. Acoust. Soc. Am. **96**, 40–55.

Nearey, T. M., and Assmann, P. F. (**1986**). "Modeling the role of inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**, 1297–1307.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllabic nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.

Polka, L. (**1995**). "Linguistic influences in adult perception of non-native vowel contrasts," J. Acoust. Soc. Am. **97**, 1286–1296.

Stack, J. W., Strange, W., Jenkins, J. J., Clarke III, W. D., and Trent, S. A. (**2006**). "Perceptual invariance of coarticulated vowels over variations in speaking rate," J. Acoust. Soc. Am. **119**, 2394–2405.

Steinlen, A. K. (**2005**). "*The influence of consonants on native and nonnative vowel production: A cross-linguistic study*" (Gunter Narr, Tübingen).

Stevens, K. N., and House, A. S. (**1963**). "Perturbation of vowel articulations by consonantal context: An acoustical study," J. Speech Hear. Res. **6**, 111–128.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., and Öhman, S. (**1969**). "Cross-language study of vowel perception," Lang Speech **12**, 1–23.

Strange, W. (**1989**). "Dynamic specification of coarticulated vowels spoken in sentence context," J. Acoust. Soc. Am. **85**, 2135–2153.

Strange, W. (**2007**). "Cross-language phonetic similarity of vowels: theoretical and methodological issues," in *Language experience in second language speech learning: In honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro (John Benjamins, Amsterdam), pp. 35–55.

Strange, W., and Bohn, O.-S. (**1998**). "Dynamic specification of coarticulated German vowels: perceptual and acoustic studies," J. Acoust. Soc. Am. **104**, 488–504.

Strange, W., Bohn, O.-S., Trent, S. A., and Nishi, K. (**2004a**). "Acoustic and perceptual similarity of North German and American English vowels," J. Acoust. Soc. Am. **115**, 1791–1807.

Strange, W., Bohn, O.-S., Nishi, K., and Trent, S. A. (**2005**). "Contextual variation in the acoustic and perceptual similarity of North German and American English vowels," J. Acoust. Soc. Am. **118**, 1751–1762.

Strange, W., Levy, E. S., and Lehnhoff, Jr., R. J. (**2004b**). "Perceptual assimilation of French and German vowels by American English monolinguals: Acoustic similarity does not predict perceptual similarity," J. Acoust. Soc. Am. **115**, 2606.

Strange, W., Weber, A., Levy, E. S., Shafiro, V., and Nishi, K. (**2002**). "Within- and across-language acoustic variability of vowels spoken in different phonetic and prosodic contexts: American English, North German, and Parisian French," J. Acoust. Soc. Am. **112**, 2384.

Syrdal, A. K., and Gopal, H. S. (**1986**). "A perceptual model of vowel recognition based on the auditory representation of American English vowels," J. Acoust. Soc. Am. **79**, 1086–1100.

van Son, R. J. J. H., and Pols, L. D. W. (**1992**). "Formant frequencies of Dutch vowels in a text, read at normal and fast rate," J. Acoust. Soc. Am. **88**, 1683–1693.

Intelligibility of speech in noise at high presentation levels: Effects of hearing loss and frequency region[a)]

Van Summers[b)] and Mary T. Cord
Army Audiology and Speech Center, Walter Reed Army Medical Center, Washington, DC 20307-5001

(Received 7 March 2006; revised 7 December 2006; accepted 29 May 2007)

These experiments examined how high presentation levels influence speech recognition for high- and low-frequency stimuli in noise. Normally hearing (NH) and hearing-impaired (HI) listeners were tested. In Experiment 1, high- and low-frequency bandwidths yielding 70%-correct word recognition in quiet were determined at levels associated with broadband speech at 75 dB SPL. In Experiment 2, broadband and band-limited sentences (based on passbands measured in Experiment 1) were presented at this level in speech-shaped noise filtered to the same frequency bandwidths as targets. Noise levels were adjusted to produce ~30%-correct word recognition. Frequency bandwidths and signal-to-noise ratios supporting criterion performance in Experiment 2 were tested at 75, 87.5, and 100 dB SPL in Experiment 3. Performance tended to decrease as levels increased. For NH listeners, this "rollover" effect was greater for high-frequency and broadband materials than for low-frequency stimuli. For HI listeners, the 75- to 87.5-dB increase improved signal audibility for high-frequency stimuli and rollover was not observed. However, the 87.5- to 100-dB increase produced qualitatively similar results for both groups: scores decreased most for high-frequency stimuli and least for low-frequency materials. Predictions of speech intelligibility by quantitative methods such as the Speech Intelligibility Index may be improved if rollover effects are modeled as frequency dependent. [DOI: 10.1121/1.2751251]

PACS number(s): 43.71.Es, 43.71.Ky, 43.71.Gv [MSS] Pages: 1130–1137

I. INTRODUCTION

Speech recognition performance by listeners with normal hearing (NH) deteriorates as presentation levels increase above moderate, conversational levels. This general pattern, referred to as rollover, has been demonstrated in numerous studies dating back to the 1920s (Fletcher, 1922; French and Steinberg, 1947; Kryter, 1946; Fletcher and Galt, 1950; Pickett and Pollack, 1958; Pollack and Pickett, 1958; Speaks *et al.*, 1967; Chung and Mack, 1979; Dirks *et al.*, 1982; Hagerman, 1982; Beattie, 1989; Goshorn and Studebaker, 1994; Studebaker *et al.*, 1999; Studebaker and Sherbecoe, 2002; Molis and Summers, 2003; Summers and Molis, 2004). The Speech Intelligibility Index (SII) (ANSI, 1997), a widely used method of predicting speech intelligibility, takes rollover into account by including a level distortion factor (LDF) as part of its calculation. The LDF systematically decreases predicted scores as critical band levels are increased above those associated with broadband speech at about 73 dB SPL.

Rollover effects have generally been viewed as similar in magnitude across frequency. Consistent with this view, the LDF adjustment in the SII is applied uniformly across the entire frequency range. However, recent data reported by Molis and Summers (2003) suggest that rollover effects are greater in mid-to-high-frequency regions beginning around 2000 Hz than in frequency regions below 1200 Hz. In that study, when sentence materials were high-pass filtered so that only high-frequency speech regions were audible, key-word recognition decreased by more than 25% from about 76% correct at presentation levels of 75 and 85 dB SPL, to about 50% correct at 105 dB SPL. Rollover was also present for low-pass filtered sentences, but increases in level from 75 to 105 dB SPL led to only a 7% reduction in key-word recognition (from about 70% to 63% correct).

Greater rollover effects for high- than low-frequency speech cues might be expected based on differences in cochlear processing of high and low frequencies by mammals (Cooper and Rhode, 1996; Rhode and Cooper, 1996; Ruggero *et al.*, 1997). At low signal levels (below about 40 dB SPL) magnitude responses of cochlear nerve fibers tuned to high frequencies (above about 1000 Hz) show high sensitivity over a restricted range of frequencies (the "tip" region of a given fiber), and greatly reduced sensitivity for other frequencies (falling in the "tail" region of the fiber). The large response in tip regions is believed to reflect the influence of active processes associated with the cochlear outer hair cells (OHCs) which amplify response to frequencies in tip regions and have little influence on responses to tail frequencies. The gain provided by this "cochlear amplifier" diminishes as levels increase and, as a result, response levels grow compressively with input level in tip regions (e.g., 0.2-dB response increase per dB increase in input) but show linear growth (dB per dB) in tail regions. With the loss of this frequency-specific amplification at high levels, frequency tuning becomes much less selective for high-frequency fibers, changing from a narrow passband response to a more broadly tuned low-pass pattern. Although active processes also influence cochlear processing in low-frequency regions, the added gain is much less frequency-specific. As a result, the

[a)]Portions of this research were presented at the 148th Meeting of the Acoustical Society of America, Vancouver, Canada, May, 2005.
[b)]Electronic mail: walter.summers@na.amedd.army.mil

TABLE I. Age and audiometric thresholds of subjects (test ear, dB HL). Available information on most likely etiology is also included for HI listeners.

	Age	250	500	1000	1500	2000	3000	4000	6000	Etiology
HI1	69	25	30	35	40	35	35	40	40	Noise induced
HI2	71	40	40	35	35	45	45	40	45	Prebycusis?[a b c]
HI3	75	30	30	35	40	35	30	40	50	Genetic or presbycusis[c d]
HI4	63	35	40	45	45	45	45	40	30	Genetic or presbycusis[c d]
HI5	79	55	45	45	40	35	40	40	40	Presbycusis?[a b c]
HI6	78	20	30	35	40	45	40	45	55	Presbycusis?[a b c]
NH1	72	15	10	10	0	10	15	20	30	
NH2	78	20	20	20	5	20	20	15	20	
NH3	72	20	20	20	15	15	10	10	20	
NH4	73	15	15	20	20	15	15	10	15	
NH5	71	10	5	10	5	20	15	20	15	
NH6	69	10	10	10	5	15	15	10	5	

[a]No family history of hearing loss.
[b]No history of exposure to excessive noise.
[c]Loss first noted after age 50.
[d]Mother had hearing loss.

responses of fibers tuned to low frequencies do not show a tip-tail pattern, and frequency selectivity is more nearly constant at low and high input levels (Cheatham and Dallos, 2001). Recent psychophysical results for NH listeners are consistent with these physiological data. In "temporal masking curve" results believed to reflect basilar membrane response growth with level, frequency selectivity was reduced at high input levels for high frequencies but was essentially level independent at low frequencies (Lopez-Poveda *et al.*, 2003; Plack and Drga, 2003). The overall pattern is that the frequency-selectivity of the cochlear response shows greater level dependence in high- than in low-frequency regions, suggesting that input level might have a greater effect on efficiency of processing for high- than low-frequency signals.

The results reported by Molis and Summers (2003) concerning the frequency dependence of rollover involved NH listeners and testing in quiet. Previous studies suggest that rollover effects may be fairly similar for NH listeners and hearing-impaired (HI) listeners, under comparable conditions of audibility (Ching *et al.*, 1998; Studebaker *et al.*, 1999). However, assuming that the results reported by Molis and Summers (2003) are closely tied to active cochlear processing, the frequency differences in rollover seen for NH listeners might not be as great for HI listeners, since the influence of active processing is generally reduced in HI listeners as a result of OHC damage. In addition, a number of previous studies have reported larger and more consistent rollover effects in the presence of competing noise than in quiet listening conditions (Pollack and Pickett, 1958; Speaks *et al.*, 1967; Studebaker *et al.*, 1999). These previous studies comparing rollover effects in quiet versus noise did not examine whether greater rollover was present in either high- or low-frequency portions of the auditory stimulus. In this study, we examined rollover for high-frequency, low-frequency, and broadband speech presented in noise. Experimental subjects had normal hearing or flat-configuration sensorineural hearing loss.

II. METHOD

The study consisted of practice/familiarization blocks followed by three experimental tasks. Practice blocks were used to verify accurate speech recognition for broadband speech presented in quiet at moderate levels. In Experiments 1 and 2, sentences were presented at moderate levels. In Experiment 1, high- and low-frequency speech bandwidths supporting ~70%-correct word recognition were determined in quiet. These high- and low-frequency bandpass stimuli, and a broadband stimulus condition, were tested in Experiment 2. For each stimulus set, competing noise levels sufficient to reduce speech scores to ~30% correct were determined. In Experiment 3, the listening conditions [i.e., frequency bandwidths and signal-to-noise ratios (SNRs)] producing constant performance at moderate presentation levels (Experiment 2) were examined at higher levels. The focus was on how speech performance in background noise changed with signal level for low-frequency, high-frequency, and broadband speech; and how differences in hearing status (NH vs. HI) influenced these patterns.

A. Listeners

Six NH listeners (69 to 78 years old; $M=72.5$ years), and 6 HI listeners (63 to 79 years; $M=72.5$ years), were tested. NH listeners had absolute thresholds at or below 20 dB HL (re: ANSI, 1996) at audiometric frequencies up to 4000 Hz (all but one NH listener also met this criterion at 6000 Hz). HI listeners had sensorineural hearing losses as verified by bone-conduction and/or immitance audiometry. Hearing losses were mild to moderate with flat configurations: in general, HI listeners had pure-tone thresholds between 30 and 45 dB HL for audiometric frequencies between

250 and 6000 Hz. Age and pure-tone audiometric thresholds for all listeners are shown in Table I (test ear only).

B. Stimuli

1. Target stimuli

Sentences from the Institute of Electrical and Electronic Engineers sentence corpus (IEEE, 1969), spoken by a female talker, were used as target stimuli in the practice block and in all three experiments. The full sentence set contains 72 lists of 10 phonetically balanced "low-context" sentences, each containing five key words (e.g., The *birch canoe slid* on the *smooth planks*.). Twenty sentences were used to produce two 10-item practice lists. Ninety of the remaining sentences were selected as target stimuli for Experiment 1, 135 sentences as targets for Experiment 2, and 405 as targets for Experiment 3. For each experiment, sentences used as targets were randomized in order and randomly assigned to experimental blocks with each block made up of 15 sentences. Each listener received a unique randomization of target sentences and each sentence was used only once per subject.

As appropriate, target sentences were initially unfiltered or were low-pass or high-pass filtered using a software implementation of a Kaiser filter (>70 dB/octave rolloff beyond the cutoff frequency, verified with a signal analyzer).

2. Masker stimuli

Unmodulated noise with a long-term spectrum matching the average spectrum of the 720 IEEE sentences was used to produce low-pass, high-pass, and broadband maskers for Experiments 2 and 3. In conditions involving low-pass or high-pass filtering, masker and target stimuli were digitally filtered to the same frequency bandwidths prior to D/A conversion. Noise maskers were 25 ms longer than target sentences, with target sentences temporally centered between masker onset and offset.

Target sentences and maskers were converted to analog form by separate channels of a D/A converter (TDT DD1), low-pass filtered at 6.9 and 8.6 kHz, respectively (TDT FT5, >70 dB/octave rolloff above the nominal cutoff frequency), attenuated separately as required for a given signal level and signal-to-noise ratio (TDT PA4), mixed (TDT SM3), amplified (Crown D-75), and led via a headphone buffer to an earphone (Telephonics TDH-49).

C. Procedure

Subjects were seated comfortably in a sound-treated booth facing a computer monitor. Target and masker stimuli were presented monaurally. Subjects were instructed that on each trial they should attend to the spoken sentence and repeat it back as accurately as possible. Subjects' verbal responses were scored immediately for key-words correct and the target sentence was displayed on the monitor. Specific procedures for the practice blocks and for each experiment are listed in the following.

Practice block: Prior to testing, practice blocks were provided to familiarize listeners with the sentence materials and evaluate recognition performance in quiet. Listeners were presented with one or two practice lists of broadband sentences in quiet at 75 dB SPL. Correct identification of at least 80% of key words in a practice list was required before proceeding to the experimental blocks. Listeners were allowed up to two attempts to reach this criterion, using the two practice lists. All but one listener reached or exceeded the 80%-correct criterion on the first practice block (listener HI3 had scores of 78% and 94% correct on the two practice lists).

Experiment 1: Target sentences were low-pass or high-pass filtered adaptively and presented in quiet. Overall stimulus levels varied with filter bandwidth but reflected narrowband levels of the original broadband sentences when presented at an overall level of 75 dB SPL. Filter cutoff frequencies were determined for low-pass and high-pass filtered sentences in separate blocks of trials. Cutoff frequencies of 1500 and 2000 Hz were used for the first presentation of the first sentence of low- and high-frequency test blocks, respectively.

The first sentence of a block was presented one or more times and all other sentences were presented once. On the initial trial of a block, if fewer than four key words were correctly identified, the sentence was repeated as necessary, with the filter cutoff altered by one-third of an octave to increase the signal bandwidth until at least four key words were identified. For the remaining sentences, filter cutoff frequencies were based on the response to the immediately preceding sentence. If the listener correctly identified at least four sentence key words, the bandwidth for the next sentence was decreased by one-third octave. If fewer than four key words were identified, the bandwidth was increased by one-third of an octave. The geometric mean of the filter cutoffs for the last 12 sentences in a given block was used as an estimate of the frequency band supporting 70% key-word recognition for that block.

Subjects took part in a total of six blocks of trials in Experiment 1 (two filter conditions × three blocks per condition). For each subject, filter condition for the first block of trials was randomly selected. Filter conditions then alternated between low and high frequency. For each filter condition, three blocks of trials were presented and the results averaged to determine the frequency bandwidths supporting ~70% accuracy.

Experiment 2: Recognition of low-pass, high-pass, and broadband speech was examined in competing noise. Experimental procedures were similar to Experiment 1. Filter cutoff frequencies for the high- and low-frequency versions of the sentence stimuli were based on the individual results of Experiment 1 (high- and low-frequency bands supporting ~70%-correct key-word recognition in quiet). Adaptive tracking was used to determine competing noise levels which limited performance to ~30%-correct key-word recognition for both filtered and unfiltered (broadband) sentences.

Broadband speech stimuli were presented at 75 dB SPL. Levels for the low- and high-frequency speech stimuli were based on broadband 75 dB SPL stimuli. As in Experiment 1, blocks of trials were composed of 15 sentences with the first sentence of a block presented one or more times and all remaining sentences presented once. The first sentence in each block was initially presented at −15 dB signal-to-noise

ratio (SNR). That is, the nominal overall noise level was 90 dB SPL on this initial presentation. If fewer than two key words were correctly identified following this presentation, the sentence was repeated as necessary, lowering the level of noise (increasing the SNR) by 2 dB with each repetition, until at least two key words were identified. For all remaining sentences, SNRs were based on the listener's response to the immediately preceding sentence. If the listener correctly identified at least two key words, the SNR for the next sentence was lowered by 2 dB. If less than two key words were identified, the SNR was raised by 2 dB. The track continued in this manner to the end of the block.

The average SNR for the last 12 sentences in the block was taken as the SNR supporting 30% key-word recognition for that trial block. For each filter condition, three blocks of trials were presented and the results averaged to estimate the SNR supporting 30% key-word recognition in that condition. Subjects participated in a total of nine blocks of trials in Experiment 2 (three filter conditions × three blocks per condition). This testing was divided into three sets of three blocks. Within each set, each filter condition (low pass, high pass, and broadband) was tested once with filter condition order randomized.

Experiment 3: Filtering conditions and SNRs associated with criterion performance in Experiment 2 were tested at higher presentation levels in Experiment 3. Low-frequency, high-frequency, and broadband versions of the IEEE sentences were presented at the levels used in Experiment 2 (representing broadband levels of 75 dB SPL) and at 12.5 and 25 dB above this level (broadband levels=87.5 and 100 dB SPL). At the higher presentation levels, both sentence levels and noise levels were increased to maintain the original SNRs.

For each subject, filter cutoff frequencies and SNRs were set to values producing ~30%-correct performance in Experiment 2 and were held constant within a block. Percent correct key-word scores were determined for each filter condition at each presentation level in separate trial blocks with each block consisting of a 15-sentence list. Subject responses were scored for the number of correct key words in each sentence and the percent correct for a block was determined based on the total number of key words identified in a list. In Experiment 3, three blocks of trials were presented for each presentation level and filter condition, for a total of 27 blocks of trials (3 level conditions × 3 filter conditions × 3 blocks per condition). For each condition, results were averaged across blocks to determine the overall percent correct.

III. RESULTS

A. Experiment 1

Figure 1 shows mean low- and high-frequency passbands supporting criterion performance by each group in Experiment 1 (performance in quiet at moderate presentation levels). For high-pass stimuli, HI listeners required a slightly broader frequency band than NH listeners to achieve equivalent scores (mean NH cutoff was 1905 Hz, about one-third of an octave higher than the 1539 Hz cutoff for HI listeners; $t(10)=-2.69$, $p<0.03$). However, HI listeners did not require a broader bandwidth than NH listeners for low-pass stimuli (presented in quiet), mean bandwidths did not differ significantly for the two groups [mean cutoffs were 1795 and 1575 Hz for NH and HI listeners, respectively; $t(10)=-1.29$ $p>0.2$]. Percent-correct scores for the final twelve trials of each block (the portion used to compute filter passbands) were similar across groups and filter conditions, ranging between 59.4% and 64.0% correct.

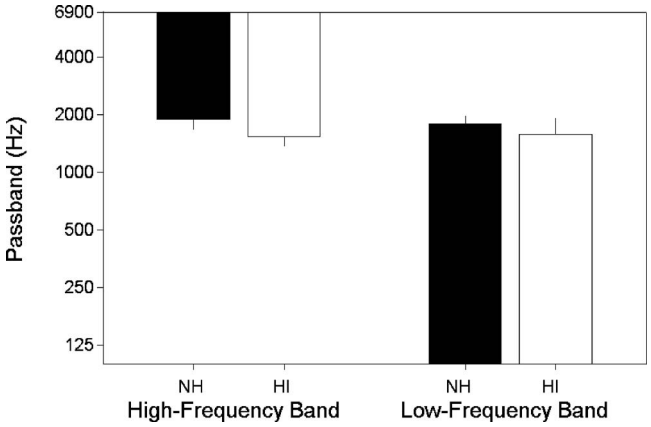


FIG. 1. Mean high-frequency and low-frequency passbands supporting criterion key-word-recognition level in Experiment 1 (in-quiet testing, nominal level=75 dB SPL). Error bars indicate the standard deviations of means.

B. Experiment 2

A two-way ANOVA was carried out to examine the effects of hearing status (NH/HI) and filter condition on SNRs supporting criterion performance in Experiment 2. Figure 2 shows mean SNRs for each group and filter condition. Listeners were able to reach criterion performance at less favorable SNRs in the broadband condition than in either the high-pass or low-pass conditions (i.e., criterion performance was possible at higher noise levels when the signal was not band-limited). Collapsing across groups, mean SNRs were about 10 dB lower for broadband stimuli than for either passband condition. The ANOVA results verified that the main effect of filter condition was significant [$F(2,20)=133.25$, $p<0.001$]. Post-hoc pairwise comparisons

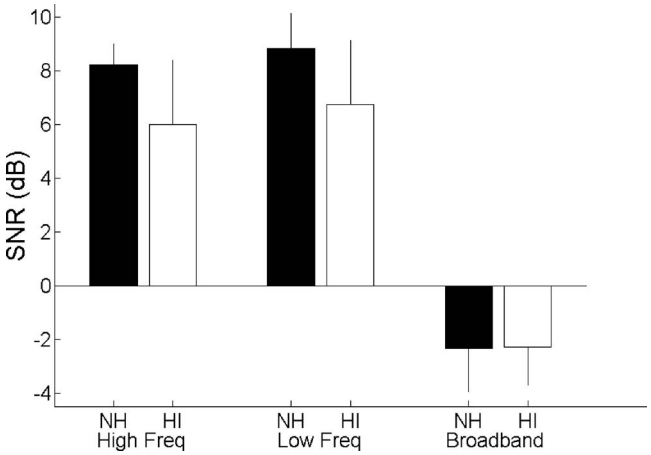


FIG. 2. Mean SNRs supporting criterion key-word-recognition level in Experiment 2. Error bars indicate standard deviations of means.

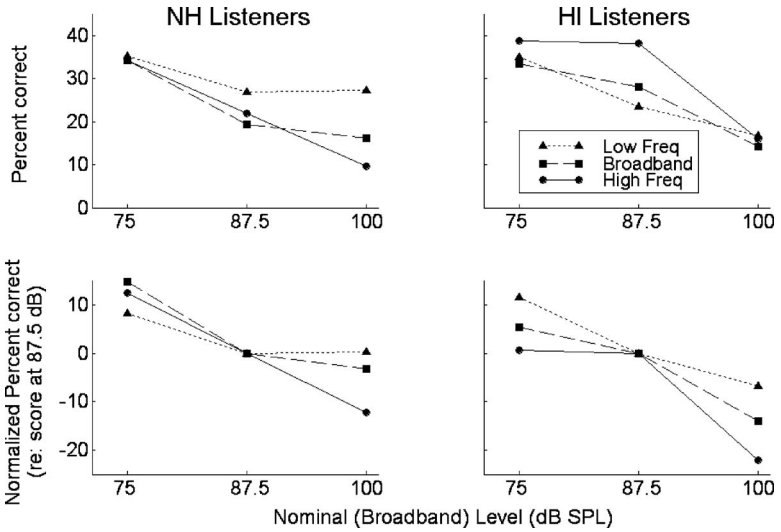


FIG. 3. Mean percent-correct scores as a function of nominal speech level for low-frequency, high-frequency, and broadband conditions in Experiment 3. Upper panels display mean scores and lower panels show normalized scores relative to performance at 87.5 dB. Means for NH listeners are plotted in the left-hand panels and means for HI listeners appear in the right-hand panels. The legend indicates frequency condition.

(Bonferroni-corrected t tests) indicated significantly lower SNRs for broadband stimuli than for either passband condition [$t(11) > 13.4$, $p < 0.001$ in each case], and no significant difference between SNRs for low-pass versus high-pass conditions [$t(11) = 0.91$, $p > 0.38$].

For both low-pass and high-pass sentences, mean SNRs were slightly (~2 dB) higher for NH listeners than HI listeners. SNRs were nearly equal for the two groups in the broadband condition. A requirement of more favorable SNRs for NH listeners than HI listeners in the passband listening conditions would be unexpected and the statistical results suggest that these results may not represent replicable group differences: the main effect of hearing status and the hearing status × filter condition interaction were not statistically significant [$F(1,10) = 3.53$, $p = 0.09$ and $F(1,20) = 1.69$, $p > 0.20$, respectively]. Mean percent-correct scores for the final 12 trials in each block (the portion used to compute SNRs) were similar across groups and filter conditions, ranging between 33.7% and 37.6% correct.

C. Experiment 3

The upper panels of Fig. 3 show mean percent-correct scores for each group and frequency-band condition at each presentation level in Experiment 3. Mean scores for NH and HI listeners are plotted in separate panels. A three-way ANOVA was carried out on these data, examining the effects of hearing status (NH/HI), presentation level, and frequency-band condition on performance (hearing status as a between-subjects variable; level and frequency-band as within-subjects factors). Effects of presentation level on performance, including any significant interactions of level with hearing status and/or frequency band condition were the focus of these analyses, rather than comparisons of absolute performance across subjects or frequency-band conditions.

Presentation level had a significant main effect on scores, with mean scores decreasing from approximately 35% correct to about 26% correct to about 17% correct as levels increased from 75 to 87.5 to 100 dB SPL [$F(2,20) = 53.7$, $p < 0.001$]. The two-way interactions between level and frequency-band condition and between level and hearing status were significant [$F(4,40) = 5.51$, $p = 0.001$; and $F(2.20) = 3.49$, $p = 0.05$, respectively]. These two interactions suggest that presentation level had somewhat different effects on performance by NH and HI listeners and across frequency-band conditions (the three-way level × hearing status × frequency-band interaction was not significant [$F(4,40) = 1.85$, $p > 0.13$]). These differences become apparent when the patterns of level effects are compared within and across the two upper panels of Fig. 3.

Post hoc analyses were carried out to probe the significant level × hearing status interaction further: the NH and HI data were separated and two-way ANOVAs examined the effects of frequency region and presentation level on performance by each group. Within each analysis, the level × frequency region interaction was significant, indicating that the influence of presentation level varied across frequency conditions [$F(4,20) = 4.0$, $p < 0.02$ (NH data); $F(4,20) = 3.3$, $p = 0.03$ (HI data)]. For NH listeners, presentation levels had a large effect on scores for high-frequency stimuli, with mean scores decreasing by about 12% with each 12.5 dB increase in level above 75 dB SPL (overall decrease of ~24%). Low-pass stimuli showed the least amount of rollover for these listeners with scores decreasing by about 8% as levels increased from 75 to 100 dB SPL. For broadband stimuli, rollover effects were intermediate, with mean scores decreasing by about 18% as levels increased by 25 dB.

For HI listeners, rather than producing rollover, the initial 12.5 dB increase in level had almost no effect on mean recognition scores for high-frequency stimuli. For low-frequency stimuli, scores decreased by about 11% when level increased to 87.5 dB SPL and for broadband stimuli, scores were about 5% lower at 87.5 than at 75 dB SPL. These changes in performance associated with the 75- to 87.5-dB level change were quite different from the changes for NH listeners. In addition, for HI listeners, the increase in level from 75 to 87.5 dB SPL had a very different effect than the further increase to 100 dB SPL. These differences may be based on increased signal audibility for HI listeners as a result of the initial level increase. This possibility is examined further using SII analyses to assess signal audibility at each presentation level in the next section.

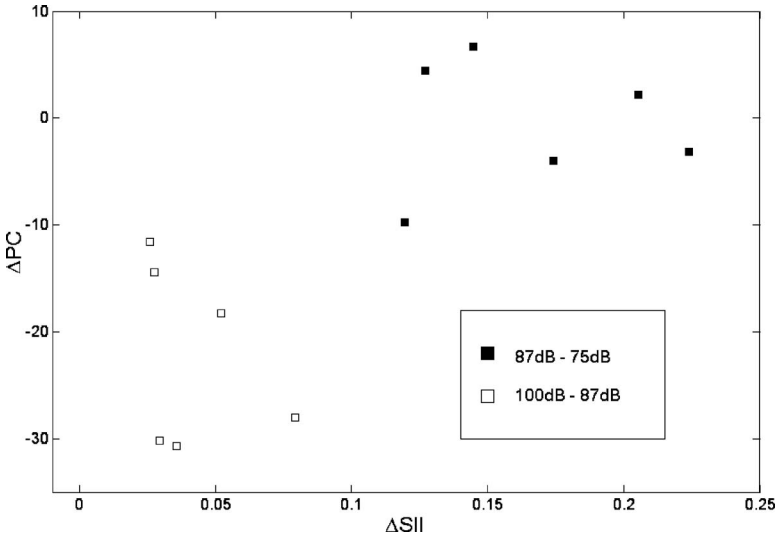


FIG. 4. Changes in mean percent-correct scores (ΔPC) as a function of change in SII scores (ΔSII) for individual HI listeners on high-frequency stimuli (75 to 87.5 dB level increase=closed symbols, 87.5 to 100 dB increase=open symbols).

For the 87.5- to 100-dB level change, the two groups showed much more similar patterns. This similarity can best be seen by replotting the data to align the mean scores for the 87.5-dB presentation level. The lower panels of Fig. 3 display these normalized data. As these panels show, for both groups of listeners, the increase in level from 87.5 to 100 dB SPL produced the largest decrease in scores for high-frequency stimuli, a smaller decrease for broadband stimuli, and the smallest change (or essentially no change) for low-frequency stimuli.

D. SII analysis—High frequency stimuli

For high-frequency stimuli, the level increase from 75 to 87.5 dB led to lower scores (i.e., rollover) for NH listeners only. This may indicate that the changes in high-frequency auditory processing associated with this level increase were greater in NH than in HI listeners. However, if increased audibility in the 87.5-dB condition provided benefit to HI listeners, this benefit may have balanced out any negative effects of distortions associated with the higher sound levels. To examine this issue, software implementing a third-octave band SII analysis (ANSI, 1997) was used to assess audibility of high-frequency stimuli for each HI listener at the three presentation levels. The unmodulated broadband noise masker (with a spectrum matching the mean long-term spectrum of the IEEE sentences) was used to represent the sentence stimuli. This "signal" stimulus was high-pass filtered using the low-frequency cutoff determined for each HI listener in Experiment 1. SII scores were then determined for each listener using audiometric thresholds and third-octave band levels in 75-, 87.5-, and 100-dB versions of the appropriate high-pass stimulus.[1]

Figure 4 shows changes in percent correct scores (ΔPC) between 75 and 87.5 dB SPL (closed symbols) and between the 87.5- and 100-dB-SPL conditions (open symbols), plotted as a function of the changes in SII scores (ΔSII) for each HI listener. SII scores increased by 0.1 or more with the level increase to 87.5 dB SPL, indicating increased audibility for each listener. The 87.5- to 100-dB increase produced only small changes in SII scores (generally less than 0.05), indicating little increase in audibility. Figure 4 shows a clear relationship between ΔSII and ΔPC when the data from the two level increases are combined. That is, level increases producing only small increases in audibility generally lead to rollover and level increases producing larger changes in audibility did not. Combining the data across the two level changes, ΔSII and ΔPC scores were significantly correlated ($r=0.73$, $p<0.007$).

IV. DISCUSSION

The basic finding of poorer speech performance as sound intensities increase from moderate to high levels is well established and is incorporated into the current SII calculation as the level distortion factor. This rollover adjustment is currently treated as independent of hearing status. That is, predicted scores are reduced at high levels in the same manner for NH and HI listeners. The LDF adjustment in the current SII is also independent of frequency: predicted scores gradually decrease as speech levels increase above moderate levels with the adjustment applied uniformly across the frequency range. The current results allow a direct comparison of rollover effects for high-frequency, low-frequency, and broadband stimuli and an examination of whether any effects of frequency region on rollover are similar for NH and HI listeners.

A. Effects of frequency region on rollover

Molis and Summers (2003) reported that for NH listeners tested in a quiet environment, high presentation levels had a more negative effect on speech recognition for high- than for low-frequency speech. The current data extend this result to performance in competing maskers: high presentation levels lowered NH listeners' scores most for high-frequency stimuli, least for low-frequency stimuli, and an intermediate amount for broadband stimuli. Studebaker and Sherbecoe (2002) reported similar findings in a study using narrow-frequency-band stimuli and in-noise testing. In that study, a 20-dB increase above moderate presentation levels had little effect on scores for low- and midfrequency stimuli but produced consistent rollover for high-frequency stimuli.

Hornsby *et al.* (2005) have also reported results consistent with this general pattern. For NH listeners, in a consonant recognition task involving competing noise, level effects were greater (more rollover was observed) for consonantal features with substantial information in high-frequency regions (e.g., place of articulation, duration) than for features strongly represented in low frequencies (e.g., voicing).

Greater rollover at high frequencies is consistent with psychophysical and physiological studies indicating differences in the effects of high sound levels on processing in basal (high-frequency) cochlear regions versus apical (low-frequency) regions. Specifically, basal regions show a clear reduction in frequency selectivity with increasing level which is not observed in the apex (Rhode and Cooper, 1996; Ruggero *et al.*, 1997; Lopez-Poveda *et al.*, 2003; Plack and Drga, 2003). This difference may make an important contribution to the greater rollover effects observed for high-frequency stimuli.

If high-frequency auditory processing is negatively influenced by high levels, this may contribute significantly to the communication problems facing many HI listeners. For these listeners, the amplification necessary to make high-frequency regions audible may lead to distortions in auditory processing that are directly linked to the sound levels involved. This could greatly limit the benefit provided by this amplification. Assuming that low-frequency processing is less level-dependent and that low frequencies are, as a result, less subject to rollover, amplification of low-frequency regions as necessary to achieve audibility may provide greater benefit than similar high-frequency gain. A number of previous studies have reported this basic pattern of results. That is, for listeners with severe high-frequency hearing loss, presentation of high-frequency speech at the levels required to achieve audibility often fails to improve performance (Ching *et al.*, 1998; Hogan and Turner, 1998) while amplification of low-frequency speech to very high levels does benefit listeners with severe low frequency losses (Turner and Brus, 2001). It should be noted that the listeners in Ching *et al.* (1998) and in Hogan and Turner (1998) had more severe high-frequency impairments than the present HI subjects and would therefore require higher signal levels to make high-frequency regions audible. The poor performance reported here for the 100-dB-SPL high-frequency signals (for both groups of listeners) is broadly consistent with the lack of benefit from high-amplitude, high-frequency cues in listeners with severe high-frequency hearing loss.

B. Effects of hearing loss on rollover

At first glance, the upper panels of Fig. 3 seem to indicate that high presentation levels had very different effects on performance for NH and HI listeners. For high-frequency stimuli, the 75 to 87.5 dB increase produced clear rollover for NH listeners but performance remained constant for HI listeners. This could indicate that distortions in high-frequency processing at high levels are greater for NH listeners. However, these distortions may actually be similar for the NH and HI groups if high presentation levels made more of the signal audible for HI listeners. That is, benefit from increased audibility may offset, or even outweigh, the negative consequences of added processing distortions at high levels. The SII results shown in Fig. 4 are consistent with this interpretation. That is, high-frequency signal audibility increased appreciably for nearly all HI listeners when levels increased to 87.5 dB and scores remained fairly constant (rollover was not observed). The further increase in level to 100 dB SPL had little effect on audibility and the decrease in performance with this level increase was actually greater for HI listeners than NH listeners. Previous studies have suggested that once audibility differences are taken in account, the negative consequences of high signal levels are similar for NH and HI subjects (Ching *et al.*, 1998; Studebaker *et al.*, 1999).

For NH listeners, greater rollover was observed for high-frequency stimuli than for broadband, or low-frequency materials. It was suggested that this might reflect differences in the consequences of reduced active processing at high levels in high-frequency versus low-frequency cochlear regions. Specifically, loss of frequency selectivity with increasing level, seen mainly in high-frequency regions, may account for the greater rollover for high frequencies. If loss of active gain makes an important contribution to rollover, it might be expected that rollover effects would be greater in NH listeners than in HI listeners, since active processing is generally reduced with hearing impairment. However, rollover effects were fairly similar in NH and HI listeners when audibility differences were taken into account. This could indicate that loss of active processing is not closely related to rollover. Alternatively, it could indicate that reductions in active gain at high levels are fairly similar for NH listeners and listeners with mild to moderate sensorineural losses. Results reported by Plack *et al.* (2004) appear to support this second possibility. In that study, temporal masking curves were used as a psychophysical means of inferring basilar membrane response functions for NH listeners and listeners with mild to moderate hearing loss. The authors concluded that: "…mild cochlear loss is associated with a reduction in the gain at the *lower input levels only*, and not across the whole range of input levels that are affected by the active mechanism" (p.1693). This characterization suggests that in the current data, the benefits of active cochlear processing may have been available and fairly similar for NH and HI listeners in the 87.5 dB SPL conditions and that reductions in this benefit with the level increase to 100 dB SPL may have also been similar across groups.

V. CONCLUSIONS

The current data are in agreement with the earlier reports (Molis and Summers, 2003) indicating greater rollover in high-frequency regions (above about 1500 Hz) than at lower frequencies. Loss of active processing at high sound levels may have contributed to this frequency-dependent rollover pattern for both NH and HI listeners.

For NH listeners, rollover effects were greatest for high-frequency stimuli, intermediate for broadband stimuli, and least for low-frequency materials. For HI listeners, the same general pattern was observed when presentation level in-

creases had little effect on audibility. However, for these listeners, the distortions in processing at high sound levels may or may not be outweighed by the benefits of increased audibility.

For HI listeners, the high sound levels required to achieve signal audibility may contribute to distortions in auditory processing. Thus, for high-frequency sounds in particular, a portion of the *D* (distortion) component of hearing loss described by Plomp (1978) may reflect the requirement that HI listeners process higher-intensity sounds than NH listeners.

These results may have implications for the SII (ANSI, 1997) and other computational methods of predicting speech intelligibility under various listening conditions. The current SII accounts for rollover with a level distortion factor that is applied equally across all frequency bands. It may be possible to improve the accuracy of the SII if the LDF were modeled as having a greater influence in high- than in low-frequency regions.

ACKNOWLEDGMENTS

Thanks to Nancy Solomon, Mitch Sommers, Chris Plack, and two anonymous reviewers for comments on earlier versions of this manuscript. This research was supported by the Clinical Investigation Service, Walter Reed Army Medical Center, under Work Unit No. 05-25019. All subjects participating in this research provided written informed consent prior to beginning the study. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

[1]The SII calculation generally includes a level distortion factor (LDF) to accommodate rollover at high sound levels. The LDF systematically reduces SII scores as levels increase above moderate levels. The ΔSII scores plotted in Fig. 4 were determined without including the LDF in the SII calculation, in order for scores to more directly reflect pure signal audibility. Inclusion of the LDF in the calculation lowered SII and ΔSII scores slightly but had little effect on correlations between ΔSII and ΔPC.

ANSI (**1996**). *Specifications for Audiometers*, ANSI S3.6–1996 (American National Standards Institute, New York).

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

Beattie, R. C. (**1989**). "Word recognition functions for the CID W-22 test in multitalker noise for normally hearing and hearing-impaired subjects," J. Speech Hear Disord. **54**, 20–32.

Cheatham, M. A., and Dallos, P. (**2001**). "Inner hair cell response patterns: Implications for low-frequency hearing," J. Acoust. Soc. Am. **110**, 2034–2044.

Ching, T. Y. C., Dillon, H., and Byrne, D. (**1998**). "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification," J. Acoust. Soc. Am. **103**, 1128–1140.

Chung, D. Y., and Mack, B. (**1979**). "The effect of masking by noise on word discrimination scores in listeners with normal hearing and with noise-induced hearing loss," Scand. Audiol. **8**, 139–143.

Cooper, N. P., and Rhode, W. S. (**1996**). "Fast travelling waves, slow travelling waves and their interaction in experimental studies of apical cochlear mechanics," Aud. Neurosci. **2**, 289–299.

Dirks, D., Morgan, D., and Dubno, J. (**1982**). "A procedure for quantifying the effects of noise on speech recognition," J. Speech Hear Disord. **47**, 114–123.

Fletcher, H. (**1922**). "The nature of speech and its interpretation," Bell Syst. Tech. J. **1**, 129–144.

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**, 90–199.

Goshorn, E. L., and Studebaker, G. A. (**1994**). "Effects of intensity on speech recognition in high- and low-frequency bands," Ear Hear. **15**, 454–460.

Hagerman, B. (**1982**). "Sentences for testing speech intelligibility in noise," Scand. Audiol. **11**, 79–87.

Hogan, C. A., and Turner, C. W. (**1998**). "High-frequency audibility: Benefits for hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 432–441.

Hornsby, B. W. Y., Trine, T. D., and Ohde, R. N. (**2005**). "The effects of high presentation levels on consonant feature perception," J. Acoust. Soc. Am. **118**, 1719–1729.

Institute of Electrical and Electronic Engineers. (**1969**). *IEEE Recommended Practice for Speech Quality Measures* (IEEE, New York).

Kryter, K. D. (**1946**). "Effects of ear protective devices on the intelligibility of speech in noise," J. Acoust. Soc. Am. **18**, 413–417.

Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (**2003**). "Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing," J. Acoust. Soc. Am. **113**, 951–960.

Molis, M. R., and Summers, V. (**2003**). "Effects of high presentation levels on recognition of low- and high-frequency speech," ARLO **4**, 124–128.

Pickett, J. M., and Pollack, I. (**1958**). "Prediction of speech intelligibility at high noise levels," J. Acoust. Soc. Am. **30**, 955–963.

Plack, C. J., and Drga, V. (**2003**). "Psychophysical evidence for auditory compression at low signal frequencies," J. Acoust. Soc. Am. **113**, 1574–1586.

Plack, C. J., Drga, V., and Lopez-Poveda, E. A. (**2004**). "Inferred basilar membrane response functions for listeners with mild to moderate sensorineural hearing loss," J. Acoust. Soc. Am. **115**, 1684–1695.

Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.

Pollack, I., and Pickett, J. M. (**1958**). "Masking of speech by noise at high sound levels," J. Acoust. Soc. Am. **30**, 127–130.

Rhode, W. S., and Cooper, N. P. (**1996**). "Nonlinear mechanics in the apical turn of the chinchilla cochlea in vivo," Aud. Neurosci. **3**, 101–121.

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.

Speaks, C., Karmen, J. L., and Benitez, L. (**1967**). "Effect of a competing message on synthetic speech identification," J. Speech Hear. Res. **10**, 390–396.

Studebaker, G. A., and Sherbecoe, R. L. (**2002**). "Intensity-importance functions for bandlimited monosyllabic words," J. Acoust. Soc. Am. **111**, 1422–1436.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). "Monosyllabic word recognition at higher-than-normal speech and noise levels," J. Acoust. Soc. Am. **105**, 2431–2444.

Summers, V., and Molis, M. R. (**2004**). "Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level," J. Speech Lang. Hear. Res. **47**, 245–256.

Turner, C. W., and Brus, S. L. (**2001**). "Providing low- and mid-frequency speech information to listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **109**, 2999–3006.

Speech signal modification to increase intelligibility in noisy environments

Sungyub D. Yoo, J. Robert Boston,[a] Amro El-Jaroudi, and Ching-Chung Li
Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261

John D. Durrant, Kristie Kovacyk, and Susan Shaiman
Department of Communication Science and Disorders, University of Pittsburgh, Pittsburgh, Pennsylvania 15261

(Received 3 February 2006; revised 2 April 2007; accepted 30 May 2007)

The role of transient speech components on speech intelligibility was investigated. Speech was decomposed into two components—quasi-steady-state (QSS) and transient—using a set of time-varying filters whose center frequencies and bandwidths were controlled to identify the strongest formant components in speech. The relative energy and intelligibility of the QSS and transient components were compared to original speech. Most of the speech energy was in the QSS component, but this component had low intelligibility. The transient component had much lower energy but was almost as intelligible as the original speech, suggesting that the transient component included speech elements important to speech perception. A modified version of speech was produced by amplifying the transient component and recombining it with the original speech. The intelligibility of the modified speech in background noise was compared to that of the original speech, using a psychoacoustic procedure based on the modified rhyme protocol. Word recognition rates for the modified speech were significantly higher at low signal-to-noise ratios (SNRs), with minimal effect on intelligibility at higher SNRs. These results suggest that amplification of transient information may improve the intelligibility of speech in noise and that this improvement is more effective in severe noise conditions.
 [DOI: 10.1121/1.2751257]

PACS number(s): 43.71.Ft, 43.71.Gv [DOS] Pages: 1138–1149

I. INTRODUCTION

There is an extensive literature on techniques to improve the intelligibility of speech in noise. Most of these techniques, such as active noise cancellation (Elliott and Nelson, 1993), signal subspace approaches (Ephraim and Van Trees, 1995), and related techniques like spectral subtraction (Boll, 1979) and Wiener-filter based approaches (Lim and Oppenheim, 1979), start with noisy speech and attempt to remove as much noise energy as possible with as little effect on the speech signal as possible. These techniques are applied to speech in noise as it arrives at the listener and assume that the properties of the noise, specifically its spectrum, are known. These can be effective strategies, well popularized, for example, in hearing aid technology, but they may not succeed under conditions in which noise is nonstationary or not known.

An alternative approach is to intentionally distort the speech signal to make it more intelligible. Cheng and O'Shaughnessy (1991) suggested that peaks in the speech-plus-noise spectrum represent vowel activity and the highest signal-to-noise ratios (SNRs). They emphasized the spectral peaks and de-emphasized the spectral valleys by using a frequency-domain filter that modeled lateral inhibition. Turicchia and Sarpeshkar (2005) implemented a frequency-specific companding strategy that also increased spectral contrast. By incorporating an implementation of two-tone suppression, their technique was reported to increase spectral contrast only in regions distant from strong spectral peaks and to leave the regions near strong peaks unaffected. These approaches are also intended to be applied to noisy speech at the listener, although they do not require specific knowledge of the noise spectrum.

Another approach to improve the intelligibility of speech in noise is to assume that noise-free speech is available for processing before it is sent to the listener operating in a noisy environment, such as might occur in radio communication between a centrally located coordinator in a quiet environment with field operators in a noisy environment. Sauert and Vary (2006) processed noise-free speech (referred to as the near-end microphone signal) before sending it to a far-end listener in a noisy environment. They assumed that the noise spectrum was known, and they amplified the speech signal based on the SNRs in different spectral regions to increase the spectral distance between speech and noise. This approach increased the intelligibility of the speech in background noise, although the algorithm used signal and noise spectra in a manner similar to traditional speech enhancement.

The above-discussed techniques focused on emphasizing the high energy segments of speech, which primarily represent vowel sounds. A number of studies have shown that

[a]Electronic mail: boston@engr.pitt.edu

consonants, or transitions in speech, are important for speech intelligibility in both normal and hearing-impaired subjects. Strange *et al.* (1983) showed that CVC syllables with vowel centers removed have about the same intelligibility as the original stimuli. Gordon-Salant (1986) manually manipulated nonsense consonant-vowel (CV) syllables to increase consonant duration and the consonant—vowel ratio (CVR) and reported improvements in consonant recognition, particularly due to increased consonant energy (increased CVR). Kennedy *et al.* (1998) studied the effect of increasing the consonant-vowel amplitude in hearing-impaired subjects and showed that these manipulations increase consonant recognition rates of VC syllables. Hazen and Simpson (1998) amplified consonant regions and vowel-onset/offset regions of nonsense syllables and sentence material. They found that relative amplification of the consonant regions provided the greatest improvement in intelligibility, particularly in the correct assessment of place and manner of articulation, although the effect was greater in nonsense syllables than in sentences. These studies used manual wave form editing to isolate consonant and vowel segments of speech sounds and cannot be used in real-time to improve intelligibility.

Nonlinear filtering techniques that can process speech automatically have been applied to emphasize speech consonants over vowels. Niederjohn and Grotelueschen (1976) used high-pass filtering to remove the first formant and amplitude compression to increase consonant energy. They reported increases in word recognition rates of up to 40% in white noise. Skowronski and Harris (2006) used an automatic voicing detector to enhance the C-V ratio by increasing the amplitude of voiceless regions of speech and decreasing voiced regions. They found increases of up to 15% in recognition rates for words presented by a range of speakers in white noise, although the results were not consistent across speech sounds. These studies demonstrated the importance to intelligibility of low energy transitional components in speech, but they are based on the classical spectral view of speech, emphasizing the association of consonants with higher frequencies. Since consonants and transitions from consonants to vowels, and even within vowels, represent changes in spectra, these events may be better represented in the time-frequency domain, and processing based on time-frequency techniques may offer additional advantages to speech intelligibility.

There is no formal definition of transitions in speech, but they are clearly evident in real-time spectral analyses of speech (Stevens, 1998). The dominant characteristic of vowels is quasi-steady or slowly changing short-time spectra. Although consonants are predominantly brief transients, some include quasi-steady-state components, which are represented as hubs. Yet, consonant-vowel, vowel-consonant, and even vowel-vowel and consonant-consonant "interfaces" pervade running speech, creating many transitions. There are even transitions within some vowel sounds, familiar as diphthongs. These transient sounds represent the trajectories of articulators as they move from one position to another, often showing brief frequency shifts that, for a given vowel, may differ among different consonant-vowel combinations (Stevens, 1998). Conventional consonant-vowel classifications and concepts of spectral composition de-emphasize such transient information.

The algorithm described in this paper identifies a speech transition component and attempts to improve speech intelligibility in noise by combining an amplified version of the transition component with the original speech. The approach assumes that noise-free speech is available and does not utilize knowledge of the noise spectrum. As suggested by results of the studies referred to earlier, the transients associated with speech transitions may play an important role in speech perception, especially in more demanding communication situations such as speech in noise. However, because speech transients represent a small proportion of the total speech energy compared to the quasi-steady-state energy in sustained vowels and consonant hubs, they may be particularly susceptible to noise. Amplification of these transients may make the intelligibility of speech more resistant to additive noise.

Traditional methods of studying the auditory system have emphasized frequency-domain techniques, a perspective that also has dominated concepts of speech intelligibility (French and Steinberg, 1947; Kryter, 1962; Fletcher and Galt, 1950). While it is generally recognized that voicing and steady vowel sounds are primarily low frequency and that consonants are dominated by higher frequencies, no single cutoff frequency uniquely separates them. Information on transients between and within vowel sounds is even more difficult to isolate using fixed-frequency filters, as this information is inherently dynamic and can be rather broad band.

Many investigators have addressed the problem of automatically identifying the start and end of phonemes or word segments for automated speech recognition, but only a few studies have focused directly on automatically identifying speech transients. Most of this work has incorporated time-domain features or time-frequency techniques to deal with the limitations of using purely frequency-domain approaches to identify transients. Zhu and Alwan (2000) showed that variable frame-rate speech processing can improve the performance of automated recognition of noisy speech. They used constant duration frames but increased the frame rate when speech models showed that the speech was changing rapidly. Daudet and Torresani (2002) described a method to estimate tonal, nontonal, and stochastic components in audio signals using a modulated cosine transform and a wavelet transform as a step to improve audio coding. Yu and Chan (2000) proposed a transient model for speech coding. The transitions were characterized by the onset time and growth rate of each harmonic component of the transient speech segment. Zhao *et al.* (1997) proposed a speech model and detected spectral transitions for applications to speech or speaker recognition based on time-frequency analysis using the Randon-Wigner and Randon-Ambiguity transforms. Although these researchers investigated the detection of speech transient information, they did not address the relation of the transient information to speech intelligibility.

The algorithm described in this paper uses a time-frequency approach, namely data-adaptive, time-varying filters, to identify speech transients, and the role of these com-

ponents in the recognition of speech in noise is investigated. A speech signal is assumed to be a superposition of quasi-steady-state (QSS) and transient components as

$$S(t) = S_{\mathrm{QSS}}(t) + S_{\mathrm{tran}}(t), \tag{1}$$

where $S(t)$, $S_{\mathrm{QSS}}(t)$, and $S_{\mathrm{tran}}(t)$ are original, QSS, and transient components, respectively. The algorithm decomposes speech into two components. One component is intended to predominately include quasi-steady-state formant activity representing steady portions of vowels and hubs of consonants, and it is referred to as the QSS component. The second component is intended to emphasize transitions between vowels and consonants and within vowels, and it is defined as the transient component.

The energy and intelligibility of the QSS and transient components produced by the proposed algorithm were compared to original speech, and the use of the transient component to modify speech to improve its intelligibility in noise was evaluated. Speech was modified by amplifying the transient component, combining it with the original speech, and normalizing the total energy to be the same as the original speech.

Evaluation of the effectiveness of modified speech in improving intelligibility in noise was measured psychoacoustically. Many of the previous studies of speech enhancement have characterized their results using measures based on the mean differences (in the time domain or the frequency domain) between original speech and modified speech. These methods are not appropriate for our approach because the modified speech signal was deliberately changed from the original speech. Modified speech can be very different and sound very different from original speech, but the important question is which form allows a subject to most effectively recognize words. Hence, the most relevant performance measure is intelligibility. A psychoacoustic listening test, using an adaptation of the modified rhyme protocol (Mackersie *et al.*, 1999), was used to measure the improvement in intelligibility provided by the modified speech.

II. METHODS

A. Speech decomposition algorithm

The approach taken to speech decomposition in this investigation was to first high-pass filter (at 700 Hz) the speech signal to remove most of the voicing energy. It is common knowledge from the high intelligibility of speech over telephones and similar nonbroadband communication devices and of whispered speech that much of the energy at the low end of the speech spectrum is effectively expendable (although important to voice identification and the musical quality of human speech). Three time-varying bandpass filters (TVBFs), whose center frequencies and bandwidths were controlled to pass most of the energy in the three largest formant components, were applied to high-pass filtered speech via digital processing (Yoo *et al.*, 2005). The center frequencies and bandwidths were estimated using information on frequency modulation (FM) and amplitude modulation or envelope obtained from each formant component (Rao and Kumaresan, 2000). The QSS component was composed of the sum of the filter outputs and was subtracted from the original speech signal to yield the transient component. The decomposition can be viewed as removing as much of the dominant QSS formant energy from the original speech signal as possible, while maintaining reasonable intelligibility in the remaining speech signal. That is, the energy of the QSS component was maximized, while keeping reasonable intelligibility in the transient component.

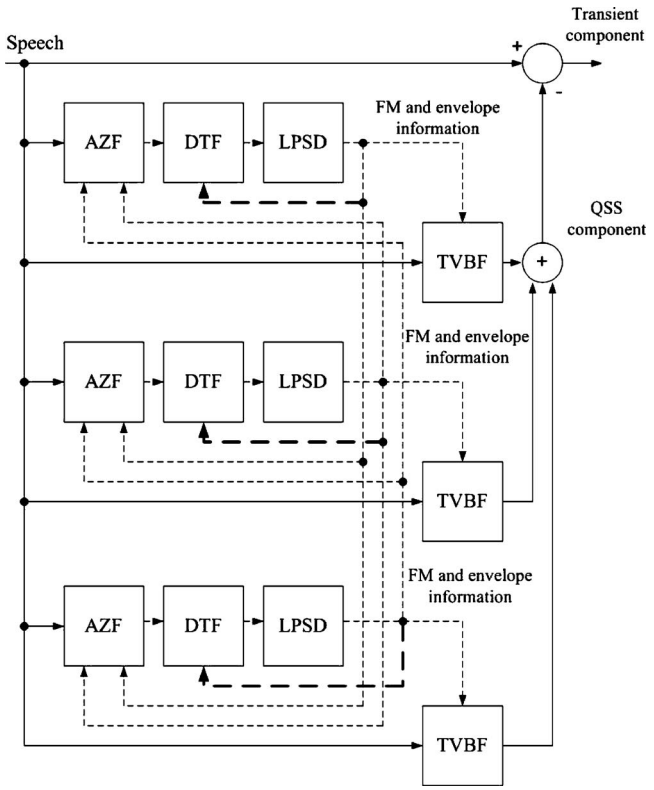


FIG. 1. Block diagram of speech decomposition algorithm. Heavy dashed lines represent DTF center frequency adjustment for a given channel. Light dashed lines represent AZF zero frequency adjustments from other channels. LPSD is linear prediction in the spectral domain.

The tracking of formants by the TVBFs was implemented using a bank of all-zero filters (AZFs) and dynamic tracking filters (DTFs), in which the center frequency of each of the DTFs tracked a formant of the speech. Each formant was decomposed into FM and envelope components using linear prediction in the spectral domain (LPSD). The FM information was used to determine the center frequencies of the TVBFs and to update the zero and pole locations of the AZFs and DTFs. The bandwidths of the TVBFs were based on the envelope information of the tracked formants obtained from the outputs of the DTFs. The output of each TVBF was considered to be an estimate of one of the formants.

A block diagram of the speech decomposition algorithm with three TVBFs is shown in Fig. 1. The input speech signal is filtered by an AZF and a DTF, and then the FM information of the output of the DTF is estimated. The estimated center frequency for the particular formant is used to specify the pole location of the DTF, and the estimated center frequencies from the other filter banks are used to specify the zero locations of the AZF. As a result, the AZF suppresses

the formants tracked by the other channels so that the DTF follows only one formant of the input speech signal.

Since the QSS component is dominated by slow-varying formants or "quasi-steady-state" components of speech, the estimated formant frequency and envelope tend to slowly change with time. In order to reduce computation time, speech was processed in 10-ms blocks, in which the formant information (FM and envelope) was calculated for the first sample in a block and then the filter defined by that information was applied to every sample in the block. This approach was compared on speech material using updates of the filters for every data sample, and the differences in the identified transient components were minimal.

The description of tracking for multiple formants proceeds as follows (Yoo, 2005). Assume there are L formants in a speech signal, and $f_l(n)$ $(l=1,2,\ldots,L)$ represent the individual formants that are to be tracked. To reduce effects of strong neighboring formants, an AZF is applied before the DTF. The zeros of the AZF are adjusted so that a particular DTF effectively filters only one formant. For example, to track the kth formant with frequency trajectory $f_k(n)$, the zeros of the kth AZF are located at $f_l(n)$ $(l=1,2,\ldots,L,\ l \neq k)$, using frequency information from the other channels. The center frequency information of the DTFs tracking $f_l(n)$ $(l=1,2,\ldots,L, l\neq k)$ are used to determine the zero locations of the kth AZF. Thus, the kth AZF's output will not have components being tracked by other filters. The kth DTF will track the dominant formant remaining in the signal.

Each TVBF in this bank was a FIR filter with 150 coefficients to provide high frequency resolution between pass and stop bands. For each filter, the positive instantaneous frequency (FM information) of the output of the corresponding DTF was used as the center frequency, so that each TVBF followed the trajectory of one formant of the speech signal.

The bandwidths of the TVBFs were determined by the envelope energy of each DTF output. In general, the steady-state portions of formants contain higher energy and the transients contain lower energy. The higher energy is expected to involve more harmonics and to be distributed over broader frequency bands. It is assumed that as energy of a formant increases, its bandwidth increases. The bandwidth of a TVBF following a particular formant should depend on the energy of that formant and should change with time as the energy of the formant changes with time. If the formant has large energy at a particular time, the TVBF had a wide bandwidth to pass the wide spread of formant energy. On the contrary, if the formant has small energy at a particular time, the TVBF had a narrow bandwidth.

The bandwidth estimate was based on the weighting function described by Li *et al.* (2001). A maximum bandwidth (B) for the TVBFs was selected, and then a function MBW(t) for the bandwidth was computed according to the SNR of the tracked formant-energy-to-reference-noise energy. The SNR was defined as

$$\mathrm{SNR} = \frac{s_e(t)}{E[n(t)^2]^{1/2}}, \tag{2}$$

where $n(t)$ is a reference noise signal recorded from quiet intervals in the utterance and $s_e(t)$ is the formant envelope (envelope of each DTF's output). The MBW(t) was computed as

$$\begin{aligned} &\mathrm{MBW}(t) = 0 \quad \text{for SNR} \leqslant \alpha, \\ &\mathrm{MBW}(t) = 1 - \frac{\alpha}{\mathrm{SNR}} \quad \text{for SNR} > \alpha, \end{aligned} \tag{3}$$

where α is the bandwidth threshold. The time-varying bandwidth BW(t) was computed as

$$\mathrm{BW}(t) = B \times \mathrm{MBW}(t). \tag{4}$$

The SNR and corresponding bandwidth were computed for each 10-ms time frame.

The MBW(t) increased from 0 to 1 as SNR increased above the threshold. Zero bandwidth corresponds to the TVBF being "off," and the filter is referred to as being closed. Once SNR exceeds the bandwidth threshold, the bandwidth increases with increasing SNR, approaching B asymptotically. When the bandwidth is nonzero, the filter is referred to as being open.

The selection of the maximum bandwidth and the bandwidth threshold parameters is important in the decomposition algorithm. The maximum bandwidth should be large enough to capture most of the energy in the spectral band being tracked but small enough to be restricted to a single band. The bandwidth threshold was based on the ratio of speech power to reference noise power in a spectral band. It should be low enough to assure that the filter is open during a sustained sound and high enough to be closed during speech transients or noise.

Pilot tests with a preliminary word set were used to determine the maximum bandwidths and bandwidth thresholds of the TVBF that most effectively removed QSS energy from the high-pass filtered (HPF) speech, while leaving good intelligibility in the remaining transient component. The bandwidth parameters were systematically varied between 700 and 1100 Hz and the bandwidth threshold between 5 and 18 dB, and intelligibility of the transient component was assessed qualitatively. A bandwidth threshold of 15 dB and maximum bandwidth of 900 Hz provided the lowest energy in the resulting transient components while maintaining good intelligibility, and those parameters were used for the results presented here. This conclusion was verified by the first psychoacoustic test described in the following.

Speech signals, sampled at 44100 Hz, were downsampled to 11025 Hz and high-pass filtered with the 700-Hz cutoff frequency. High-pass filtering was used because, in unfiltered speech, the first DTF usually tracked a QSS component below 700 Hz. The power near the center frequency of the first tracker was usually large enough to hold the filter open throughout the speech signal, and the first TVBF effectively functioned as a low-pass filter with approximately a 700-Hz cutoff frequency. The energy of the QSS and transient components obtained from unfiltered speech with four

trackers and HPF speech with three trackers were essentially the same, and the intelligibility of the QSS and transient components was not different between the two methods. High-pass filtering does not affect speech intelligibility (Lim and Oppenheim, 1979) but it significantly improved the computational efficiency of the decomposition. The QSS component was obtained as the sum of the outputs of the three TVBFs, and the transient component was obtained by subtracting the QSS component from the HPF speech signal.

The number of DTFs was set to three because most vowel sounds that have been high-pass filtered at 700 Hz are composed of two or three dominant formant components. For each filter, the maximum bandwidth was set to 900 Hz, and the bandwidth threshold was set to 15 dB SNR.

If two adjacent formant components are too close in frequency, the bandwidths of these bandpass filters may overlap in some time intervals, and the outputs of these adjacent filters may contain some energy from the same formant. This overlapping of energy results in the QSS component having too much energy. This situation was avoided by limiting the bandwidth of one of the bandpass filters to avoid overlap between bandwidths. Specifically, if two adjacent bandwidths are close enough to overlap, the low-end bandwidth of the filter tracking the higher formant frequency is increased to prevent overlapping.

B. Generation of modified speech

The basic concept of improving speech intelligibility is that the transient component is critical to the perception of speech and that noise affects the transient component more than it affects the QSS component because of the low energy of the transient component. To generate modified speech, original words were decomposed into QSS and transient components, as described earlier. The transient component was multiplied by an amplification factor k and then recombined with a base speech as

$$S_{\text{mod}}(t) = m^*(S_{\text{base}}(t) + k^* S_{\text{tran}}(t)), \tag{5}$$

where $S_{\text{mod}}(t)$, $S_{\text{base}}(t)$, and $S_{\text{tran}}(t)$ represent the modified speech, base speech, and transient component, respectively, and m is an energy factor to adjust the energy of modified speech to be equal to the energy of the base speech as

$$m = \sqrt{\frac{\int S_{\text{base}}(t)^2 dt}{\int (S_{\text{base}}(t) + k^* S_{\text{tran}}(t))^2 dt}}.$$

Amplification factors from 4 to 20 and two different base speech types (recombining with the original and with HPF speech) were preliminarily evaluated by informal listening comparisons. Based on these evaluations, an amplification factor of 12 and use of original speech as the base were selected for psychoacoustic testing.

C. Psychoacoustic testing

A word recognition test used commonly in clinical speech audiometry was used to determine the relative intelligibility of original speech, HPF speech, and the QSS and transient components. A word monitoring test was used to compare the intelligibility of modified speech to that of original speech. These procedures are described in the following.

1. Word recognition test

Monosyllabic words were used to compare speech intelligibility among the original speech, HPF speech, QSS component, and transient component (Robinson and Watson, 1973; Stevens, 1998). Three hundred consonant-vowel-consonant (CVC) words from the NU-6 word lists (Tillman and Carhart, 1966) were processed as described earlier to provide HPF, QSS, and transient components for each word. These word lists have well-established psychometric characteristics and have been widely used in clinical research and hearing tests. Test words (original, HPF, QSS, and transient versions) were presented in a quiet background to five volunteer subjects with negative otologic histories and hearing sensitivity of 15 dB HL or better by conventional audiometry (250–8000 Hz). Subjects sat in a sound-treated booth, and test words were delivered monaurally through headphones (right ear, chosen arbitrarily; Telephonic TDH-39). Subjects were asked to repeat the words presented, and the number of errors in word identification was recorded by skilled examiners under supervision of a certified clinical audiologist. For each component, stimuli were presented at five intensity levels from 0% recognition until recognition reached 100% or did not increase. This approach avoided any *a priori* assumptions about the sound pressure level of word presentation to achieve PB_{max} (asymptotic score), which is only estimated by the typical clinical test method.

Following principles of classical psychophysics and common concepts of speech recognition ability as a function of word presentation level, recognition results for each subject were fit to an error function (cumulative normal distribution or *ogive*), using the nonlinear least-squares fit routine "lsqcurvefit" (MATLAB, The Mathworks, Inc.). The function minimum was set to zero, and estimates of the PB_{max}, midpoint (50% of PB_{max}), and slope (measured by the standard deviation parameter associated with the *ogive*) were obtained. The mean squared difference between the fitted function and the original data divided by the total mean square of the data (R^2) was calculated to assess the adequacy of the fit, with $R^2 > 0.8$ taken to indicate a satisfactory fit. The parameters for the function fits for original, HPF, QSS, and transient versions of the words were tested for differences across versions. The Friedman test was used as a nonparametric analysis of variance because of concerns that the recognition results were not normally distributed. Significant results ($p < 0.05$) were followed by using Wilcoxon paired comparisons to identify specific pairs of parameters with statistically significant differences.

2. Word monitoring test

The modified rhyme word monitoring test (Mackersie *et al.*, 1999) was adapted to compare speech intelligibility between original and modified speech. This test provides a direct quantitative measure of the intelligibility of a message spoken over any system and requires minimal training of the

listeners. In addition, the test stimuli can be repeatedly used with the same listeners with minimal learning effects.

Three hundred monosyllabic words (50 sets of rhyming words) were recorded by a male native speaker of English. Each stimulus set consisted of six monosyllabic rhyming CVC words. Additional words recorded by the same male speaker were processed for training purposes. Test words were presented with six different SNRs (−25, −20, −15, −10, −5, and 0 dB) of speech-weighted background noise, which approximates the long-term sound pressure spectrum level of speech (ANSI S3.6, 1996). The spectrum level of the noise was constant from 100 to 1000 Hz and decreased at a rate of 12 dB/octave from 1000 to 5512 Hz (half the sampling rate). Length of the experiment prohibited use of multiple noise types. Thus, only speech-weighted background noise was tested because it is representative of commonly encountered noise from the environment—the cacophony of sound from multiple talkers.

Each word was normalized to unit root-mean-square amplitude. The background noise was presented for 1.83 s and gated by a Tukey window for a smooth onset and offset. The window rise and fall times were 0.25 s. The amplitude of the background noise was adjusted to one of six SNRs for the word, and the words were presented with this background noise. Each SNR was defined by the amplitude ratio of the word and noise over the same time interval. The interval between each speech token in noise was 0.25 s. The order of presentations and SNRs were randomized under computer control. Subject responses were recorded by the computer.

Eleven volunteer subjects with negative otologic histories and hearing sensitivity of 15 dB HL or better by conventional audiometry (250–8000 Hz) were tested. Subjects sat in a sound-treated booth, and test words with background noise were monaurally delivered though headphones (right ear, chosen arbitrarily; Telephonic TDH-39). Each trial involved one set of six rhyming words, with one of the words selected as the target. At the beginning of each trial, the target word appeared on the computer monitor and remained until all six alternative words were presented. The first word among six alternative words was presented 1 s after the target word appeared on the computer monitor. The subjects were asked to press a mouse button as soon as they heard the target word.

After a training trial to familiarize the subject with the task, 50 sets of rhyming words were repeated 6 times for each subject (total 300 sets of rhyming words per subject). One hundred fifty of the 300 sets were presented as original speech and the remaining 150 sets were presented as modified speech. The target words were randomly selected from the 300 monosyllabic words and the selected target word was excluded as a target in future trials (the same word was not used as a target more than once). The sets were presented at six different SNR levels of speech-weighted noise (25 sets for each noise level and speech type).

Although recognition rates at several SNRs were measured, the null hypothesis was specified for significance testing that original and modified speech would be equally recognizable at SNR=−20 dB. This hypothesis was based on a preliminary study in two additional subjects that showed

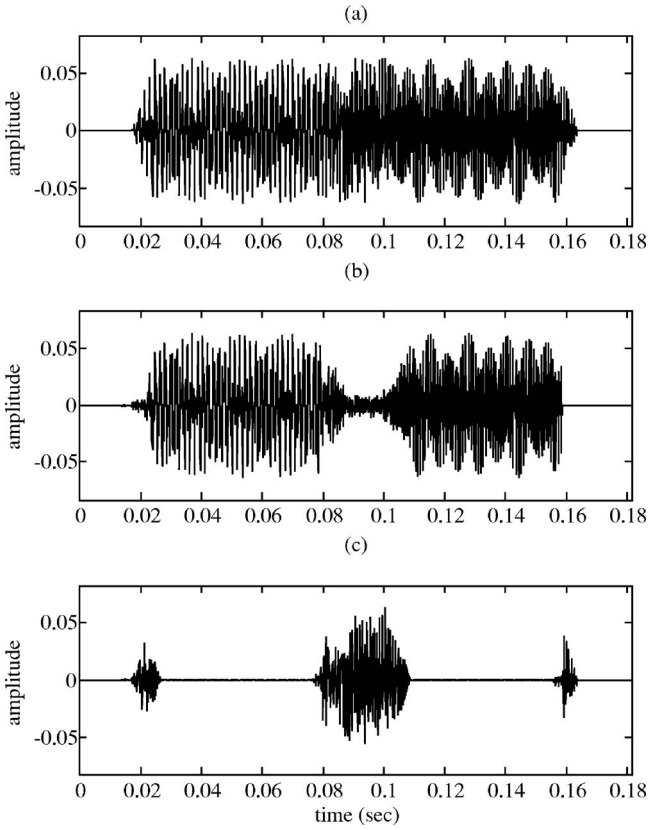


FIG. 2. Synthetic signal wave forms used to illustrate the algorithm: (a) original, (b) QSS, and (c) transient components.

that, at SNR=−20 dB, recognition rates for both versions were reasonably high, with modified speech being substantially more recognizable. The distributions of recognition scores for the original and modified speech from the 11 subjects were examined and confirmed to be normally distributed, and consequently a paired t-test was used to test the null hypothesis, with $p<5$ taken to indicate a significant difference. Recognition rates at other SNRs are presented in terms of means, 95% confidence intervals, and p-values obtained from paired t-tests.

III. RESULTS

A. Analysis of a synthetic signal

A synthetic chirp signal was analyzed to illustrate how the proposed TVBFs are formed and how the decomposition algorithm can extract transient information and to check the performance of the algorithm with deterministic signals. The synthetic signal was synthesized at a sampling rate of 11 025 Hz with a duration of 180 ms. It consisted of three tones (frequencies at F1, F2, and F3), followed by three positive chirps, and then followed by three tones (frequencies at F4, F5, and F6). The duration of each tone+chirp+tone was 140 ms, and each onset and offset was 7 ms. This synthetic signal was chosen because these three tones and onsets and offsets are similar to the vowel sounds in a simple speech signal, and the three chirps are similar to formant transitions. The frequencies F1, F2, and F3 were 574, 1786, and 2999 Hz, respectively, and the frequencies F4, F5, and F6 were 2514, 3726, and 4939 Hz, respectively. The steady

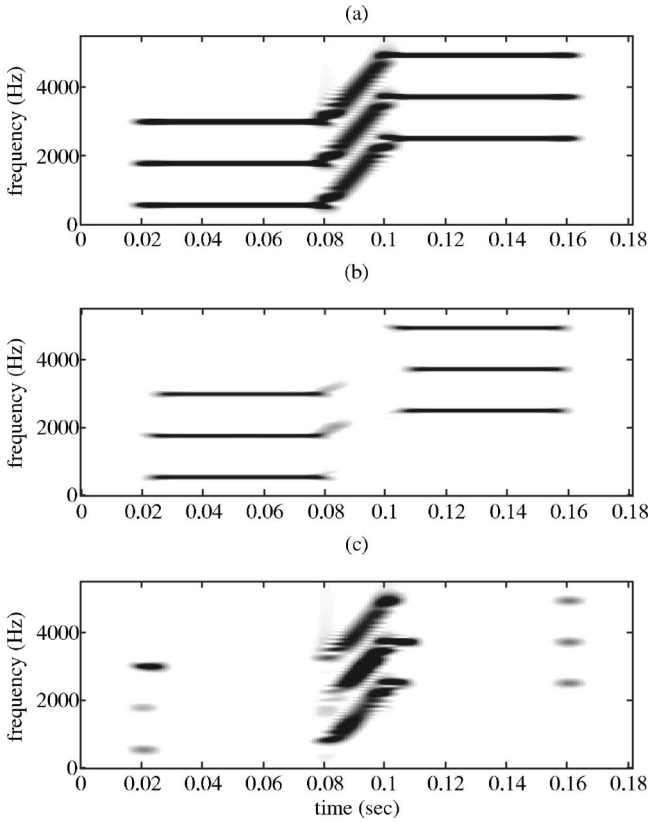


FIG. 3. Synthetic signal spectrograms used to illustrate the algorithm: (a) original, (b) QSS, and (c) transient components.

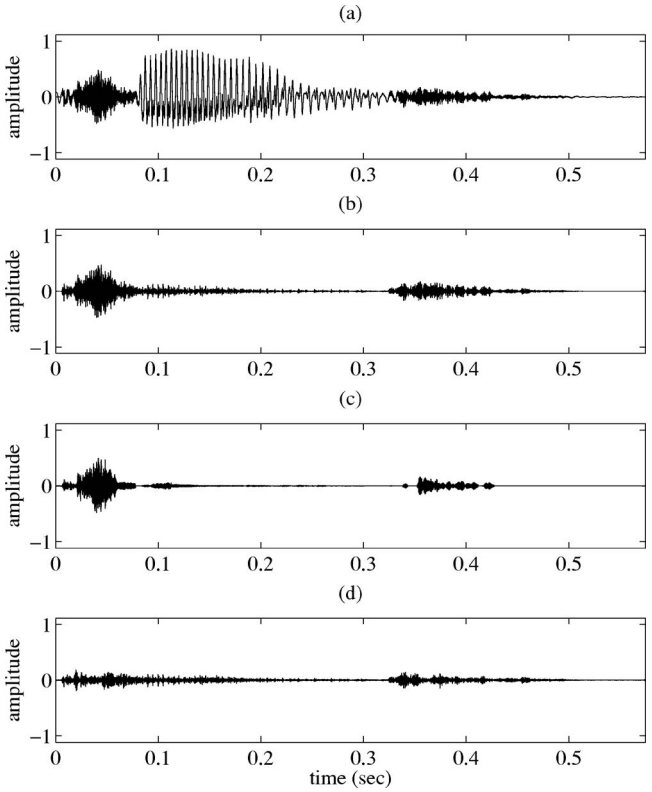


FIG. 4. Wave forms of decomposed real speech signal "Juice" spoken by a female speaker: (a) original, (b) HPF, (c) QSS, and (d) transient components.

tones lasted for 60 ms with a 20-ms chirp interval between them. The chirp rate for this signal was 97 Hz/cm.

The QSS component was expected to contain the quasi-steady-state energy of the synthetic signal. The transient component should be dominated by onset and offset parts of the tones as well as the chirps between tones and should contain relatively little energy. The number of DTFs in the filter bank was set to 3 to match the number of tones and chirps. The maximum bandwidth was set to 900 Hz and bandwidth threshold was set to 15 dB SNR. A constant reference noise energy, derived from silent parts of a single speech phrase, was used as a reference noise signal for SNR calculation. The reference noise signal was not added to the original synthetic signal. In essence, time-varying bandwidth was computed based on signal power.

The original, QSS, and transient wave forms decomposed by TVBFs are shown in Fig. 2. The steady-state parts of the three tones are effectively passed through the TVBFs, and the sum of these filter outputs comprise the QSS component. The difference between original and QSS components is the transient component. The onsets, offsets, and chirps are appropriately filtered out by the TVBFs, and they are shown as the transient component. The QSS and transient components contain 82% and 18% of the total energy of the synthetic signal, respectively.

The time-varying characteristics of the decomposed QSS and transient components are illustrated in spectrograms in Fig. 3. The spectrograms were calculated using a Hanning window with length of 1/10 of the signal, and the spectrum of the windowed signal was estimated by a fast Fourier transform. The estimated spectrum formed one time section of the spectrogram. The window was translated 1 ms and then the above-mentioned processing was repeated until the sliding window covered the entire synthetic signal. The spectrograms confirm that QSS and transient components clearly separate the tonal and transitional parts of the signal, demonstrating that the proposed algorithm is able to identify the transient component.

B. Example of real speech decomposition

An example of decomposition of a speech signal spoken by a female speaker is illustrated in Figs. 4 (time analyses) and 5 (speech spectrograms). A monosyllabic word ("Juice," represented phonetically as /dzu:s/) was decomposed into QSS and transient components as described earlier. The original, HPF, QSS, and transient components decomposed by TVBFs are shown in Fig. 4. The energy in the HPF speech is 9% of the energy in the original speech. The energy in the QSS component is 78% of the energy in the HPF speech (7% of the original speech energy). The QSS component is dominated by the consonant hub (/dz/) at approximately 0.01–0.07 s, and it also includes most of the fricative sound (/s/) at around 0.37 s. The remaining 22% of the energy of the HPF speech is in the transient component (2% of the original speech energy) and includes energy associated with the onset and offset of the consonant hub at around 0.01 and 0.07 s and the beginning and ending of the fricative sound at around 0.30 and 0.38 s.

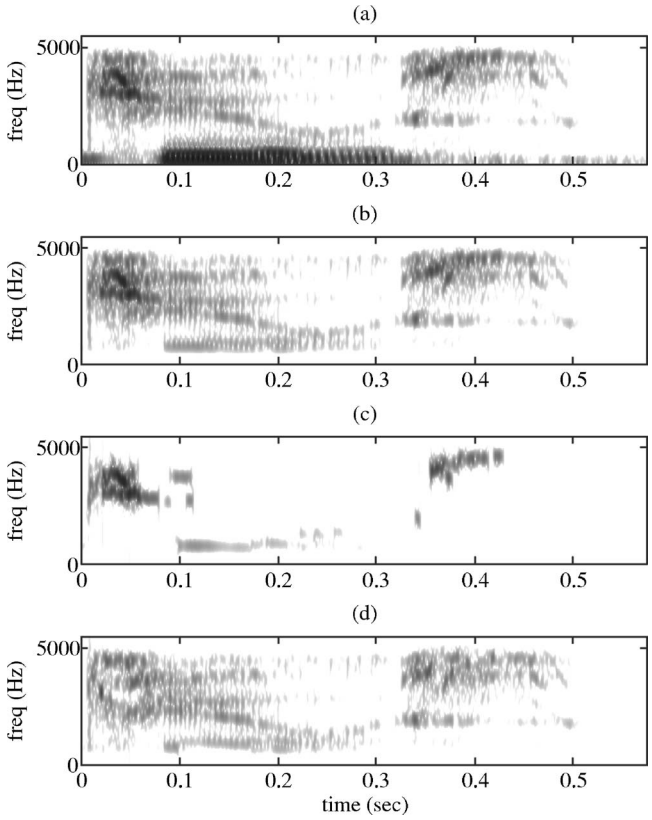


FIG. 5. Spectrograms of decomposed real speech signal "Juice" spoken by a female speaker: (a) original, (b) HPF, (c) QSS, and (d) transient components.

The speech spectrograms of these signals (Fig. 5) were calculated using the procedures presented in the previous section. The spectrograms are consistent with Fig. 4, demonstrating that most of the sustained consonant hub /dz/ is included in the QSS component and most of the transition activities, onset and offset of the consonant hub, and beginning and ending of fricative sound are in the transient component.

The QSS component was very garbled and not readily identifiable as the word "Juice" from informal evaluations by the experimenters. On the contrary, the transient component was nearly as easily recognized as the HPF version, despite having much less energy.

For the 300 words used for the word recognition test, the energy of each component of the decomposition was calculated and compared to the original and HPF speech. The transient components averaged 2% of the original speech energy (18% of the HPF speech energy), and the QSS component averaged 18% of the original speech energy (82% of the HPF speech energy). The perceptual loudness of the QSS component was approximately equal to the HPF speech, but the transient component sounded less loud, as would be expected due to the lower energy.

TABLE I. Mean growth function parameters (standard deviation in parentheses).

	Midpoint	PB_{max}	Standard deviation
Original speech	0.3 (2.7)	98.7 (3.0)	7.1 (3.2)
HPF speech	−11.2 (3.8)[a]	96.5 (2.1)	7.2 (2.5)
QSS component	2.2 (11.3)	45.1 (19.3)[a]	5.6 (8.5)
Transient component	0.5 (4.6)	84.9 (14.4)	12.1 (6.3)

[a] $p<0.05$ for pairwise comparisons with other components.

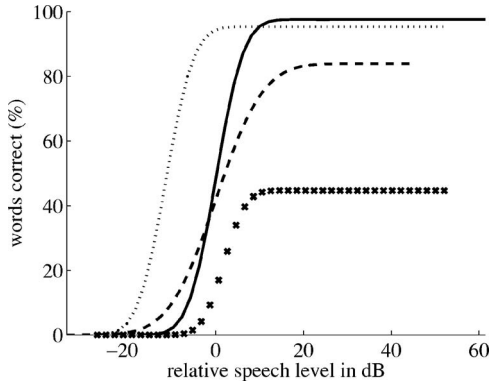


FIG. 6. Growth of word recognition based on mean *ogive* function parameters. Solid line: original speech; Dotted line: HPF speech; x-x: QSS component; dashed line: transient component. Abscissa is component energy in decibels relative to the energy in original speech at the 50% word recognition level.

C. Psychoacoustic results: Relative intelligibility of components

The word recognition test was used to compare intelligibility of original, HPF, QSS, and transient speech components. Growth in word recognition rates with increasing SNR was successfully fit for each subject to *ogive* functions. Of the 20 sets of data (4 different word versions for 5 subjects), 18 were fit with $R^2>0.9$ and 2 with R^2 between 0.8 and 0.9. Means and standard deviations of the ogive function parameters are summarized in Table I. Figure 6 illustrates growth functions for each speech component using the mean parameter values, where speech level for each component is expressed in decibels with respect to the level for the midpoint (i.e., 50% of PB_{max}) of the growth function for original speech.

The Friedman test applied to each parameter showed that the midpoints ($p=0.026$) and PB_{max} ($p=0.016$) were significantly different. Wilcoxon paired comparison results for these two parameters are summarized in Table II. The midpoint for the HPF speech was at a significantly lower SPL than midpoints for the other components, which were not different from each other. For PB_{max}, the QSS component was significantly different from the other three versions. The

TABLE II. *p*-values obtained from Wilcoxon paired comparison tests.

	Midpoint	PB_{max}
Original—High-pass filtered	0.043[a]	0.144
Original—QSS	0.893	0.043[a]
Original—Transient	0.893	0.109
High-pass filtered—QSS	0.043[a]	0.043[a]
High-pass filtered—Transient	0.043[a]	0.225
QSS—Transient	0.715	0.043[a]

[a] $p<0.05$ for pairwise comparisons with other components.

TABLE III. Differences (modified speech—original speech) in word recognition scores.

SNR (dB)	Mean difference	Standard deviation of difference	95% confidence interval of difference
−25	32.0	12.1	23.85 to 40.15
−20	25.5	7.4	20.46 to 30.45
−15	17.8	12.2	9.64 to 26.00
−10	10.5	18.6	−1.96 to 23.05
−5	−2.5	6.3	−6.76 to 1.66
0	0	9.3	−6.24 to 6.24

QSS component, despite having most of the energy of HPF speech, had significantly lower PB_{max} than the other components, confirming the preliminary observation that this component has poor intelligibility. The standard deviations of growth functions were the same for all versions, indicating that the slopes of the growth functions were not significantly different across components.

These results showed that HPF speech is about as intelligible as original speech and that the transient component is only slightly less intelligible than the original and HPF speech. The QSS component, however, is much less intelligible. These results support the suggestion that transient information in speech may be important to speech perception.

D. Psychoacoustic results: Intelligibility of modified speech

The modified rhyme protocol was used to compare the intelligibility of original speech in noise to modified speech in noise, produced using the procedure described in Sec. II B. The null hypothesis that there is no difference in speech recognition rates between original and modified speech at SNR=−20 dB was tested using a paired t-test. The mean difference at this noise level across subjects was 25.5%, with standard deviation 7.4%, which is significantly different from zero at $p<0.05$. The null hypothesis was thus rejected.

Table III shows mean paired differences and 95% confidence intervals between recognition rates of original and modified speech at all SNRs tested. At lower SNRs (−25, −20, and −15 dB), the 95% confidence intervals for the differences in recognition scores did not include zero, indicating that the subjects could identify the modified speech better than the original speech. The intervals did include zero for higher SNRs (−10, −5, and 0 dB), suggesting no difference in word recognition rates at these levels. Differences among SNRs naturally are delimited by ceiling effects at these more favorable conditions, but of particular interest was that the signal processing method did not degrade the inherently good speech recognition ability at high SNRs.

To compare overall recognition results, means and 95% confidence intervals for word recognition scores for original speech, averaged across subjects, and modified speech, averaged across subjects, are shown for all SNRs tested in Fig. 7, where the dashed line represents intelligibility of modified speech and the solid line represents intelligibility of original speech. For both speech versions, percent correct scores decreased as the SNRs decreased from 0 to −25 dB. At higher SNRs, the recognition scores for modified speech were similar to the recognition scores for the original speech. As the SNRs decreased, the recognition scores for original speech decreased more rapidly than the recognition scores for modified speech. These results suggest that at lower SNRs (−25, −20, and −15 dB), speech recognition can be improved by selectively amplifying the transient component.

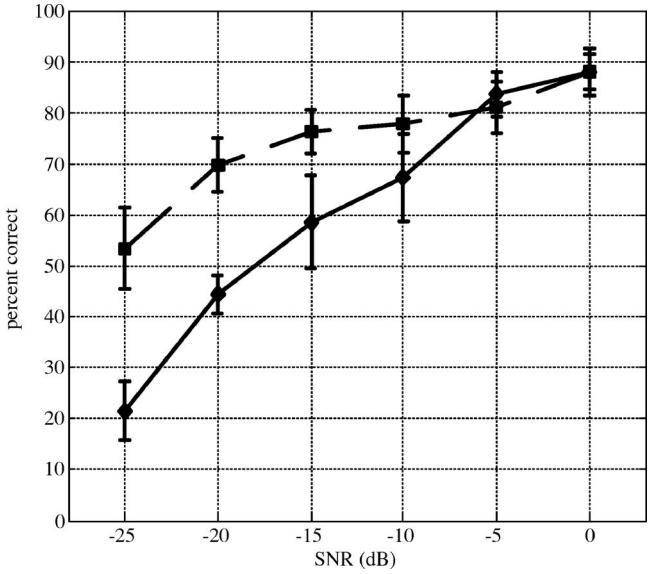


FIG. 7. Means and 95% confidence intervals of word recognitions (%) for original (solid line) and modified (dashed line) speech, averaged across subjects.

IV. DISCUSSION

Since speech transients can be expected to be distributed across time and frequency, fixed frequency filters may not be adequate to optimally identify them. Our algorithm employed a bank of adaptive bandpass filters, with time-varying center frequencies and bandwidths, to remove steady-state energy from the signal, leaving a transient component. Although the remaining signal represents a small proportion of the total speech energy, psychometric measures of maximum word recognition rates showed that the transient components were almost as intelligible as original speech, while the steady-state signal that was removed was much less intelligible. These results demonstrate the importance of the transient component to speech perception and suggest that an emphasis of the transient might improve speech recognition in noisy conditions.

To produce modified speech, the transient component was amplified and recombined with the original speech, and the intelligibility in noise of the modified speech was compared to that of original speech using a modified rhyme test protocol. The results demonstrated that the modified speech provided significant improvement in word recognition rates at very low SNRs without compromising performance under much more favorable SNR conditions (i.e., approaching quiet).

It is difficult to directly compare these results with earlier work. Hazan and Simpson (1998) tested their manually edited stimuli in speech-weighted noise and obtained in-

creases in intelligibility scores of 6% at 0 dB and 12% at −5 dB SNR for nonsense syllables, similar to the results obtained by Gordon-Salant in speech babble. Improvements for sentence material were lower, but with modifications to the original algorithm, they were able to obtain a 4% improvement at 0 dB. Our method did not begin to show improvement until SNR was below −5 dB. Differences in the noise spectrum, differences in the intelligibility test procedure, and limitations in our algorithm to extract speech transitions may contribute to the reduced performance. Our algorithm is an automated procedure, and if Hazan and Simpson's improvement continues to increase at the lower SNR levels that we are interested in, their results suggest that further improvement in the algorithm may be possible.

Two evaluations of automated algorithms that increased the consonant-vowel ratio have been tested in white noise. While Skowronski and Harris (2006) obtained intelligibility increases of up to 15% at SNR=0 and −10 dB on some speech sounds, their results were variable across sounds and speakers. They did not report an average result. Niederjohn and Groteluesch en (1976) obtained increases of up to 40% at SNR=0 dB but they had no improvement at −10 dB. Because of the different types of noise used, it is not clear at what noise levels comparisons would be most relevant. It is compelling that our results continue to show improvement in release from masking with decreasing SNR (that is, increasing noise level), conditions under which the auditory system will be more challenged and the redundancy of running speech less likely to make up for direct interference of noise on time-frequency cues in speech.

Most traditional studies of speech enhancement have focused on noise, either to directly reduce the noise by filtering or to modify the speech to take advantage of frequency regions that are less affected by the noise. The focus in this study was to modify the original speech signal itself. Our algorithm operates independently of the noise at the far end listener, and speech enhancement techniques based on noise reduction applied at the listener's end could be combined with our approach, potentially providing the same SNR improvement to our modified speech as they do to original speech. However, whether these noise reduction techniques would have the same impact on intelligibility as they do for original speech is not known and should be investigated.

The problem we were considering is radio communication between a centrally located coordinator in a quiet environment with field operators in a noisy environment, and we assumed that the noise-free speech is available for processing before it is sent to the far end listener. We did not investigate the performance of our algorithm on noisy speech. Since the transient speech component has very low energy, we would expect it to be difficult to identify in noisy speech, especially at the SNRs of interest in this study (−10 dB and below). We are investigating modifications of the algorithm that would be effective with noisy speech, with traditional speech enhancement to minimize noise being a first step.

We used a psychoacoustic listening test to evaluate the improvement in speech intelligibility provided by modified speech. Quantitative measures such as mean square error between original and modified speech or between their spectra are not appropriate performance measures because the modified speech signal was deliberately changed from the original speech. Our modified speech looks and sounds very different from the original speech, but the relevant performance measure is recognition rates for words in noise. Indices such as the Articulation Index and similar measures may provide quantitative measures of modified speech, but they are based on long-term spectra and do not incorporate temporal changes in the spectra (Sauert and Vary, 2006). The spectral differences between our modified speech and original speech are primarily temporal, and differences in the long-term spectra are minimal. That is, long-term spectral measures do not capture the time-varying frequency characteristics of the differences between modified speech and original speech. Hence, we concluded that direct intelligibility tests were most appropriate to characterize the improvement provided by modified speech. The modified rhyme protocol that we used requires minimal training of the listeners, and the test stimuli can be repeatedly used with the same listeners with minimal learning effects. It can be used with different speakers and different environmental noises. However, because testing under a given set of conditions is a lengthy process (only one set of conditions can realistically be evaluated in a given experiment), we tested intelligibility using only speech-weighted noise for this demonstration of concept study.

Although these results are compelling, there are, naturally, limitations that restrict our ability to predict success of the same degree in the real world. The modified rhyme test protocol provides a well-controlled artificial environment, and the ability to achieve this level of improvement in natural communication is not clear. Subjects had a well-defined task, and, because they could see the target word before stimulus presentation, strong expectations of what the stimulus word would be. The test involved simple speech material, a single speaker, and a constant type of noise that may have allowed subjects to adapt to the stimulus conditions. The word lists were not balanced to the phonetic distribution of conversational speech, and some sounds may be over- or under-represented in the stimulus material. However, testing in a more natural context would require less control over experimental conditions and would be expected to result in greater variability in the results (more difficult to achieve statistical significance).

On the other hand, the task involved subtle differences in stimulus words and hence was very demanding. Subjects were presented the words with no context, which in normal conversation would provide strong cues for correct word recognition. These effects might improve a subject's ability to understand a modified version of normal conversation in noise. Still more extensive evaluations with a variety of speakers, types of noise, and range of conversational material are necessary to determine the potential effectiveness of this technique.

English contains a mix of vowel and consonant sounds, resulting in a mix of transient and steady-state energy. Other languages have different mixes, which would be expected to affect the utility of this approach to modified speech. There are certainly transients associated with vowels, but the role

of transients in a language like Korean, which relies extensively on vowel sounds, may be different from their role in English. A language such as Polish with more pronounced consonant sounds may emphasize transients more than English. Further study is needed to understand the relative roles of transient and steady-state speech energy in different languages and the effectiveness of improving speech recognition based on amplification of transients in these languages. The parameter settings (e.g. number of TVBFs, amplification factor, etc.) used in this study were optimized in English. For different languages, different parameter settings may be required for best performance.

The motivation for this investigation was improvement of speech recognition, perhaps (ultimately) in conjunction with noise-reduction signal-processing and/or transducers, for normal-hearing listeners working in very noisy environments. Much related work in the past has focused on improved speech-in-noise performance of hearing-impaired listeners. This population was avoided in this study. Try as one may to replace dysfunction (e.g., associated with the loss of hair cells), such as by using a combination of multiband compression amplification and frequency-response equalization, the receiver of the modified speech signal is a defective organ. This is in contrast to the vast majority of users of corrective eyeglasses, whose function is largely completely restored, as the problem addressed is presensory. Indeed, even the most successful hearing-aid users have difficulties in very noisy environments. Still, it seems likely that such a processing algorithm as presented here could have application in amplification for the hearing impaired, as conventional approaches, while increasingly flexible, may still suffer from fixed-filter approaches. It thus is speculated that compression amplification might be more effective given a pretreatment of the signal that more dynamically channels the speech sounds of greatest importance for intelligibility.

The time required for computation of the transient component using this algorithm is approximately 40 times the duration of the speech itself. The goal of this study was to evaluate the potential of this approach to improve speech intelligibility rather than to develop a practical implementation, and the algorithm, which was coded in MATLAB, was not optimized for real-time processing. Coding directly in "C" and using specialized hardware, such as a Digital Signal Processor chip, would provide some improvement. Increasing the block size to greater than the 10 ms used here could also decrease computation time. We are currently investigating the use of wavelet-based time frequency analysis, which can be implemented very efficiently, to approximate the transient component obtained by this algorithm.

V. CONCLUSION

A new dynamic method to extract transient information from speech has been described. Time-varying bandpass filters whose center frequencies and bandwidths are controlled to pass most of the energy in the three largest formant components in HPF speech were designed to extract QSS energy. The signal with the QSS component removed was referred to as the transient component of speech. The transient component retained most of the intelligibility of the original speech, while the QSS component was much less intelligible. Modified speech was formed by emphasizing the transient component, and psychometric measures of word intelligibility demonstrate that the modified speech can provide significant improvement in speech intelligibility at low SNR levels. The results suggest that amplification of transient information can improve speech recognition in noise. Whether equally significant improvements can be obtained in other languages and/or hearing-impaired listeners remains to be seen.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research under Grant No. N000140310277, J. R. Boston, principle investigator and J. D. Durrant, co-investigator. The authors would like to express their appreciation to Ken Morton and Robert Nickl, who prepared the speech material for the psychoacoustic tests.

American National Standards Institute (**1996**). ANSI S3.6 "American National Standard specification for audiometers," (American National Standards Institute, New York).

Boll, S. F. (**1979**). "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. **27**, 113–120.

Cheng, Y. M., and O'Shaughnessy, D. (**1991**). "Speech enhancement based conceptually on auditory evidence," IEEE Trans. Signal Process. **39**, 1943–1954.

Daudet, L., and Torresani, B. (**2002**). "Hybrid representations for audiophonic signal encoding," Signal Process. **82**, 1595–1617.

Elliott, S. J., and Nelson, P. A. (**1993**). "Active noise control," IEEE Signal Process. Mag. **10**, 12–35.

Ephraim, Y., and Van Trees, H. L. (**1995**). "A signal subspace approach for speech enhancement," IEEE Trans. Acoust., Speech, Signal Process. **3**, 251–266.

Fletcher, H., and Galt, R. (**1950**). "The perception of speech and its relation to telephony," J. Acoust. Soc. Am. **22**, 89–151.

French, N., and Steinberg, J. (**1947**). "Factors governing the intelligibility of speech," J. Acoust. Soc. Am. **19**, 90–114.

Gordon-Salant, S. (**1986**). "Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing," J. Acoust. Soc. Am. **80**, 1599–1607.

Hazan, V., and Simpson, A. (**1998**). "The effect of cue-enhancement on the intelligibility of nonsense word and sentence materials presented in noise," Speech Commun. **24**, 211–226.

Kennedy, E., Levitt, H., Neuman, A. C., and Weiss, M. (**1998**). "Consonant-vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners," J. Acoust. Soc. Am. **103**, 1098–1114.

Kryter, K. (**1962**). "Methods for the calculation and use of the articulation index," J. Acoust. Soc. Am. **34**, 1689–1697.

Li, M., McAllister, H., Black, N., and De Perez, T. (**2001**). "Perceptual time-frequency subtraction algorithm for noise reduction in hearing aids," IEEE Trans. Biomed. Eng. **48**, 979–988.

Lim, J., and Oppenheim, A. (**1979**). "Enhancement and bandwidth compression of noisy speech," Proc. IEEE **67**, 1586–1604.

Mackersie, C., Neuman, A., and Levitt, H. (**1999**). "A comparison of response time and word recognition measures using a word-monitoring and closed-set identification task," Ear Hear. **20**, 140–148.

Niederjohn, R. J., and Grotelueschen, J. H. (**1976**). "The enhancement of speech intelligibility in high noise levels by high-pass filtering followed by rapid amplitude compression," IEEE Trans. Acoust., Speech, Signal Process. **24**, 277–282.

Rao, A., and Kumaresan, R. (**2000**). "On decomposing speech into modulated components," IEEE Trans. Speech Audio Process. **8**, 240–254.

Robinson, D., and Watson, C. (**1973**). "Psychophysical methods in modern psychoacoustics," *Foundation of Modern Auditory Theory*, edited by J. Tobias (Academic, New York).

Sauert, B., and Vary, P. (**2006**). "Near end listening enhancement: Speech intelligibility improvement in noisy environments," Proceedings International Conference on Acoustics, Speech and Signal Processing (ICASSP)

1, 493–496.
Skowronski, M. D., and Harris, J. G. (**2006**). "Applied principles of clear and Lombard speech for automated intelligibility enhancement in noisy environments," Speech Commun. **48**, 549–558.
Stevens, K. (**1998**). *Acoustic Phonetics* (MIT, Cambridge).
Strange, W., Jenkins, J. J., and Johnson, T. L. (**1983**). "Dynamic specification of coarticulated vowels," J. Acoust. Soc. Am. **74**, 695–705.
Tillman, T., and Carhart, R. (**1966**). "An expanded test for speech discrimination utilizing CVC monosyllabic words," Northwestern University Auditory Test No 6, Technical Report.
Turicchia, L., and Sarpeshkar, R. (**2005**). "A bio-inspired companding strategy for spectral enhancement," IEEE Trans. Speech Audio Process. **13**, 243–253.
Yoo, S. (2005). "Speech decomposition and enhancement," Ph.D. dissertation, University of Pittsburgh, Pittsburgh, PA.
Yoo, S., Boston, J., Durrant, J., Kovacyk, K., Karn, S., Shaiman, S., El-Jaroudi, A., and Li, C. (**2005**). "Relative energy and intelligibility of transient speech information," ICASSP **1**, 69–72.
Yu, E., and Chan, C. (**2000**). "Phase and transient modeling for harmonic +noise speech coding," Proceedings International Conference on Acoustics, Speech and Signal Processing (ICASSP) **3**, 1467–1470.
Zhao, Q., Gao, Q., and Chi, H. (**1997**). "Detection of spectral transition for speech perception based on time-frequency analysis," Proceedings International Conference on Information Communications and Signal Processing (ICICS) **97**, 522–525.
Zhu, Q., and Alwan, A. (**2000**). "On the use of variable frame rate analysis in speech recognition," Proceedings International Conference on Acoustics, Speech and Signal Processing (ICASSP) **3**, 1783–1786.

Effects of noise and distortion on speech quality judgments in normal-hearing and hearing-impaired listeners[a)]

Kathryn H. Arehart[b)]
University of Colorado, Department of Speech Language and Hearing Sciences, 409 UCB, Boulder, Colorado 80309

James M. Kates
GN ReSound and University of Colorado, Department of Speech Language and Hearing Sciences, 409 UCB, Boulder, Colorado 80309

Melinda C. Anderson
University of Colorado, Department of Speech Language and Hearing Sciences, 409 UCB, Boulder, Colorado

Lewis O. Harvey, Jr.
University of Colorado, Department of Psychology, 345 UCB, Boulder, Colorado 80309, USA

(Received 19 September 2006; revised 8 June 2007; accepted 11 June 2007)

Noise and distortion reduce speech intelligibility and quality in audio devices such as hearing aids. This study investigates the perception and prediction of sound quality by both normal-hearing and hearing-impaired subjects for conditions of noise and distortion related to those found in hearing aids. Stimuli were sentences subjected to three kinds of distortion (additive noise, peak clipping, and center clipping), with eight levels of degradation for each distortion type. The subjects performed paired comparisons for all possible pairs of 24 conditions. A one-dimensional coherence-based metric was used to analyze the quality judgments. This metric was an extension of a speech intelligibility metric presented in Kates and Arehart (2005) [J. Acoust. Soc. Am. **117**, 2224–2237] and is based on dividing the speech signal into three amplitude regions, computing the coherence for each region, and then combining the three coherence values across frequency in a calculation based on the speech intelligibility index. The one-dimensional metric accurately predicted the quality judgments of normal-hearing listeners and listeners with mild-to-moderate hearing loss, although some systematic errors were present. A multidimensional analysis indicates that several dimensions are needed to describe the factors used by subjects to judge the effects of the three distortion types. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2754061]

PACS number(s): 43.71.Gv, 43.66.Ts, 43.71.Ky, 43.66.Sr [AJO] Pages: 1150–1164

I. INTRODUCTION

Sound quality plays an important role in hearing-aid outcomes. Of the several factors that Kochkin (2005) identified as being strongly correlated to overall satisfaction by hearing-aid users, three are related to sound quality: clarity, naturalness, and richness/fidelity. The perceived quality of the signal emerging from a hearing aid is affected by noise present in the input signal (e.g., speech in a noisy background) as well as by noise and distortion within the hearing aid itself. Digital signal-processing algorithms in hearing aids are often nonlinear and generate unwanted distortion along with the desired signal modifications (e.g., Kates, 1993; Hansen, 1999; Stelmachowicz *et al.*, 1999; Souza *et al.*, 2006). Furthermore, the hearing-aid circuits and transducers generate additional nonlinear distortion that can reduce sound quality (Palmer *et al.*, 1995). This study investigates the perception and prediction of sound quality for additive noise and simple forms of distortion such as peak clipping and center clipping.

The first aim of this study is to quantify speech quality in normal-hearing (NH) and hearing-impaired (HL) listeners for a representative set of noise and distortion conditions. The signal modifications that appear at the output of a hearing aid involve both the effects of environmental noise on the speech and the effects of hearing-aid processing on the noisy speech. In a previous study investigating speech intelligibility (Kates and Arehart, 2005), three signal-degradation conditions were used: 1) stationary speech-shaped noise, 2) symmetric peak clipping, and 3) symmetric center clipping. The two clipping conditions were included because they are related to distortion mechanisms found in hearing aids. Peak clipping is related to arithmetic, amplifier, and transducer saturation. Center clipping is related to numeric underflow and to the effects of noise-suppression signal processing in reducing the intensity of low-level signal components. The same three conditions are used in the present quality study to

[a)]Portions of this work were presented at the 151st Meeting of the Acoustical Society of America, Providence, RI, June 2006.
[b)]Author to whom correspondence should be addressed. Electronic mail: kathryn.arehart@colorado.edu

facilitate comparisons of the quality results with the Kates and Arehart (2005) intelligibility results. The forms of degradation included here are not intended to duplicate "real-world" environmental noise or hearing-aid processing. Instead, as simplified and easily controlled forms of noise and distortion, they provide a starting point for analyses which may eventually lead to an understanding of the complexities of real-world listening through hearing aids.

Quality judgments by hearing-impaired listeners, like those for normal-hearing listeners (Kates and Kozma-Spytek, 1994; Tan *et al.*, 2003), show decreasing quality ratings with increasing distortion level for amplifier saturation (Palmer *et al.*, 1995), symmetric peak clipping (Crain, 1992; Kozma-Spytek *et al.*, 1996; Stelmachowicz *et al.*, 1999), and signal rectification (Lawson and Chial, 1982). Sound quality for speech processed by hearing aids has also been shown to be multidimensional in its structure (e.g., Yonovitz *et al.*, 1978; Gabrielsson and Sjögren, 1979; Punch and Beck, 1980; Gabrielsson *et al.*, 1988; Versfeld *et al.*, 1999). Several studies (Lawson and Chial, 1982; Stelmachowicz *et al.*, 1999) have found that hearing-impaired listeners are less sensitive to the differences across distortion conditions than are normal-hearing listeners, and a comparison of the normal-hearing data in Kates and Kozma-Spytek (1994) with the hearing-impaired data in Kozma-Spytek *et al.* (1996) leads to the same conclusion. Quality judgments made by hearing-impaired listeners appear to have moderate test-retest reliability (Gabrielsson *et al.*, 1988; Narendran and Humes, 2003).

The second aim of this study is to develop a metric to predict the listeners' quality judgments. A metric designed for hearing aids must meet several requirements. The metric must be applicable to systems, like hearing aids, that have frequency-dependent magnitude and phase transfer functions. The method should be applicable to speech as the test signal; much of the nonlinear signal processing in hearing aids is specifically designed for speech as the input and will respond differently to other types of excitation such as pure tones, multi-tone complexes, or broadband noise. The method must be accurate for the many different kinds of noise and distortion that can occur in hearing aids. Finally, the method must also be able to represent hearing loss and its effects on quality judgments.

The metric developed by Kates and Arehart (2005) satisfies the above requirements for predicting speech intelligibility. A goal of the present study is to determine if a similar approach can be effective in predicting speech quality. The Kates and Arehart (2005) procedure combines coherence with the speech-intelligibility index (SII) (ANSI S3.5, 1997). Developed for predicting speech intelligibility, the SII measures the signal-to-noise ratio (SNR) on a dB scale in each frequency band. The calculation procedure then adjusts for auditory threshold and for frequency-domain masking effects and sums the weighted SNR across frequency to produce the intelligibility estimate. While the SII has not been generalized to predict speech quality in hearing-impaired listeners, Eisenberg *et al.* (1998) found that there was a high degree of correlation between clarity judgments, intelligibility ratings, and SII values for both normal-hearing and hearing-impaired listeners for low-pass and high-pass filter conditions. Preminger and Van Tasell (1995) have also shown that quality judgments are correlated with intelligibility scores, so a quality metric using the same analytical framework as the previous intelligibility metric would be expected to have a similar degree of success. The Kates and Arehart (2005) intelligibility metric can be interpreted as estimating the audibility of speech sounds occurring within different intensity regions. Similarly, a quality metric based on the same approach would estimate the audibility of the noise or distortion products found at different intensities of speech.

Objective quality metrics have been developed for nonhearing-aid applications (CCITT, 1986, 1987; Quakenbush *et al.*, 1988; Czerwinski *et al.*, 2001a, b; Tan *et al.*, 2003, 2004; Geddes and Lee, 2003). Digital audio technologies (e.g., telephones, cellular telephones, digital distribution of music and speech) often require that the audio signals be digitally coded and decoded at low data rates due to limitations in the capacity of storage devices and/or transmissions systems (Kondoz, 2004). The sound quality standards for wideband audio systems (ITU BS.1387-1, 2001) and narrowband speech coding and transmission systems (ITU P.862, 2001) are derived using models of the auditory periphery. Both the wideband (Thiede *et al.*, 2000) and narrowband (Beerends *et al.*, 2002) perceptual models include auditory frequency analysis and masking effects, and incorporate excitation pattern or loudness models. The sound quality metric is based on the difference between the perceptual model outputs for the original and processed signals. Any change in the processed system output, including the amplification provided by a hearing aid to overcome a hearing loss, causes a difference in the model outputs and would be interpreted as a reduction in sound quality. Furthermore, neither ITU standard includes hearing loss. The ITU standards, in their present form, would therefore be difficult to apply to hearing aids.

Another approach to estimating sound quality is to compute the coherence between the clean input and the degraded output signals (ANSI S3.42, 1992; Kates, 1992). The coherence indicates the amount of noise or distortion that is present, independent of linear system modifications such as the hearing-aid gain-*vs*-frequency characteristic. Several investigators have reported correlations between the physical coherence measurements and subjective ratings of hearing aid distortions. Palmer *et al.* (1995), in comparing two hearing-aid amplifiers, found a good correlation between listeners' quality judgments and the coherence measured at 3 kHz as the distortion in the amplified signal increased with increased amplification. Stelmachowicz *et al.* (1999) found that there was a high degree of correlation between clarity judgments and the signal-to-distortion ratio computed from the coherence for peak-clipping distortion for normal hearing subjects, but that the correlation between clarity judgments and signal-to-distortion ratio was lower for the hearing-impaired subjects. Kates and Kozma-Spytek (1994) found that a frequency-weighted coherence measurement could accurately model the effects of peak-clipping distortion on sound quality for normal-hearing listeners, and Kozma-Spytek *et al.* (1996) showed that the same approach was

effective for hearing-impaired listeners, although the frequency-dependent weights appear to depend on the degree of hearing loss.

The coherence analysis results cited above all deal with peak-clipping distortion or amplifier saturation. However, Tan *et al.* (2004) have shown that a metric based on coherence can accurately predict sound quality for normal-hearing listeners exposed to a variety of signal degradation conditions, including hard peak clipping, soft peak clipping, center clipping, and instantaneous dynamic-range compression and expansion. In their approach, the clean input and degraded output signals are both passed through auditory filter banks. The signals are divided into 30-ms-long segments, and the coherence between the output and input segments is computed in the time domain in each auditory frequency band. The frequency bands are then weighted, with all output signal levels within 40 dB of the most-intense output band receiving a gain of 1 and output levels between 40 and 80 dB below the peak level receiving a weight that varies linearly with level in dB from 1 down to 0. The weighted coherence values are summed across frequency and normalized by the sum of the weights. The normalized coherence values are then averaged over the segments constituting the test signal to produce the quality estimate.

Both the Tan *et al.* (2004) quality metric and Kates and Arehart (2005) intelligibility metric use coherence to measure the changes in the signal caused by noise and distortion. However, the Tan *et al.* (2004) metric was designed to test digital communications over cellular telephones, and was not intended to deal with impaired hearing and hearing aids. The auditory threshold is not incorporated into their calculation, and hearing-aid amplification will change the shape of the spectrum and hence the weights used in combining the frequency bands. In contrast, the Kates and Arehart (2005) approach explicitly includes the impaired auditory threshold in the SII calculations. The objective of this paper is to determine if the intelligibility approach developed in Kates and Arehart (2005) can be extended to quality judgments.

The remainder of this paper begins with a presentation of the experimental design, including descriptions of the subjects, test materials, the noise and distortion conditions, and the paired-comparison test procedure. The subject preference results are then presented, followed by the model results using the coherence SII approach and a comparison of the quality model with the previous intelligibility model. The results are followed by a discussion of the validity of using coherence to estimate speech quality and presentation of a multidimensional unfolding analysis of the paired-comparison data used to determine the structure of the underlying perceptual space. The multi-dimensional scaling results suggest that several perceptual dimensions are needed to describe the factors used by the subjects in judging the effects of the different signal degradation mechanisms.

II. METHODS

A. Listeners

Participants in this study included 14 listeners with normal hearing (mean age=38 years; age range=23–58 years) and 18 listeners with hearing loss of presumed cochlear origin (mean age=55 years; age range=20–76 years). Normal hearing is defined here as thresholds of 20 dB hearing level (ANSI S3.6, 1989) or better at octave frequencies from 250 to 8000 Hz, inclusive, and otoacoustic emissions and admittance testing that were consistent with normal hearing. Listeners underwent an audiometric evaluation during their initial visit. All of the listeners with hearing loss demonstrated test results that were consistent with cochlear impairment: normal tympanometry, absence of excessive reflex decay, absence of air-bone gap exceeding 10 dB at any frequency, and absence of otoacoustic emissions in regions of threshold loss. Table I provides a summary of the audiometric thresholds of the test ear of the listeners with hearing loss. The severity of hearing loss ranged from mild to severe. All participants were recruited from the Boulder/Denver metropolitan area and were native speakers of American English.

All listeners were tested monaurally and individually in a double-walled sound-treated booth. For normal-hearing listeners, the left ear was selected as the test ear. For the listeners with hearing loss, the test ear was chosen based on the ear with a threshold configuration allowing the best digital filter design for linear amplification (see below). Listeners were compensated $10/h for their participation.

B. Test materials and degradation conditions

The stimuli were two sets of concatenated sentences from the hearing-in-noise-test (HINT) (Nilsson *et al.*, 1994), with one concatenated two-sentence set spoken by a male talker and another two-sentence set spoken by a female talker. The male-talker sentences were taken from the commercially available HINT materials, while the female-talker sentences were recorded by Nilsson (2005) under similar conditions. The sentences were digitized at a 44.1 kHz sampling rate and down sampled to 22.05 kHz to reduce computation time.

The sentences were subjected to three types of degradation: symmetric peak-clipping distortion, symmetric center-clipping distortion, and additive stationary speech-shaped noise. The two forms of clipping were chosen as examples of memoryless nonlinearities that have been used in previous investigations (Licklider, 1946). As noted above, peak clipping is related to arithmetic, amplifier, and receiver saturation in a hearing aid and center clipping is related to noise-suppression systems that reduce the amplitude of low-level portions of the signal.

Previous studies that have compared intelligibility results with subjective judgments of quality (Preminger and van Tasell, 1995; Eisenberg *et al.*, 1998) have found that at low levels of intelligibility, quality and intelligibility are highly correlated. This correlation decreases at higher levels of intelligibility, however, and quality still continues to vary as noise and/or distortion are manipulated. Levels of degradation were chosen to give speech intelligibility of 75% or better in normal-hearing listeners, based on the results of recent intelligibility experiments (Kates and Arehart, 2005). These levels of degradation corresponded to speech intelligibility of approximately 50% or better in the hearing-impaired

TABLE I. Auditory thresholds for the test ear of listeners with hearing loss. NR=No Response at the limits of the testing equipment.

			Frequency						
Listener	Age	Ear	250 Hz	500 Hz	1 KHz	2 KHz	4 KHz	6 KHz	8 KHz
HL_1	43	R	25	30	50	60	75	60	65
HL_2	55	L	10	10	15	45	85	80	70
HL_3	71	R	5	10	10	15	50	65	60
HL_4	58	R	5	15	15	15	60	55	55
HL_5	66	L	15	30	30	40	35	50	35
HL_6	20	L	5	5	10	10	80	80	75
HL_7	55	L	15	10	10	15	40	85	75
HL_8	58	L	10	20	40	30	45	65	70
HL_9	26	R	30	25	30	35	40	55	55
HL_10	59	R	25	30	40	35	25	25	45
HL_11	46	L	0	0	5	10	75	70	80
HL_12	76	R	20	25	45	85	105	100	90
HL_13	56	L	15	15	10	50	85	80	60
HL_14	31	L	15	15	20	30	50	55	40
HL_15	73	L	20	25	30	35	60	60	NR
HL_16	64	L	50	50	60	60	60	55	65
HL_17	21	R	25	40	35	50	70	90	NR
HL_18	64	L	15	15	25	30	35	50	55

listeners included in Kates and Arehart (2005). The amounts of degradation were chosen to include both the higher amounts of noise and distortion where intelligibility could be a factor in quality as well as reduced levels of degradation where intelligibility is high but quality differences can still be observed.

Each two-sentence set was subjected to 24 noise and distortion conditions. Specifically, each two-sentence set was combined with eight levels each of additive noise, symmetric peak-clipping distortion, or symmetric center-clipping distortion. The distortion levels were as follows: peak clipping with the clipping threshold set to $\{0.001, 40, 60, 80, 90, 95, 98, 100\}$ percent of the cumulative level histogram of each sentence; center clipping with the clipping threshold set to $\{80, 75, 70, 65, 60, 50, 30, 0\}$ percent of the cumulative level histogram; and additive noise with the SNR set to $\{4, 6, 8, 10, 15, 20, 30, 100\}$ dB. Note that the effects of peak clipping are reduced as the clipping threshold is increased, while the effects of center clipping are increased as the clipping threshold is increased. In the noise condition denoted by 100 dB SNR, no noise was added to the stimulus; however, there may still be some residual noise from the audio recording process.

The peak-clipping and center-clipping distortion thresholds were set using the magnitude level histogram for each sentence. The silent intervals at the beginning and end of each sentence were discarded, and the cumulative level distribution of the absolute values of the signal samples was then computed for the sentence. The clipping threshold was then set as a percent of the cumulative level histogram for the sentence. For symmetric peak clipping, the clipping operation is given by

$$y(n) = \begin{cases} c, & x(n) > c \\ x(n), & -c \leq x(n) \leq c \\ -c, & x(n) < -c \end{cases}, \tag{1}$$

where $x(n)$ is the speech input, $y(n)$ is the distorted output, and c is the clipping threshold. The symmetric center-clipping operation is given by:

$$y(n) = \begin{cases} x(n), & x(n) > c \\ 0, & -c \leq x(n) \leq c \\ x(n), & x(n) < -c \end{cases}. \tag{2}$$

The additive noise was extracted from the opposite channel of the HINT compact disk, and matched the long-term spectrum of the male talker. The SNR was determined by computing the root-mean-squared (rms) power of each sentence, ignoring the silent intervals at the beginning and end of the sentence, and adjusting the noise power over the same interval to give the desired SNR.

The two-sentence sets that were distorted with peak clipping and with center clipping were readjusted to give an average level of 65 dB sound pressure level (SPL) for normal-hearing listeners. Similarly, for the two-sentence sets subjected to additive noise, the combined signal-plus-noise power over the duration of the two-sentence set was also adjusted to give a presentation level of 65 dB SPL for normal-hearing listeners. Since the intensity of the noisy speech was kept at a constant level, the speech intensity was incrementally reduced as the noise level was increased. In the worst case, to the SNR of 4 dB, the speech level was reduced by about 1.5 dB relative to that of the speech without any noise. This small shift in the intensity of the speech would have only a minimal effect on intelligibility, and was deemed less of a confounding effect on quality than changes

in the overall signal amplitude, especially when comparing stimuli subjected to different degradation conditions. It is also possible that the larger amounts of peak clipping could change the loudness of the level-normalized stimuli compared to the loudness of the clean speech. However, informal listening tests indicated that no differences in loudness were apparent when comparing different levels of distortion.

The 65 dB SPL stimuli were amplified for each listener with hearing loss using the NAL-R prescriptive formula based on individual thresholds (Byrne and Dillon, 1986). By way of example, a listener with a flat 50 dB HL hearing loss would receive approximately 22 dB of gain at 2000 Hz. The amplification was implemented through digital filtering (via a 128-point linear-phase Finite Impulse Response (FIR) filter) prior to the experiment. Stimuli were customized for each individual listener and were stored on a personal computer prior to presentation. A linear amplification scheme was chosen over a compression scheme to avoid introducing additional distortion and processing artifacts that could not be controlled in the experimental design.

For listener presentation, the digitally stored speech stimuli went through a digital-to-analog converter (TDT RX6), an attenuator (TDT PA5) and a headphone buffer amplifier (TDT HB7). Finally, the stimuli were presented monaurally to the listeners' test ear through a Sennheiser HD 25-1 earphone.

C. Procedure

Listeners participated in three 1 h sessions. During each session, listeners were presented with three blocks of 72 paired-comparison trials. The first block in the first session was a practice block and included some of the stimuli that were included in the formal testing. No feedback was provided during practice or formal testing.

On each trial, listeners compared two presentations of the same concatenated two-sentence set spoken by the same talker, with each presentation representing one of the 24 processing conditions. For each comparison, the subject indicated which of the two stimuli had the better sound quality. The instructions given to the subject are described in the Appendix. Selections were made using a customized computer interface in which listeners used a point-and-click method to record and verify their preferences. The timing of presentation was controlled by the subject.

Pairwise quality judgments were obtained from each subject for all possible comparisons of the degraded signals. The 576 possible combinations can be visualized with Tables III and IV. Ignoring for the moment the specific numbers in each cell, half the cells in each matrix are shaded gray while the other half are unshaded. For half of the listeners, the comparisons in the shaded cells were made for the two-sentence sets spoken by the male talker and the comparisons in the unshaded cells were made for the sentence sets spoken by the female talker. For the other half of the listeners, the gender assignment to the matrix cells was reversed. As indicated by the cells with a dash (-) along the central diagonal of Tables III and IV, each listener was presented 24 trials in which the stimuli compared were the same, with 12 spoken by the male talker and 12 spoken by the female talker.

For each listener, the order of the 576 pairs was randomized across talker gender and across distortion conditions. The 576 pairs were then played out in this randomized order in blocks of 72 trials. Thus, each block included both gender talkers and multiple distortion-level conditions. For the shaded conditions in Tables III and IV, the stimulus condition presented in the top row was presented first in a paired-comparison trial and for the unshaded cells, the stimulus condition on the left hand margin was presented first in the paired comparison trial. For the other half of the listeners, the gender assignment to the matrix cells was reversed.[1,2]

III. RESULTS

Pairwise comparisons can be analyzed within a degradation type and across degradation types. Analysis *within* a degradation type considers only paired comparisons made between one level of a degradation type and another level of the same signal degradation (e.g., noise compared against noise; peak clipping compared to peak clipping; and center clipping compared to center clipping). Analysis *across* degradation types includes all paired comparison trials (e.g., noise compared to noise, to peak clipping and to center clipping).

One objective of this study is to model and predict speech quality for arbitrary degradation mechanisms that might be present in a hearing aid. The *across*-degradation comparisons are designed to provide the data set needed for this analysis. Another approach to analyzing the comparisons is to focus on a single degradation mechanism. If the two subject groups responded in different ways to any single mechanism of signal degradation, then one would expect to see differences in the preference scores computed *within* that specific degradation type.

The aggregate pairwise quality preferences for the 14 subjects with normal hearing and for the 18 subjects with hearing loss are shown *within* a distortion type in Table II and *across* distortion types in Tables III and IV. Table II shows the preferences for comparisons made between a particular distortion-level condition and all other levels of that same distortion type for the 14 normal-hearing subjects (left side of table) and for the group of 18 hearing-impaired subjects (right side of table). Data are shown for both the male talker and for the female talker (in parentheses). The aggregate pairwise quality preferences for paired-comparison trials combined *across* distortion types are shown in Table III for the 14 subjects with normal hearing and in Table IV for the 18 subjects with hearing loss. The preference values are combined across the two talker genders. The number in each cell is the number of times the distortion-level condition listed by the column heading was preferred to the distortion-level condition listed by the row heading, summed over the total number of responses made for each distortion-level comparison in each subject group (number of subjects $\times$ two repetitions of each paired comparison (one for each gender talker)). For example, when compared to the 80% peak clipping threshold, the 20 dB SNR condition in Table III was

TABLE II. Preferences for comparison *within* each distortion type (between a particular distortion-level condition and all other levels of that same degradation type for the NH group (left) and the HL group (right) and for the male talker and female talker (parentheses).

		SNR, dB (Normal Hearing Group)									SNR, dB (Hearing-Impaired Group)							
		4	6	8	10	15	20	30	100		4	6	8	10	15	20	30	100
SNR, dB	4	*	11 (10)	11 (12)	13 (13)	14 (13)	14 (14)	14 (13)	14 (14)	4	*	12 (13)	15 (11)	16 (17)	18 (18)	16 (18)	17 (18)	18 (17)
	6	3 (4)	*	10 (11)	12 (12)	13 (14)	13 (13)	14 (14)	14 (14)	6	6 (5)	*	13 (12)	13 (13)	18 (16)	17 (18)	18 (17)	18 (18)
	8	3 (2)	4 (3)	*	11 (10)	12 (13)	14 (14)	14 (14)	14 (14)	8	3 (7)	5 (6)	*	13 (12)	14 (15)	15 (18)	17 (17)	18 (17)
	10	1 (1)	2 (2)	3 (4)	*	11 (12)	13 (13)	14 (13)	14 (13)	10	2 (1)	5 (5)	5 (6)	*	14 (12)	15 (16)	17 (17)	18 (18)
	15	0 (1)	1 (0)	2 (1)	3 (2)	*	10 (10)	12 (14)	14 (14)	15	0 (0)	0 (2)	4 (3)	4 (6)	*	12 (12)	17 (16)	17 (16)
	20	0 (0)	1 (1)	0 (0)	1 (1)	4 (4)	*	10 (12)	14 (14)	20	2 (0)	1 (0)	3 (0)	3 (2)	6 (6)	*	15 (14)	16 (17)
	30	0 (1)	0 (0)	0 (0)	0 (1)	2 (0)	4 (2)	*	11 (13)	30	1 (0)	0 (1)	1 (1)	1 (1)	1 (2)	3 (4)	*	13 (12)
	100	0 (0)	0 (0)	0 (0)	0 (1)	0 (0)	0 (0)	3 (1)	*	100	0 (1)	0 (0)	0 (1)	0 (0)	2 (2)	2 (1)	5 (6)	*
	Pref	0.07	0.19	0.27	0.41	0.57	0.69	0.83	0.97	Pref	0.11	0.18	0.33	0.40	0.57	0.63	0.84	0.94
	Score	(0.09)	(0.16)	(0.29)	(0.41)	(0.57)	(0.67)	(0.83)	(0.98)	Score	(0.11)	(0.21)	(0.27)	(0.40)	(0.56)	(0.69)	(0.83)	(0.91)
		Peak Clip Thr, %									Peak Clip Thr, %							
		0	40	60	80	90	95	98	100		0	40	60	80	90	95	98	100
Peak Clip Thr, %	0	*	13 (14)	14 (14)	14 (14)	14 (14)	14 (14)	14 (14)	14 (14)	0	*	16 (17)	18 (18)	18 (18)	17 (18)	18 (18)	18 (18)	18 (18)
	40	1 (0)	*	12 (13)	14 (14)	14 (14)	14 (14)	14 (14)	14 (14)	40	2 (1)	*	13 (14)	17 (18)	18 (17)	18 (18)	18 (18)	18 (18)
	60	0 (0)	2 (1)	*	14 (13)	14 (14)	14 (13)	14 (14)	14 (14)	60	0 (0)	5 (4)	*	15 (15)	17 (18)	18 (18)	16 (17)	17 (16)
	80	0 (0)	0 (0)	0 (1)	*	10 (11)	13 (13)	14 (14)	14 (14)	80	0 (0)	1 (0)	3 (3)	*	13 (15)	16 (15)	17(15)	18 (14)
	90	0 (0)	0 (0)	0 (0)	4 (3)	*	11 (14)	12 (10)	13 (11)	90	1 (0)	0 (1)	1 (0)	5 (3)	*	12 (10)	14 (12)	16 (12)
	95	0 (0)	0 (0)	0 (1)	1 (0)	3 (5)	*	8 (8)	10 (10)	95	0 (0)	0 (0)	0 (0)	2 (3)	6 (8)	*	11 (12)	11 (12)
	98	0 (0)	0 (0)	0 (0)	0 (0)	2 (4)	6 (6)	*	9 (8)	98	0 (0)	0 (0)	2 (1)	1 (3)	4 (6)	7 (6)	*	12 (10)
	100	0 (0)	0 (0)	0 (0)	0 (0)	1 (3)	4 (4)	5 (6)	*	100	0 (0)	0 (0)	1 (2)	0 (4)	2 (6)	7 (9)	6 (8)	*
	Pref	0.01	0.15	0.27	0.48	0.59	0.78	0.83	0.90	Pref	0.02	0.17	0.30	0.46	0.61	0.76	0.79	0.87
	Score	(0.00)	(0.15)	(0.30)	(0.45)	(0.66)	(0.76)	(0.82)	(0.87)	Score	(0.01)	(0.17)	(0.30)	(0.51)	(0.70)	(0.75)	(0.79)	(0.77)
		Ctr. Clip Thr., %									Ctr. Clip Thr.,%							
		80	75	70	65	60	50	30	0		80	75	70	65	60	50	30	0
Ctr. Clip Thr., %	80	*	12 (9)	12 (11)	14 (14)	13 (14)	14 (14)	14 (14)	14 (14)	80	*	16 (14)	18 (15)	17 (18)	18 (16)	18 (16)	18 (18)	17 (18)
	75	2 (5)	*	9 (11)	13 (12)	12 (11)	14 (14)	14 (14)	14 (14)	75	2 (4)	*	13 (8)	14 (11)	14 (16)	18 (17)	17 (17)	18 (18)
	70	2 (3)	5 (3)	*	7 (12)	9 (11)	14 (14)	14 (14)	14 (14)	70	0 (3)	5 (10)	*	12 (7)	12 (10)	17 (14)	17 (18)	17 (17)
	65	0 (0)	1 (2)	7 (2)	*	13 (11)	14 (11)	14 (14)	14 (14)	65	1 (0)	4 (7)	6 (11)	*	11 (11)	16 (11)	18 (17)	17 (17)
	60	1 (0)	2 (3)	5 (3)	1 (3)		13 (9)	14 (14)	14 (14)	60	0 (2)	4 (2)	6 (8)	7 (7)	*	14 (12)	16 (13)	17 (16)
	50	0 (0)	0 (90)	0 (0)	0 (3)	1 (5)	*	14 (14)	14 (14)	50	0 (2)	0 (1)	1 (4)	2 (7)	4 (6)	*	17 (12)	16 (16)
	30	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	*	11 (14)	30	0 (0)	1 (1)	1 (0)	0 (1)	2 (5)	1 (6)	*	6 (11)
	0	0 (0)	0 (0)	0 (0)	0 (0)	1 (0)	0 (0)	3 (0)	*	0	1 (0)	0 (0)	1 (1)	1 (1)	1 (2)	2 (2)	12 (7)	*
	Pref	0.05	0.20	0.34	0.36	0.50	0.70	0.89	0.96	Pref	0.03	0.24	0.37	0.42	0.49	0.68	0.91	0.86
	Score	(0.08)	(0.17)	(0.28)	(0.45)	(0.53)	(0.63)	(0.86)	(1.0)	Score	(0.09)	(0.28)	(0.37)	(0.41)	(0.52)	(0.62)	(0.81)	(0.90)

preferred 16 times out of 28 by the group of 14 normal-hearing subjects.

In most speech quality metrics, the listener preferences are reduced to a one-dimensional scale giving the mean opinion score (MOS). The MOS facilitates comparisons both within and across different forms of distortion. To summarize the subject quality judgments in the present study, the pairwise comparisons were reduced to a preference score (David, 1963; Rabiner *et al.*, 1969). The preference score ranges from 0 (never chosen) to 1 (always chosen). The preference scores were calculated in two ways. First, preference scores were calculated *within* each distortion condition by summing the number of times each distortion-level condition was preferred to the other seven levels of the same distortion type and then dividing by the total number of trials in which that distortion-level condition was compared to all other levels of the same distortion type. Second, preference scores were calculated *across* the three degradation types by summing the number of times each distortion-level condition was preferred to all of the other 23 conditions, and then dividing by the total number of comparisons for that distortion-level condition. The total number of comparisons for a distortion-level condition does not include the comparison of the condition against itself. The number of comparisons for the preference score calculations *within* each distortion type was 196 for the normal-hearing (NH) group (14 subjects $\times 7$ distortion-level conditions $\times 2$ repetitions) and 252 for the hearing loss (HL) group (18 subjects $\times 7$ distortion-level conditions $\times 2$ repetitions). The preference score calculations *across* distortion type were based on 644 comparisons for the NH group (14 subjects $\times 23$ distortion-level conditions $\times 2$ repetitions) and 828 comparisons for the HL group (18 subjects $\times 23$ distortion-level conditions $\times 2$ repetitions). The preference scores are listed in the bottom rows of Tables II–IV. By way of example, the preference score *across* distortion type for the normal-hearing subjects for the 4 dB SNR condition is 0.295, which is the column total (190) divided by the total number of comparisons (644).

Because the levels of distortion were not equivalent for all three types of distortion (e.g., dB SNR *vs* percent distortion), preference scores (with an arcsin transformation) for each of the three types of distortion for the *within*-distortion

TABLE III. Preferences for comparisons made *across* distortion type for the NH group. Pairwise quality preferences were combined over the male-male and female-female comparisons and summed over the 14 NH subjects. The number in each cell is the number of times the distortion-level condition listed by the column heading was preferred to the distortion-level condition listed by the row heading, summed over the total number of responses made for each distortion-level comparison in each subject group (number of subjects x two repetitions of each paired comparison (one for each gender talker)). The shading indicates how stimuli of different gender talkers were assigned to the matrix. For half of the listeners, the comparisons in the shaded cells were made for the two-sentence sets spoken by the male talker and the comparisons in the unshaded cells were made for the sentence sets spoken by the female talker. For the other half of the listeners, the gender assignment to the matrix cells was reversed.

		SNR dB								Peak Clip Threshold, %								Center Clip Threshold, %							
		4	6	8	10	15	20	30	100	0	40	60	80	90	95	98	100	80	75	70	65	60	50	30	0
SNR, dB	4	…	21	23	26	27	28	27	28	0	3	14	27	28	28	27	28	0	2	12	10	18	21	28	28
	6	7	…	21	24	27	26	28	28	0	5	15	25	28	27	27	27	0	2	3	9	13	19	27	27
	8	5	7	…	21	25	28	28	28	0	0	13	27	26	28	28	27	0	2	4	8	7	17	25	28
	10	2	4	7	…	23	26	27	27	0	3	13	21	23	27	26	28	2	0	1	5	4	17	27	28
	15	1	1	3	5	…	20	26	28	0	0	7	17	24	23	27	26	1	0	0	0	2	8	23	27
	20	0	2	0	2	8	…	22	28	0	0	1	12	18	24	25	27	0	0	1	0	1	6	19	26
	30	1	0	0	1	2	6	…	24	0	0	0	5	9	14	19	22	0	0	0	0	0	1	7	21
	100	0	0	0	1	0	0	4	…	1	0	1	1	1	6	7	16	0	1	0	1	0	1	7	16
Peak Clip Thr, %	0	28	28	28	28	28	28	28	27	…	27	28	28	28	28	28	28	20	25	27	28	28	28	28	28
	40	25	28	28	25	28	28	28	28	1	…	25	28	28	28	28	28	3	13	18	22	25	24	28	28
	60	14	13	15	15	21	27	28	27	0	3	…	27	28	27	28	28	1	4	3	12	17	19	26	28
	80	1	3	1	7	11	16	23	27	0	0	1	…	21	27	28	28	0	0	0	0	3	7	20	26
	90	0	0	2	5	4	10	19	27	0	0	0	7	…	20	22	24	0	0	1	0	2	4	15	25
	95	0	1	0	1	5	4	14	22	0	0	1	1	8	…	16	20	0	0	0	0	0	0	7	24
	98	1	1	0	2	1	3	9	21	0	0	0	0	6	12	…	17	0	0	0	0	0	1	3	17
	100	0	1	1	0	2	1	6	12	0	0	0	0	4	8	7	…	0	0	0	0	0	0	7	13
Ctr. Clip. Thr., %	80	28	28	28	26	27	28	28	28	8	25	27	28	28	28	28	28	…	21	23	28	27	28	28	28
	75	26	26	26	28	28	28	28	27	3	15	24	28	28	28	28	28	7	…	20	25	23	28	28	28
	70	16	25	24	27	28	27	28	28	1	10	25	28	27	28	28	28	5	8	…	19	20	28	28	28
	65	18	19	20	23	28	28	28	27	0	6	16	28	28	28	28	28	0	3	9	…	24	25	28	28
	60	10	15	21	24	26	27	28	28	0	3	11	25	26	28	28	28	1	5	8	4	…	22	28	27
	50	7	9	11	11	20	22	27	27	0	4	9	21	24	28	27	28	0	0	0	3	6	…	28	28
	30	0	1	3	1	5	9	21	21	0	0	2	8	13	21	25	21	0	0	0	0	0	0	…	25
	0	0	1	0	0	1	2	7	12	0	0	0	2	3	4	11	15	0	0	0	0	1	0	3	…
Pref. Score		0.29	0.36	0.41	0.47	0.58	0.66	0.80	0.90	0.02	0.16	0.36	0.61	0.71	0.81	0.85	0.90	0.06	0.13	0.20	0.27	0.34	0.47	0.73	0.90

analysis were subjected to separate repeated measures analyses of variance (ANOVA). The analyses considered whether significant differences existed in preference scores between different levels of distortion (within-subject factor of level), between the male and female talkers (within-subject factor of gender), and between the NH group and the HL group (between-group factor of group). The analyses showed no significant differences in preference scores between the male and female talkers and between the NH group and the HL group. The analyses revealed significant effects for levels of distortion for noise ($F_{(7,210)}=314.61$, $p<0.01$), for peak clipping ($F_{(7,210)}=464.68$, $p<0.01$) and for center clipping ($F_{(7,210)}=318.80$, $p<0.01$).

Figure 1 shows the average preference scores for the *within*-distortion comparisons for the (HL) group and for the HL group. Because significant differences were not observed for talker gender, the data shown in Fig. 1 are averaged across the male and female talkers. The left panel shows the preference scores for the comparisons involving just the different SNRs, the center panel shows the preference scores for those comparisons involving just the different peak-clipping thresholds, and the right panel shows the preference scores for those comparisons involving just the different center-clipping thresholds. All of the data show monotonically increasing quality scores with decreasing amounts of noise or distortion. (Recall that the deleterious effects of peak clipping are greatest for small thresholds of peak clipping whereas the deleterious effects of center clipping are greatest for large clipping thresholds). Consistent with the lack of significant differences between the NH group and the HL group, the shapes of the monotonic functions are quite similar for both subject groups. The two subject groups appear to respond in similar ways to each of the mechanisms of signal degradation.

The preference scores calculated *across* conditions are plotted in Fig. 2. Data are shown for both the NH group and the HL group as a function of a) SNR for additive stationary speech-shaped noise (left panel), b) clipping threshold for peak clipping (center panel), and c) clipping threshold for center clipping (right panel). Similar to the *within*-distortion preference scores shown in Fig. 1, the *across*-distortion type preference scores show monotonically increasing quality scores with decreasing amounts of noise or distortion.

Preference scores (with an arcsin transformation) for each of the three types of distortion for the *across*-distortion analysis were also subjected to separate repeated measures analyses of variance (ANOVA). The analyses considered whether significant differences existed in preference scores

TABLE IV. As Table III, but for the 18 subjects in the HL group.

		SNR, dB								Peak Clip Threshold, %							Center Clip Threshold, %								
		4	6	8	10	15	20	30	100	0	40	60	80	90	95	98	100	80	75	70	65	60	50	30	0
SNR, dB	4	…	25	26	33	36	34	35	35	8	15	24	31	35	36	35	36	10	15	21	26	30	30	35	36
	6	11	…	25	26	34	35	35	36	4	11	17	28	33	34	35	36	10	13	20	24	28	33	36	36
	8	10	11	…	25	29	33	34	35	1	7	18	28	34	33	34	36	7	13	16	20	21	27	34	34
	10	3	10	11	…	26	31	34	36	2	5	14	25	31	34	34	35	6	10	12	20	23	26	36	36
	15	0	2	7	10	…	24	33	33	1	4	10	18	22	30	35	34	2	6	11	14	16	19	32	34
	20	2	1	3	5	12	…	29	33	1	1	8	14	17	25	28	31	1	3	6	4	9	14	24	30
	30	1	1	2	2	3	7	…	25	0	1	0	10	13	15	18	24	0	1	2	6	6	4	18	25
	100	1	0	1	0	3	3	11	…	0	0	0	5	8	11	17	19	1	2	2	0	2	5	11	19
Peak Clip Thr, %	0	28	32	35	34	35	35	36	36	…	33	36	36	35	36	36	36	31	33	35	36	36	36	36	36
	40	21	25	29	31	32	35	35	36	3	…	27	35	35	36	36	36	11	23	26	31	29	31	36	36
	60	12	19	18	22	26	28	36	36	0	9	…	30	35	36	33	33	5	14	14	18	26	27	34	35
	80	5	8	8	11	18	22	26	31	0	1	6	…	28	31	32	32	2	5	4	9	14	20	30	32
	90	1	3	2	5	14	19	23	28	1	1	1	8	…	22	26	28	1	2	5	5	6	13	23	25
	95	0	2	3	2	6	11	21	25	0	0	0	5	14	…	23	20	2	1	1	3	3	9	17	23
	98	1	1	2	2	1	8	18	19	0	0	3	4	10	13	…	22	0	1	1	4	6	8	20	17
	100	0	0	0	1	2	5	12	17	0	0	3	4	8	16	14	…	0	2	1	4	4	3	13	19
Ctr. Clip. Thr., %	80	26	26	29	30	34	35	36	35	5	25	31	34	35	34	36	36	…	30	33	35	34	34	36	35
	75	21	23	23	26	30	33	35	34	3	13	22	31	34	35	35	34	6	…	21	25	30	35	35	36
	70	15	16	20	24	25	30	34	34	1	10	22	32	31	35	35	35	3	15	…	19	22	31	35	34
	65	10	12	16	16	22	32	30	36	0	5	18	27	31	33	32	32	1	11	17	…	22	27	35	34
	60	6	8	15	13	20	27	30	34	0	7	10	22	30	33	30	32	2	6	14	14	…	26	29	33
	50	6	3	9	10	17	22	32	31	0	5	9	16	23	27	28	33	2	1	5	9	10	…	29	32
	30	1	0	2	0	4	12	18	25	0	0	2	6	13	19	16	23	0	2	1	1	7	7	…	17
	0	0	0	2	0	2	6	11	17	0	0	1	4	11	13	19	17	1	0	2	2	3	4	19	…
Pref. Score		0.22	0.28	0.35	0.40	0.52	0.64	0.78	0.85	0.04	0.19	0.34	0.55	0.68	0.77	0.81	0.85	0.13	0.25	0.33	0.40	0.47	0.57	0.79	0.84

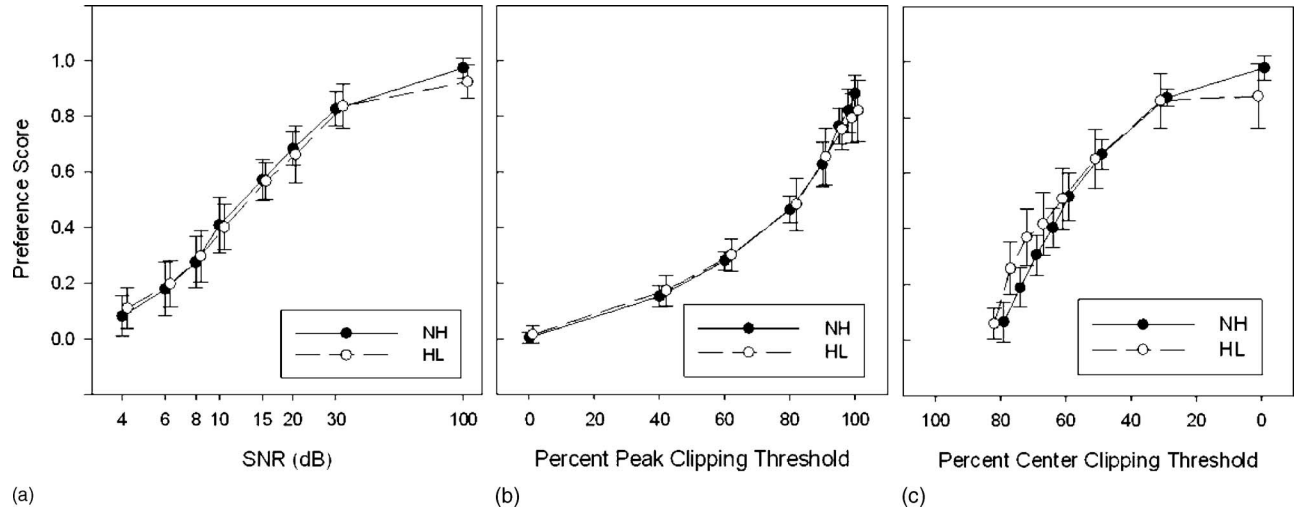


FIG. 1. Subject preference score (proportion of the times a signal degradation condition was preferred to conditions *within* the same distortion type) for listeners with normal hearing (NH) and listeners with hearing loss (HL) as a function of a) SNR for additive stationary speech-shaped noise (left panel), b) clipping threshold for peak clipping (center panel), and c) clipping threshold for center clipping (right panel). Error bars show the standard deviation of the preference scores for each listener group for each level-distortion condition. Symbols and error bars are offset slightly for clarity.

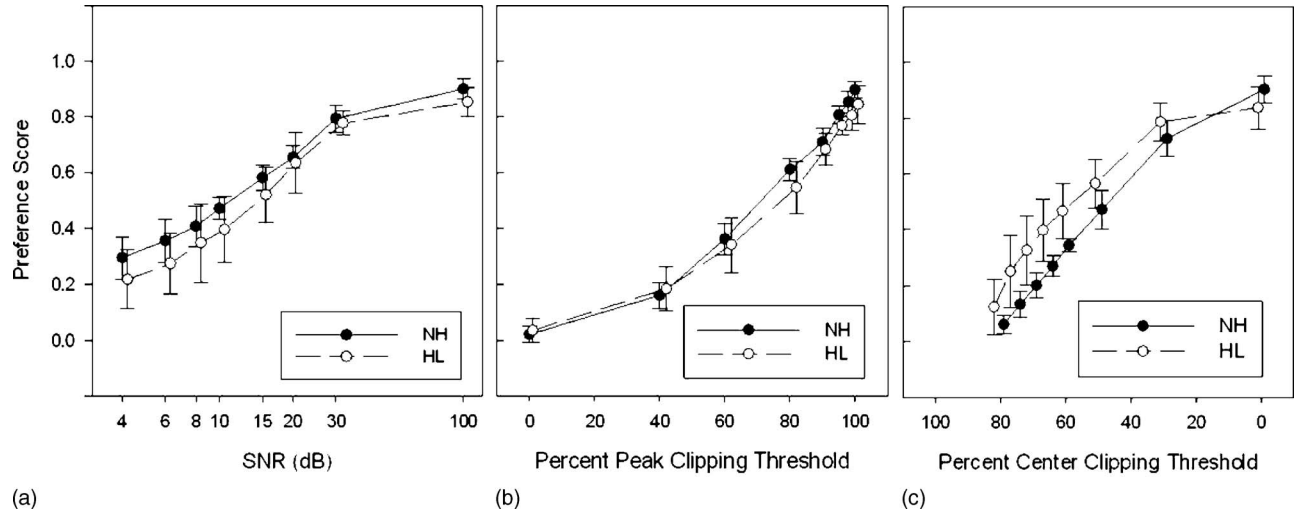


FIG. 2. Subject preference score calculated *across* distortion types (proportion of the times a signal degradation condition was preferred to all of the other conditions) for listeners with normal hearing (NH) and listeners with hearing loss (HL) as a function of a) SNR for additive stationary speech-shaped noise (left panel), b) clipping threshold for peak clipping (center panel), and c) clipping threshold for center clipping (right panel). Error bars show the standard deviation of the preference scores for each listener group for each level-distortion condition. Symbols and error bars are offset slightly for clarity.

between different levels of distortion (within-subject factor of level) and between the NH group and the HL group (between-group factor of group). However, in contrast to the *within*-distortion analysis, preference scores calculated *across* distortion types were significantly different between the two groups for noise ($F_{(1,210)}=6.58$, $p<0.05$), for peak clipping ($F_{(1,210)}=5.12$, $p<0.05$) and for center clipping ($F_{(1,210)}=20.04$, $p<0.01$). For both noise and peak clipping, the ratings by the HL group are on average lower than for the NH group. In contrast, the HL group shows higher average preference ratings for center clipping than the NH group, except for the center-clipping threshold of 0%. This latter result is consistent with the group-by-level interaction for center clipping ($F_{(1,210)}=6.81$, $p<0.001$). As evidenced by the larger standard deviations (shown by error bars in Fig. 2), the preference scores of the HL group are characterized by greater variability than the preference scores of the NH group. For all the types of distortion, the maximum scores assigned by the HL group for the no-distortion stimuli are not quite as high as those assigned by the NH group for the same stimuli, which suggests that the quality differences between the speech stimuli may be less discernible to the hearing-impaired listeners.

The analysis *within* a degradation type did not reveal significant differences between groups, suggesting that listener groups respond to a single degradation mechanism in similar ways. In contrast, significant between-group differences were evident when preference scores were analyzed *across* different distortion mechanisms. This difference suggests that the perceptual basis of the quality judgments may be different in subjects with hearing loss when judging *across* degradation mechanisms. However, it is important to note that the *across*-condition preference scores have a constraint in that a subject preference for one form of signal degradation will of necessity reduce the preference scores for all of the comparison forms of degradation. Therefore, a small preference for center clipping by the HL group would drive down the preferences for peak clipping and additive noise, which could contribute to the between-group statistical significance observed in the *across*-condition analysis.

IV. MODEL ANALYSIS

Hearing aids contain many possible sources of noise and distortion that can affect sound quality. One objective of this study is to model and predict speech quality for arbitrary degradation mechanisms that might be present in a hearing aid. To accomplish this end, a data set is required that compares quality judgments across different forms of signal degradation. The *across*-degradation analysis considered in this experiment is a simplification of the more complex signal processing environment found in a hearing aid. As such, the *across*-degradation analysis represents a starting point for the modeling and prediction of hearing-aid sound quality.

One approach to modeling the preference results is to analyze the degraded signals and use the results of the analysis to predict the subject quality judgments. This approach collapses the potentially multi-dimensional structure of the preference decisions into a single dimension indicating overall preference. Implicit in this approach are assumptions as to what may be the dominant factor(s) in forming the quality judgments.

In the one-dimensional analysis presented in this section, we assume that audibility may be an important factor in listeners' perceptual judgments of quality. The audibility of the distortion depends on the intensity of the distortion products relative to that of the speech. Additive noise and center clipping, for example, primarily affect the low-level portions of the signal. These forms of signal degradation will be most audible during low portions of the speech and will be masked during high-level portions. Peak clipping, on the other hand, only affects the high-level portions of the signal and does not modify the low-level portions. Peak clipping will therefore only be audible during the more-intense por-

tions of the signal. The coherence speech intelligibility index (CSII) approach (Kates and Arehart, 2005) gives a procedure that takes into account these audibility differences in the signal degradation effects as a function of signal intensity.

The NAL-R formula provides frequency-dependent amplification of the signal. This amplification affects the audibility of both the speech and the noise or the distortion products caused by signal clipping. The audibility of different speech components and the interference caused by noise and distortion is assumed to be a factor in the subject preferences. Thus, a different prescriptive formula could result in somewhat different preferences for the different signal degradation mechanisms.

Kates and Arehart (2005) used the three CSII values to model speech intelligibility scores for HINT sentences subjected to the same three conditions of signal degradation as were used for the quality experiments discussed in this paper. A complete description of the CSII calculation procedure is given in Kates and Arehart (2005). Each undegraded test sentence was divided in three amplitude regions: at or above the rms sentence level, 0–10 dB below the rms level, and 10–30 dB below the rms level. The signal envelope was computed using a Hamming-windowed segment size of 16 ms. The coherence comparing the degraded output to the clean input was computed as a function of frequency for the signal segments in each amplitude region. The SII for the each region was then computed for the normal-hearing listeners in each of the three level regions, with the signal-to-distortion ratio (SDR) calculated from the coherence used to replace the conventional SNR in the SII calculation. The result is coherence SII (CSII) values for low-, mid-, and high-level portions of the speech. A minimum mean-squared error fit to the proportion of sentences identified correctly in the intelligibility study was given by

$$c = -3.47 + 1.84\mathrm{CSII}_{\mathrm{Low}} + 9.99\mathrm{CSII}_{\mathrm{Mid}} + 0.0\mathrm{CSII}_{\mathrm{High}}$$

$$I_3 = \frac{1}{1+e^{-c}} \tag{3}$$

for the normal-hearing subjects. For the hearing-impaired listeners, the SII calculation used each subject's audiogram, and the signal level was adjusted for the NAL-R amplification used to compensate for each subject's hearing loss. The width of the auditory filters and the upward spread-of-masking function used for the hearing-impaired subjects were the same as for the normal-hearing subjects, as were the coefficients used to weight the CSII values.

The same CSII approach was used to model the quality data from the present study. The pair of sentences produced by the male talker was concatenated with the pair of sentences produced by the female talker. The combined group of four sentences was subjected to additive speech-shaped noise, peak-clipping distortion, or center-clipping distortion, and the CSII values computed for the low-, mid-, and high-level segments of the stimulus. A minimum mean-squared error fit to the normal-hearing subjects' quality ratings is given by

$$c = -4.56 + 2.41\mathrm{CSII}_{\mathrm{Low}} + 2.16\mathrm{CSII}_{\mathrm{Mid}} + 1.73\mathrm{CSII}_{\mathrm{High}}$$

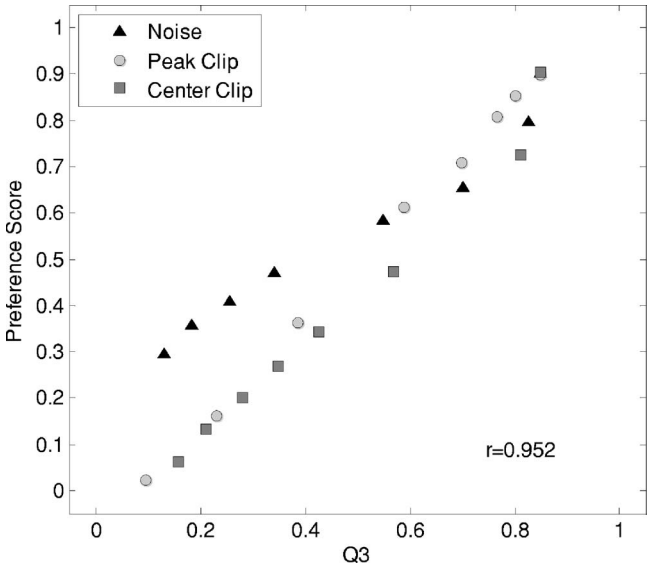


FIG. 3. Subject preference score (proportion of the times a signal degradation condition was preferred to all of the other conditions) plotted as a function of the Q_3 quality metric for normal-hearing listeners.

$$Q_3 = \frac{1}{1+e^{-c}}. \tag{4}$$

The quality model given by Eq. (4) has roughly equal weights on all three amplitude regions, as opposed to the intelligibility model of Eq. (3) where the greatest weight was on the mid-level region and the high-level region had a weight of 0. This difference in the CSII coefficients suggests that different perceptual properties of the degraded speech signal are used for the quality judgments than are used for intelligibility. The fit of the Q_3 metric to the average ratings of subjects from the NH group is presented in Fig. 3. The metric fits the peak-clipping and center-clipping data very well. For the additive speech-shaped noise, there appears to be a small bias at low quality ratings where the metric predicts lower quality ratings than found by the subjects. The metric fits the subject ratings with an overall correlation coefficient of $r=0.952$.

The metric given by Eq. (4) was also fit to the average ratings of subjects from the HL group. The same CSII calculation procedure and model parameters were used for the listeners in the HL group as were used for the listeners in the NH group, but with the HL subject's audiogram and the NAL-R amplification taken into account in computing the SII. The fit of the Q_3 metric to the average ratings of the HL group is presented in Fig. 4. The metric fits the HL group ratings with an overall correlation coefficient of $r=0.985$, which is actually better than the fit to the NH group data for which the metric was derived. The tendency of the metric to predict lower quality ratings at high levels of speech-shaped noise is also reduced in comparison with the predictions for the normal-hearing listeners.

A further question is whether the severity of the hearing loss affects the accuracy of the model prediction. This question was investigated by comparing the accuracy of the Q_3 metric calculated for individual listeners in the HL group with the degree of hearing loss. For each hearing-impaired

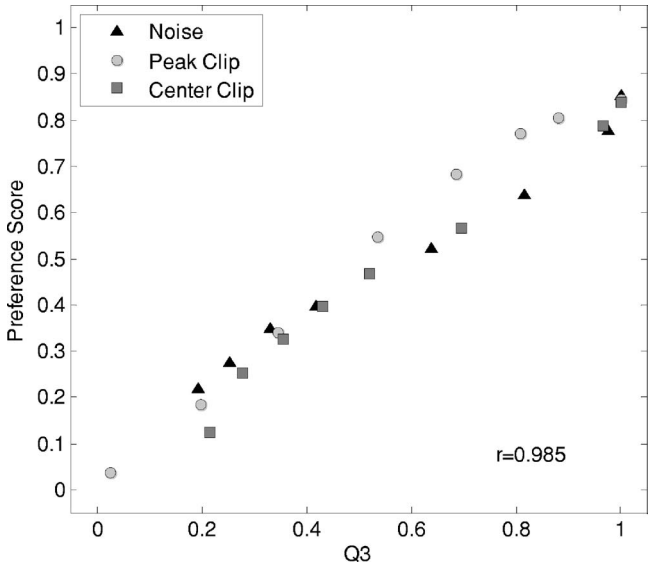


FIG. 4. Subject preference score (proportion of the times a signal degradation condition was preferred to all of the other conditions) plotted as a function of the Q_3 quality metric for hearing-impaired listeners.

listener in the experiment, the Q_3 metric was computed for each of the 24 noise and distortion conditions, taking into account the individual audiogram and the NAL-R gain compensation. The correlation coefficient comparing the Q_3 metric values to the listeners' preference scores was then calculated for the 24 degradation conditions. This correlation value is plotted in Fig. 5 for each of the 18 hearing-impaired listeners as a function of each listener's three-frequency average hearing loss (the loss in dB averaged across 0.5, 1, and 2 kHz). Each correlation coefficient value indicates how accurately the Q_3 metric models the quality judgments made by that listener.

Figure 5 shows a general loss of modeling accuracy with increasing hearing loss. The correlation coefficient between the modeling accuracy and average loss is $r=0.600$($p=0.008$). However, much of this correlation is due to the two listeners having the greatest average loss, identified as HL 12 and HL 16 in the figure. If the data points for these two listeners are removed from consideration, the correlation coefficient for the remaining 16 subjects is $r=0.174$($p=0.520$). These results suggest that the Q_3 metric is accurate for normal hearing through moderate hearing loss, but begins to lose accuracy for more severe losses.

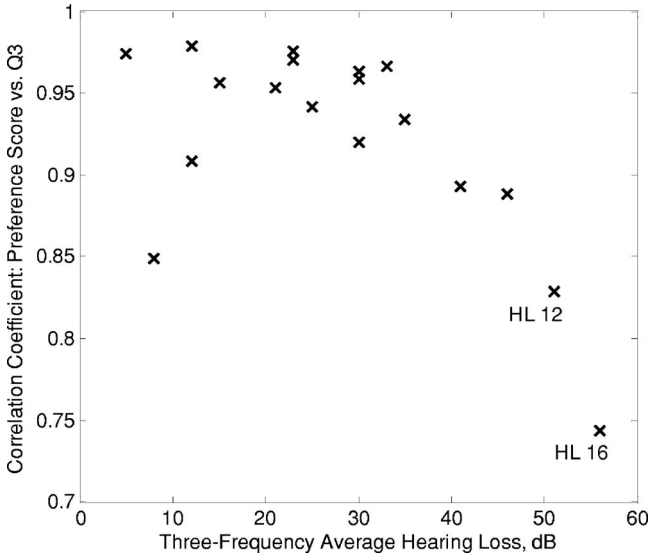


FIG. 5. Accuracy with which the Q_3 metric models the individual hearing-impaired subject preference scores, plotted as a function of the three-frequency average hearing loss.

V. DISCUSSION

The listener preference scores for peak-clipping, center-clipping and additive noise decrease monotonically with increasing signal degradation. While preference scores for listeners in the NH and HL groups follow similar trends, the scores calculated *across* distortion type are significantly different between the two groups of listeners. One objective is to predict the quality scores for both groups of listeners. To this end, we considered the coherence-based one-dimensional Q_3 model to predict preference scores. This analysis considers the audibility of the noise or distortion relative to the instantaneous signal level as a dominant factor in determining listener preferences, especially since the experimental design controlled for overall signal intensity. In this section, we discuss the CSII results in the context of previous studies using coherence calculations, as well as other studies that have considered audibility as a factor in quality judgments. In addition, we will discuss how additional factors beyond audibility may be required to fully model the quality judgments for noise and distortion in listeners with normal hearing and with hearing loss.

A. Coherence and audibility

Much of the previous work on using coherence to analyze or predict sound quality in hearing aids (Kates and Kozma-Spytek, 1994; Kozma-Spytek *et al.*, 1996; Versfeld *et al.*, 1999; Stelmachowicz *et al.*, 1999; Palmer, 2001) has concentrated on computing SDR for the entire signal rather than for portions of the signal in different amplitude regions. If only one form of distortion is present, for example, peak clipping (Kates and Kozma-Spytek, 1994; Kozma-Spytek *et al.*, 1996; Stelmachowicz *et al.*, 1999), the entire-signal coherence decreases monotonically with increasing distortion and an accurate prediction of the quality can be derived from the coherence. However, when various types of signal degradation were compared in the same experiment, the coherence of the entire signal was much less effective at predicting the quality judgments than the SII (Versfeld *et al.*, 1999). A problem with using the entire-signal coherence is that the calculation is dominated by the ratio of the strongest signal components (the vowel nuclei in speech) to the level of the noise and distortion, and it cannot directly measure the effects of the signal degradation on the weaker speech components.

Accurate predictions of quality can be produced across different noise and distortion conditions when the coherence calculation is modified to reflect the effects of the signal degradation on the different levels of the speech. Tan *et al.* (2004), for example, were able to predict the quality of de-

graded speech (correlation coefficient of 0.93) and music (correlation coefficient of 0.98) for normal-hearing listeners. Their procedure is based on a coherence calculation in which the test signal in each frequency band is divided into segments, the coherence computed for each segment, and the coherence values averaged across segments. Averaging the coherence values in this way gives equal weight to the loss of signal quality at low as well as high signal levels. Because the Tan *et al.* approach does not consider hearing loss, comparisons with the analysis used in this study were not pursued.

A further concern is not just the presence of noise and distortion, but rather the audibility of the signal degradation. For example, Eisenberg *et al.* (1998) found a strong correlation between the SII and subject quality judgments for speech clarity and intelligibility for bandpass filtered speech in additive speech-shaped noise. Thus, one can argue that the Q_3 metric is effective because it measures the effects of the noise and distortion on speech at different signal intensities and combines these measurements with estimates of the audibility of the noise and distortion at each signal intensity. Similar approaches have also proven effective for predicting speech intelligibility for noise and distortion (Kates and Arehart, 2005) and in fluctuating noise (Rhebergen and Versfeld, 2005).

The Q_3 metric is accurate in predicting quality, which indicates that the audibility of the noise and distortion is an important factor in the quality judgments and that coherence is a viable procedure for determining the amount of signal degradation. However, there are still some systematic errors in the predictions. For example, the metric tends to predict lower quality ratings for additive noise at poor SNRs than were given by the subjects. This type of discrepancy suggests that a one-dimensional reduction of the subject preferences may not capture all of the factors that entered into the preference judgments.

The audibility of speech and noise and distortion products is a factor in both the experimental design and the Q_3 metric. In our experimental protocol, the signal normalization and the amplification scheme (NAL-R) provide a specific procedure for adjusting the level (and hence, audibility) of the signal relative to impaired auditory thresholds. Different strategies for adjusting the level of the signals could change the relative audibility of different signal components and lead to different subject preferences. The Q_3 metric includes only a partial model of audibility, based on the SII. Better estimates of audibility, which might include modifying the filter bandwidth and upper spread of masking incorporated in the SII calculation or developing a more extensive peripheral model of cochlear hearing loss, may lead to more accurate quality predictions for listeners with more severe hearing loss.

B. Multidimensional unfolding

The possibility that several perceptual factors influence the quality judgments can be addressed by applying the methods of multi-dimensional unfolding (Coombs, 1950; Borg and Groenen, 2005) to the preference judgments. In the unfolding model, listeners and stimuli are represented as points in a space containing one or more dimensions. These dimensions can be interpreted as independent perceptual factors that contribute to the listeners' judgments. The position of each listener in the space is called their ideal point, which gives the location of that person's most-preferred stimulus. The location of each stimulus in the space gives a measure of the contribution of each of the perceptual factors in forming the preference judgment. In this model, stimulus preference is inversely proportional to the distance of each stimulus from the listener's ideal point (Borg and Groenen, 2005). We computed a variety of different scaling solutions based on different model assumptions. They all gave substantially similar results. We therefore report here the scaling solutions computed from ratio-scale transformation of the raw preference judgments using the robust PREFSCAL algorithm developed by the Data Theory Scaling System Group at the University of Leiden, included as part of SPSS 14.0 (Borg and Groenen, 2005; Meulman *et al.*, 2005).

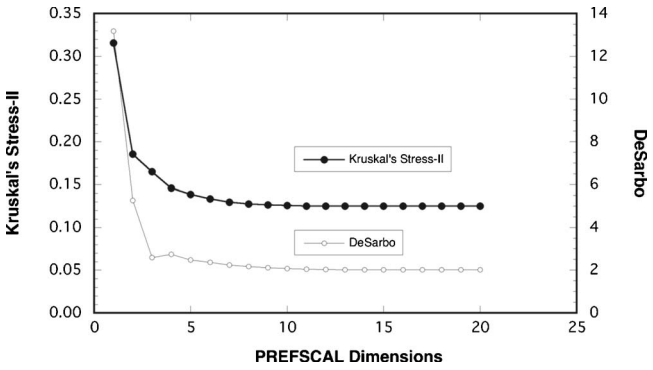


FIG. 6. Kruskal's Stress-II test (on the left axis) and DeSarbo's Degeneracy test (on the right axis) plotted as a function of the number of dimensions in the unfolding analysis.

The basic data for the multi-dimensional unfolding analysis are the number of times each condition was not chosen over the other 23 distortion-level conditions. Larger numbers correspond to lower preferences for that stimulus. The data were arranged in a 32 (listeners) by 24 (speech samples) matrix that was then analyzed by the unfolding procedure. The general goal was to find the lowest-dimensional solution that would give interpretable results while avoiding degeneracy (a mathematically correct but trivial result). Scaling solutions for 1 through 20 dimensions were computed using SPSS 14.0. The badness of fit (Kruskal Stress II) as a function of the number of dimensions is shown in Fig. 6. There is a monotonic improvement of the fit between the data and the predictions of the model as the number of dimensions increases, improving rapidly at first and then leveling off after several dimensions. We also considered the DeSarbo measure of mathematical degeneracy in the solutions (DeSarbo *et al.*, 1997). The DeSarbo measure considers the degree to which the points in the solution space from the speech samples are intermixed with those of the listeners. A high value of the DeSarbo intermixedness index suggests that the solution is degenerate. Figure 6 also shows the DeSarbo index plotted as a function of the number of dimensions in the scaling solution. The values of the stress

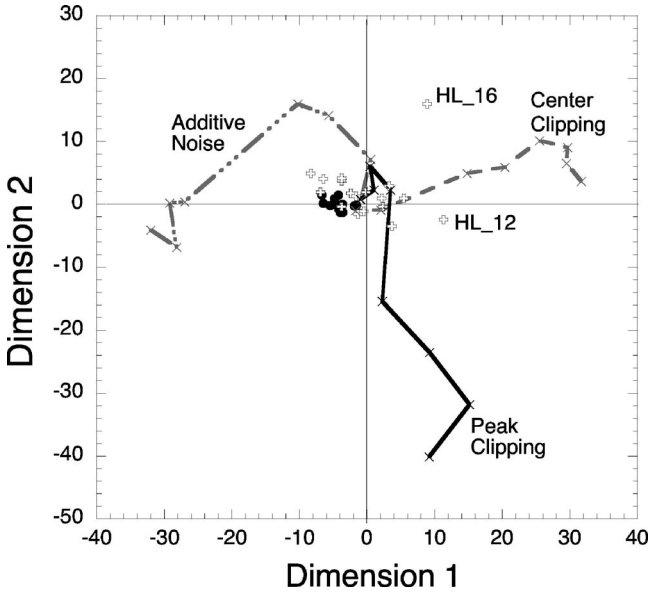


FIG. 7. Subject ideal points (filled circles=NH; open crosses=HL) and stimulus locations in the first two dimensions of the four-dimension solution.

and of the DeSarbo index led us to choose the four-dimensional solution to examine in greater detail.

The first two dimensions of the four-dimensional solution account for the largest proportion of the variance. Figure 7 shows the ideal points of the listeners and the position of each stimulus in the two-dimensional subspace. Three features of the figure deserve additional comment. The first is that the ideal points of the hearing-impaired listeners are shifted relative to those of normal-hearing listeners. The second is that the stimuli for each type of distortion move systematically away from the center as distortion increases, indicating that preference decreases with increased distortion. The third feature is that the peak-clipping stimuli move away from the center in a direction roughly orthogonal to the directions of the additive noise and center-clipping stimuli.

The filled circles near the center of the plot are the ideal points for the normal-hearing subjects, while the open crosses are the ideal points for the subjects with hearing loss. The ideal points represent the location of the stimulus characteristics that are most preferred by the individual subjects. The ideal points for the normal-hearing subjects cluster tightly close to the origin, while those for the subjects with hearing loss show more scatter. This increase in the scatter for the HL listeners is consistent with the greater variability in the HL data plotted in Figs. 1 and 2. While no systematic relationship was observed between degree of hearing loss and ideal point location (calculated by Euclidean distance), the two subjects with the greatest hearing loss (HL_12 and HL_16) had the greatest Euclidean distances.

Also shown in Fig. 7 (by "x" symbols) are the locations of stimuli in the two-dimensional subspace. The three undistorted stimuli are located close to the center of this space, and as stimuli become progressively degraded their position in the space moves progressively outward from the center. The stimuli degraded by additive noise first move upward and then progressively to the left; stimuli degraded by center clipping move to the right; and stimuli degraded by peak clipping move downward. Additive noise and center clipping appear to vary along dimension 1 and peak clipping along dimension 2. Therefore, dimension 1 appears to depend on the signal degradations that most strongly affect the low-level portions of the speech (as occurs for additive noise and for center clipping). In contrast, dimension 2 appears to depend on the high-level portions of the signal (as occurs for peak clipping).

The preference scores calculated *across* degradation types and plotted in Fig. 2 show that the judgments for the NH and HL listeners were similar for the peak-clipping distortion, but that the HL listeners rated the additive noise lower than the NH listeners but rated the center clipping higher than the NH listeners. One possible explanation for these differences is that both groups of listeners respond in a similar manner to modifications of the high-level portions of the signal, but that the HL listeners respond differently than the NH listeners to the low-level portions of the signal. Additive noise and center clipping differ in how they change the low-level portions of the signal; the noise masks the signal, while center clipping replaces the signal with silence. The coherence calculation used for the low-level CSII cannot distinguish between these two conditions since they both reduce the coherence value in similar ways. However, the plot of Fig. 7 clearly shows that the subjects perceive additive noise differently from center clipping. Even though noise and center clipping lie along the same dimension, the stimulus points for noise and center clipping move further away from each other with increasing amounts of signal degradation.

This distinction between additive noise and center clipping is also reflected in the locations of the ideal points for the listeners. The ideal points for the NH listeners lie to the left of the origin in Fig. 7, which indicates a slight preference for a small amount of additive noise. On the other hand, the ideal points for the HL listeners lie to the right of the origin, which indicates a slight preference for a small amount of center clipping. This finding is consistent with the preference scores presented in Fig. 2, which show that HL listeners rated center clipping higher and additive noise lower than did the NH listeners.

The Q_3 metric measures the audibility of the noise and distortion, and the accuracy of the Q_3 predictions indicates that audibility is an important factor in forming quality judgments. As the noise level is increased the noise becomes more audible at low signal levels, and as the center-clipping threshold is increased the distortion becomes more audible. This interpretation is similar to the importance of audibility in speech intelligibility, where an increased noise level masks the low-level speech components and an increased center-clipping threshold removes the low-level speech components; in either case the low-level speech is rendered inaudible. However, the Q_3 metric predicts that the reductions in speech quality for additive noise and center clipping will be identical while the multi-dimensional unfolding results show that they are not the same. Thus audibility alone is insufficient to explain quality judgments.

VI. CONCLUSIONS

- The listener preference scores for additive noise, for peak-clipping distortion, and for center-clipping distortion decrease monotonically with increasing signal degradation. While preference scores for listeners in the NH and HL groups follow similar trends, the scores from the two groups are significantly different when analyzed *across* distortion types.
- The coherence-based one-dimensional Q_3 metric accurately predicted the quality judgments of both normal-hearing and hearing-impaired listeners, suggesting that audibility may be a factor in listeners' quality judgments. This result is similar to a coherence-based one-dimensional I_3 metric used to predict speech intelligibility for the same noise and distortion conditions (Kates and Arehart, 2005). However, differences in the weights used in the quality and intelligibility analyses suggest that different perceptual properties of the degraded speech signal are used for the quality judgments than are used for intelligibility.
- Despite its accuracy, the Q_3 metric shows some systematic errors. The accuracy of the model appears to be lower for those subjects with the most severe hearing losses. This loss of accuracy may be related to an incomplete model of the audibility of the speech and the noise and distortion caused by signal degradation.
- Additional factors beyond audibility may also contribute quality perception. A multi-dimensional analysis indicates that several dimensions are needed to fully describe the factors used by the subjects to judge the effects of the multiple distortion types considered here. It may be possible to derive a more accurate quality metric by modeling not just the subject quality preference scores, but instead by modeling the underlying perceptual factors that are suggested by the multi-dimensional scaling. The factor predictions would then be combined into a single rating to produce a one-dimensional preference score.
- The Q_3 metric presented here includes a simplified model of the auditory periphery. A more complete model of peripheral hearing loss may lead to more accurate quality predictions. Peripheral factors might include broader auditory filters and increased upward spread of masking associated with hearing loss. Further considerations would be changes in the suprathreshold signal characteristics (e.g., loudness, envelope modulation).

ACKNOWLEDGMENTS

This work was supported in part by a grant from GN Resound Corporation. The information provided in this paper was also supported in part by Grant/Cooperative Agreement No. UR3/CCU824219 from the Centers for Disease Control and Prevention (CDC). The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of CDC. The authors thank Jessica Rossi-Katz for assistance in the data collection, and to Amit Das, Vinod Prakash and Ramesh Kumar Muralimanohar for development of the software used for the listener tests and for the data analysis. The authors also thank Michael Nilsson for providing the female HINT sentences.

APPENDIX

Instructions given to subject: In this experiment, you will be listening to different speech samples. Your task will be to decide which sample sounds better. In each trial, you will listen to speech Sample A and speech Sample B. After listening to both samples, you will select which speech sample you think sounds better. At times, you may find it more difficult to decide which sample sounds better. In all situations, we encourage you to take your best guess as to whether Sample A or Sample B sounds better.

[1]Ideally, the order with which the stimuli to be compared were played out should have been randomized within a trial. For each type of signal degradation (additive noise, peak clipping, or center clipping), the degradation condition can be numbered from one to eight, with a lower-numbered condition having less degradation than a high-numbered condition. In the experiment, a potential bias emerged: for one gender talker, the stimulus with a lower-degradation number was always presented in interval one and for the other gender talker, the stimulus with a lower-degradation number was always presented in interval two. Thus, listeners may have learned to use this potential bias in their responses, by preferentially choosing interval one with one gender talker and interval two with the other gender talker. To examine this potential bias, we statistically analyzed the pattern of responses for listeners on the trials in which the two stimuli within a pair were the same (cells indicated with dashes along the diagonals in Tables III and IV). If listeners were biased toward picking interval one for the male talker and interval two for the female talker (or vice versa), then this bias would be evidenced as a significant pattern in the same-stimuli pairs. However, no significant difference ($p=0.181$) was observed, which suggests that listeners' responses were not biased.

[2]In the analysis of the possible interval bias, we also examined whether there was a general bias towards interval 1 or interval 2. Listeners did show an overall bias for interval 2. We determined this bias using a repeated-measures ANOVA in which we compared how often interval 1 was chosen vs how often interval 2 was chosen. Listeners chose interval 2 more often than interval 1 [$F_{(1,31)}$; 13.9; $p=0.001$]. This bias was evident for both male and female talkers and is consistent with a recent proceedings report by Wickelmaier and Choisel (2006), who showed that given audio samples of similar quality, subjects are more likely to favor the second presentation.

ANSI S3.42. (**1992**). "American National Standard: Testing Hearing Aids with a Broadband Noise Signal" (American National Standards Institute, New York).

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

ANSI S3.6. (**1989**). "Specifications for audiometers," American National Standards Institute, New York.

Beerends, J. G., Hekstra, A. P., Rix, A. W., and Hollier, M. P. (**2002**). "Perceptual evaluation of speech quality (PESQ), the new ITU standard for end-to-end speech quality assessment, Part II—Psychoacoustic model," J. Audio Eng. Soc. **50**, 765–778.

Borg, I., and Groenen, P. (**2005**). *Modern Multidimensional Scaling: Theory and Applications*, 2nd ed. (Springer, New York).

Byrne, D., and Dillon, H. (**1986**). "The National Acoustics Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid," Ear Hear. **7**, 257–265.

CCITT. (**1986**). "Objective evaluation of non-linear distortion effects on voice transmission quality," CCITT Study Group XII, Communication-XII No. 8.

CCITT. (**1987**). "Re-evaluation of the objective method for measurement of non-linear distortion," CCITT Study Group XII, Communication XII-175-E.

Coombs, C. H. (**1950**). "Psychological scaling without a unit of measurement," Psychol. Rev. **57**, 145–158.

Crain, T. (**1992**). "The effect of peak clipping on the speech recognition threshold," Ph.D. thesis, University of Minnesota.

Czerwinski, E., Voishvillo, A., Alexandrov, S., and Terekhov, A. (**2001a**). "Multitone testing of sound system components—Some results and conclusions, Part 1: History and Theory," J. Audio Eng. Soc. **49**, 1181–1192.

Czerwinski, E., Voishvillo, A., Alexandrov, S., and Terekhov, A. (**2001b**). "Multitone testing of sound system components—some results and conclusions, Part 2: Modeling and application," J. Audio Eng. Soc. **49**, 1011–1042.

David, H. A. (**1963**). *The Method of Paired Comparisons* (Hafner, New York).

DeSarbo, W. S., Young, M. R., and Rangaswarmy, A. (**1997**). "A parametric multidimensional unfolding procedure for incomplete nonmetric preference/choice set data in marketing research," J. Marketing Res. **34**, 499–516.

Eisenberg, L. S., Dirks, D. D., Takayanagi, S., and Martinez, A. S. (**1998**). "Subjective judgments of clarity and intelligibility for filtered stimuli with equivalent speech intelligibility index predictions," J. Speech Lang. Hear. Res. **41**, 327–339.

Gabrielsson, A., and Sjögren, H. (**1979**). "Perceived sound quality of sound reproducing systems," J. Acoust. Soc. Am. **65**, 1018–1033.

Gabrielsson, A., Shenkman, B. N., and Hagerman, B., (**1988**). "The effects of frequency responses on sound quality judgments and speech intelligibility," J. Speech Hear. Res. **31**, 166–177.

Geddes, E., and Lee, L. (**2003**). "Auditory perception of nonlinear distortion—Theory," *Audio Eng. Soc. 115th Convention*, New York, paper No. 5890.

Hansen, J. H. L. (**1999**). "Speech enhancement," *Encyclopedia of Electrical and Electronics Engineering* (Wiley, New York), Vol. **20**, 159–175.

ITU BS.1387-1. (**2001**). "Method for objective measurements of perceived audio quality" (International Telecommunications Union, Geneva).

ITU P.862. (**2001**). "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and codecs" (International Telecommunications Union, Geneva).

Kates, J. M. (**1992**). "On using coherence to measure distortion in hearing aids," J. Acoust. Soc. Am. **91**, 2236–2244.

Kates, J. M. (**1993**). "Hearing-aid design criteria," J. Speech Lang. Path. and Audiology Monograph Suppl. **1**, 15–23.

Kates, J. M., and Arehart, K. H. (**2005**). "Coherence and the speech intelligibility index," J. Acoust. Soc. Am. **117**, 2224–2237.

Kates, J. M., and Kozma-Spytek, L. (**1994**). "Quality ratings for frequency-shaped peak-clipped speech," J. Acoust. Soc. Am. **95**, 3586–3594.

Kochkin, S. (**2005**). "Customer satisfaction with hearing instruments in the digital age," Hear. J. **58**, 30–37.

Kondoz, A. M. (**2004**). *Digital Speech Coding for Low bit-Rate Communication Systems* (Wiley, Hoboken, NJ).

Kozma-Spytek, L., Kates, J. M., and Revoile, S. G. (**1996**). "Quality ratings for frequency-shaped peak-clipped speech: Results for listeners with hearing loss," J. Speech Hear. Res. **39**, 1115–1123.

Lawson, G. D., and Chial, M. R. (**1982**). "Magnitude estimation of degraded speech quality by normal- and impaired-hearing listeners," J. Acoust. Soc. Am. **72**, 1781–1787.

Licklider, J. (**1946**). "Effects of amplitude distortion upon the intelligibility of speech," J. Acoust. Soc. Am. **18**, 429–434.

Meulman, J. J., Heiser, W. J., and SPSS, I. (**2005**). *SPSS Categories® 14.0*, SPSS, Chicago.

Narendran, M. M., and Humes, L. E. (**2003**). "Reliability and validity of judgments of sound quality in elderly hearing aid wearers," Ear Hear. **24**, 4–11.

Nilsson, M., Soli, S. D., and Sullivan, J. (**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.

Palmer, C. V., Killion, M. C., Wilber, L. A., and Ballad, W. J. (**1995**). "Comparison of two hearing aid receiver-amplifier combinations using sound quality judgments," Ear Hear. **16**, 587–598.

Palmer, C. V. (**2001**). "The impact of hearing loss and hearing aid experience on sound quality judgments," Semin. Hear. **22**, 125–138.

Preminger, J. E., and Van Tasell, D. J. (**1995**). "Quantifying the relation between speech quality and speech intelligibility," J. Speech Hear. Res. **38**, 714–725.

Punch, J., and Beck, E. (**1980**). "Low-frequency response of hearing aids and judgments of aided speech quality," J. Speech Hear Disord. **45**, 325–335.

Quakenbush, S. R., Barnwell, T. P., and Clements, M. A. (**1988**). *Objective Measures of Speech Quality* (Prentice–Hall, Englewood Cliffs, NJ).

Rabiner, L. R., Levitt, H., and Rosenberg, A. E. (**1969**). "Investigation of stress patterns for speech synthesis by rule," J. Acoust. Soc. Am. **45**, 92–101.

Rhebergen, K. S., and Versfeld, N. J. (**2005**). "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **117**, 2181–2192.

Souza, P., Jenstad, L., and Boike, K. (**2006**). "Measuring the acoustic effects of compression and amplification on speech in noise," J. Acoust. Soc. Am. **119**, 41–44.

Stelmachowicz, P. G., Lewis, D. E., Hoover, B., and Keefe, D. H. (**1999**). "Subjective effects of peak clipping and compression limiting in normal and hearing-impaired children and adults," J. Acoust. Soc. Am. **105**, 412–422.

Tan, C.-T., Moore, B. C. J., and Zacharov, N. (**2003**). "The effect of nonlinear distortion on the perceived quality of music and speech signals," J. Audio Eng. Soc. **51**, 1012–1031.

Tan, C.-T., Moore, B. C. J., Zacharov, N., and Mattila, V.-V. (**2004**). "Predicting the perceived quality of nonlinearly distorted music and speech signals," J. Audio Eng. Soc. **52**, 699–711.

Thiede, T., Treurniet, W. C., Bitto, R., Schmidmer, C., Sporer, T., Beerends, J. G., Colomes, C., Keyhl, M., Stoll, G., Brandenberg, K., and Feiten, B. (**2000**). "PEAQ – The ITU standard for objective measurement of perceived audio quality," J. Audio Eng. Soc. **48**, 3–29.

Versfeld, N. J., Festen, J. M., and Houtgast, T. (**1999**). "Preference judgments of artificial processed and hearing-aid transduced speech," J. Acoust. Soc. Am. **106**, 1566–1578.

Wickelmaier, F., and Choisel, S. (**2006**). "Modeling within-pair order effects in paired comparison judgments," *Proceedings of Fechner Day 2006, 22nd Annual Meeting of the International Society for Psychophysics*, July 25–28, St. Albans, Hertfordshire, England, pp. 89–94.

Yonovitz, A., Bickford, B., Lozar, J., and Ferrell, D. (**1978**). "Electroacoustic distortions: Multidimensional analysis of hearing aid transduced speech and music," *IEEE International Conference of Acoustics, Speech, and Signal Processing*, 270–274, Tulsa.

Factors influencing glimpsing of speech in noise

Ning Li and Philipos C. Loizou[a)]
Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75083-0688

(Received 15 December 2006; revised 4 April 2007; accepted 22 May 2007)

The idea that listeners are able to "glimpse" the target speech in the presence of competing noise has been supported by many studies, and is based on the assumption that listeners are able to glimpse pieces of the target speech occurring at different times and somehow patch them together to hear out the target speech. The factors influencing glimpsing in noise are not well understood and are examined in the present study. Specifically, the effects of the frequency location, spectral width, and duration of the glimpses are examined. Stimuli were constructed using an ideal time-frequency (*T-F*) masking technique that ensures that the target is stronger than the masker in certain *T-F* regions of the mixture, thereby rendering certain regions easier to glimpse than others. Sentences were synthesized using this technique with glimpse information placed in several frequency regions while varying the glimpse window duration and total duration of glimpsing. Results indicated that the frequency location and total duration of the glimpses had a significant effect on speech recognition, with the highest performance obtained when the listeners were able to glimpse information in the $F1/F2$ frequency region (0–3 kHz) for at least 60% of the utterance. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2749454]

PACS number(s): 43.72.Dv, 43.72.Ar [DOS] Pages: 1165–1172

I. INTRODUCTION

The notion that listeners can "glimpse" the target speech when listening in noise dates back to an early study by Miller and Licklider (1950) on the intelligibility of interrupted speech masked by noise. Miller and Licklider (1950) assessed the intelligibility of interrupted speech produced by gating the speech signal on and off at a range of modulation frequencies. They found that high levels of speech understanding can be obtained when the modulation frequency (rate of interruption) was around 10 Hz even though 50% of the signal was gated off. Miller and Licklider (1950) concluded that listeners were able to piece together glimpses of the target speech available during the uninterrupted ("on" segments) portions of speech. In this study, listeners had access to the full spectrum during the uninterrupted portion. Other studies (e.g., Howard-Jones and Rosen, 1993; Buss *et al.*, 2003) used a "checkerboard" type of noise masker to investigate whether listeners were able to integrate asynchronous glimpses present in disjoint segments of the spectrum. Howard-Jones and Rosen (1993) showed that listeners were able to piece together asynchronous glimpses, provided the spectral region of the glimpses was wide enough.

Evidence of glimpsing was also reported in intelligibility studies investigating the difference in performance between identifying words in the presence of steady-state noise and in the presence of a single competing talker. Several studies (e.g., Festen and Plomp, 1990; Miller, 1947) confirmed that performance is lower in steady-state noise than in a single competing talker, and the difference in speech reception threshold (SRT) was large (6–10 dB). This difference was attributed to the fact that listeners were exploiting the silent gaps or waveform "valleys" in the competing signal to recognize the words in the target sentence. These gaps presumably enabled listeners to "glimpse" entire syllables or words of the target voice, since the local SNR is quite favorable during those gaps.

The listening-in-the-gaps account of speech segregation falls apart, however, when there are large numbers (more than 4) of competing voices present since the masker waveform becomes nearly continuous, leaving no silent gaps in the waveform (Miller, 1947). A different view of glimpsing was proposed by Cooke (2003, 2005) extending and generalizing the above idea of listening in the gaps. This new view was based on a different definition of what constitutes a *glimpse*: "a time-frequency region which contains a reasonably undistorted 'view' of local signal properties" (Cooke, 2005). Useful signal properties may include signal energy or presence of reliable $F0$ and/or formant frequency information. Glimpses of speech in background noise might, for instance, comprise of all time-frequency (*T-F*) bins or regions having a local SNR exceeding a certain threshold value (e.g., 0 dB). This definition of glimpse is henceforth adopted in the present study. The assumption is that listeners are able to first detect "useful" glimpses of speech, possibly occurring at different times and occupying different regions of the spectrum, and then somehow integrate those glimpses to hear out the target speech.

Computational models of glimpsing were developed for computational auditory scene analysis (CASA) algorithms and for robust automatic speech recognition by modifying the recognition process to allow for the possibility of "missing data" (Cooke *et al.* 1994, 2001). Despite the attractive appeal of glimpsing as a means of speech segregation in competing noise sources, there remain several issues to be resolved. Foremost among those issues is the question of what constitutes a useful glimpse and whether glimpses contain sufficient information to support identification of the tar-

[a)]Electronic mail: loizou@utdallas.edu

get signal. Several studies (Roman *et al.*, 2003; Roman and Wang, 2006; Cooke, 2006; Brungart *et al.*, 2006; Anzalone *et al.*, 2006) have attempted to answer these questions and demonstrated that speech synthesized from the ideal binary mask is highly intelligible even when extracted from multisource mixtures (Roman *et al.*, 2003) or under reverberant conditions (Roman and Wang, 2006). The ideal binary "mask" takes values of 0 and 1, and is constructed by comparing the local SNR in each *T-F* unit against a threshold (e.g., 0 dB). The ideal mask is commonly applied to the *T-F* representation of a mixture signal and eliminates portions of a signal (those assigned to a "0" value) while allowing others (those assigned to a "1" value) to pass through intact. Roman *et al.* (2003) assessed the performance of an algorithm that used location cues and an ideal time-frequency binary mask to synthesize speech. Large improvements in intelligibility were obtained from partial spectro-temporal information extracted from the ideal time-frequency mask. Similar findings were also reported by Brungart *et al.* (2006), for a range of SNR thresholds (−12 to 0 dB) used for constructing the ideal binary mask. A different method for constructing the ideal binary mask was used by Anzalone *et al.* (2006) based on comparisons of the speech energy detected in various bands against a preset threshold. The threshold value was chosen such that a fixed percentage (99%) of the total energy contained in the entire stimulus was above this threshold. Results with the ideal speech energy detector indicated significant reductions in speech reception thresholds (SRTs) for both normal-hearing and hearing-impaired listeners. Cooke (2006) used a computational model of glimpsing along with behavioral data collected from normal-hearing listeners on a consonant identification task. Several different glimpsing models were tested differing in the local SNR used for detection, the minimum glimpse size, and the use of information in the masked regions. Close fits to listener's performance on a consonant task were obtained with local SNR thresholds in the range of −2 to 8 dB.

The ideal time-frequency mask used in the above intelligibility studies for synthesizing speech makes the implicit assumption that all *T-F* units falling below a prescribed SNR threshold (e.g., 0 dB) are not detectable and should therefore be eliminated. While this assumption is valid in situations wherein there is little or no spectral overlap between the masker and the target signal in individual *T-F* units, it is not valid for speech babble or other broadband type of maskers where there exists a great deal of spectral overlap between the masker and the target. It is very likely that the masker has enough energy to distort the signal, but not to the point that it makes the target signal undetectable. Nonsimultaneous masking effects, for instance, are not taken into account when zeroing out the *T-F* units falling below the SNR threshold. Furthermore, it is known from intelligibility studies (Drullman, 1995) that the weak elements of speech lying below the noise level do contribute to some extent (up to −4 dB) to intelligibility and should therefore be preserved.

A different approach is taken in this paper to address the above limitations of using the ideal binary mask as a tool to study speech segregation or auditory scene analysis. In the proposed approach, rather than eliminating completely any *T-F* unit falling below the SNR threshold, we consider retaining those units. The *T-F* mask is no longer binary but takes real values. In the proposed approach, speech is synthesized by retaining all *T-F* units falling below the local SNR threshold while carefully controlling the duration and frequency region of the *T-F* units above the SNR threshold. The synthesized stimuli better approximate the acoustic stimuli encountered by normal-hearing listeners in a real-world noisy scenario. Under this framework, the present study aims to answer the question of what is a useful glimpse and examine the various factors that could potentially influence glimpsing in noise.

The total duration of glimpsing is one of many factors hypothesized to influence performance. In most CASA-based methods, it is assumed that glimpsing opportunities are available throughout the utterance. In practice, only a portion of the signal might be glimpsed, which in turn raises the question: What is the minimum duration of glimpsing required to achieve high levels of performance? An experiment is conducted in the present study to answer this question. In the study by Miller and Licklider (1950), 50% of the stimulus was uninterrupted and available for glimpsing, with performance steadily improving as the total duration increased. Listeners, however, had access to the full spectrum during the uninterrupted portions of speech, an assumption that generally does not hold in a complex listening situation. Only a portion of the spectrum is typically available to listeners for glimpsing in noisy environments depending on the temporal/spectral characteristics of the masker. This, in turn, raises another question: What is the influence of the location and/or width of the frequency region that is available for glimpsing? Clearly, the glimpse window width (i.e., glimpse window duration) will affect the answer to this question, and for that reason we examine systematically in experiment. 1 the influence of glimpse window width for different frequency regions of glimpsing. Previous studies showed that listeners can exploit glimpse window widths lasting as long as a phoneme for sentence/word recognition tasks (e.g., Miller and Licklider, 1950), and as short as 10 ms for a double-vowel identification task (Culling and Darwin, 1994). In most of these studies, however, listeners had either access to the full spectrum or disjoint segments of the spectrum (i.e., "checkerboard" noise) occurring periodically in time. These conditions might not reflect the true scenario in noisy environments faced by listeners wherein glimpsing opportunities may occur randomly in both time and frequency.

The findings from the present study have important implications for CASA and speech enhancement algorithms aiming to improve speech intelligibility. In many of the above studies, it is assumed that an ideal binary mask is available throughout the utterance and across the whole spectrum. In a practical system, the binary mask needs to be estimated from the noisy data, and that is a challenging task, particularly in adverse noisy conditions. Since it is practically impossible to compute accurately the ideal binary mask for all frames and all frequencies, it is of interest to determine at the very least the region in the spectrum that is perceptually most important and also the minimum duration of glimpsing required to synthesize highly intelligible

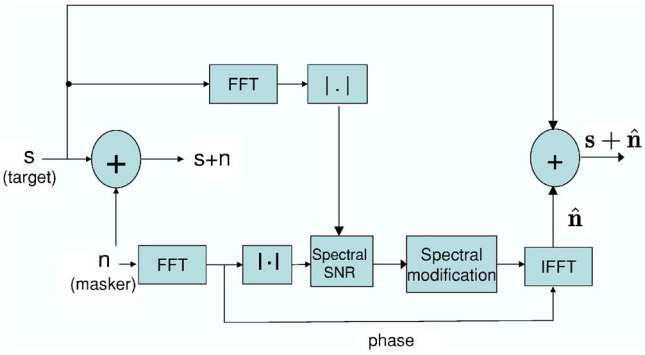


FIG. 1. (Color online) Block diagram of the signal processing technique used for constructing stimuli with glimpse information injected in prescribed frequency bands.

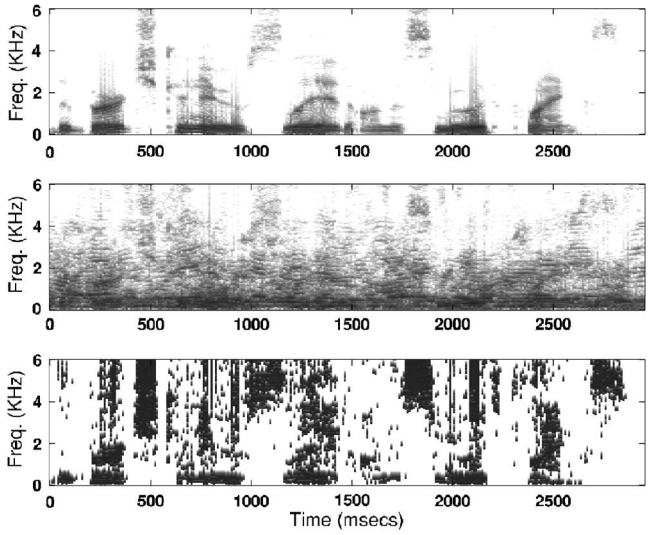


FIG. 2. Top panel shows the spectrogram of a sentence in quiet from the IEEE corpus. Middle panel shows the spectrogram of the sentence embedded in multitalker babble at −5-dB SNR. Bottom panel shows the ideal binary mask using an SNR threshold of 0- dB, with white pixels indicating a 0 (target weaker than the masker) and black pixels indicating a 1 (target stronger than the masker).

speech. These questions are addressed in the present paper.

II. EXPERIMENT 1: EFFECT OF GLIMPSE WINDOW WIDTH AND FREQUENCY LOCATION ON SPEECH INTELLIGIBILITY

A. Methods

1. Subjects

Nine normal-hearing listeners participated in this experiment. All subjects were native speakers of American English, and were paid for their participation. Subject's age ranged from 18 to 40 years, with the majority being undergraduate students from the University of Texas at Dallas.

2. Stimuli

The speech material consisted of sentences taken from the IEEE database (IEEE, 1969). All sentences were produced by a male speaker. The sentences were recorded in a sound-proof booth (Acoustic Systems) in our lab at a 25-kHz sampling rate. Details about the recording setup and copies of the recordings are available in Loizou (2007). The IEEE database consists of 72 phonetically balanced lists, each consisting of ten sentences. The sentences were corrupted by a 20-talker babble (Auditec CD, St. Louis) at −5-dB SNR. This SNR level was chosen to avoid floor effects (i.e., performance near zero).

3. Signal processing

To create stimuli with glimpses present in certain frequency regions, we spectrally modified the masker signal according to the diagram shown in Fig. 1. Our definition of glimpse is similar to that used by Cooke (2006): a time-frequency (*T-F*) region wherein the speech power is greater than the noise power by a specific threshold value (see the example in Fig. 2). In our study, we used a threshold of 0 dB, which is the threshold typically used for constructing ideal binary masks (Wang, 2005). Different SNR thresholds are considered later.

As shown in Fig. 1, the masker signal (20-talker babble) is first scaled (based on the rms energy of the target) to obtain a desired −5-dB SNR level. The target and scaled masker signals are segmented (using rectangular windows with no overlap) into 20-ms frames. A fast Fourier transform (FFT) is applied to each frame of the scaled masker to obtain the magnitude masker spectrum. Two different types of scaling are done to the masker spectrum depending on whether the *T-F* units fall within a prescribed region of the spectrum (i.e., the glimpse region) or outside the glimpse region. For all *T-F* units falling within the prescribed frequency region (glimpse region), the masker spectrum is appropriately scaled to ensure that the target *T-F* units are greater or equal (since the SNR threshold is 0 dB) in magnitude to the masker *T-F* units. The scaling (see the Appendix A for more details) is done independently in all *T-F* units in the masker spectrum which are in the glimpse region and are larger in magnitude than the corresponding target *T-F* units. No spectral modifications are done to individual *T-F* units in the masker spectrum if the target *T-F* units happen to be larger in magnitude than the masker *T-F* units. For all *T-F* units falling outside the prescribed frequency region (i.e., outside the glimpse region), the masker spectrum is appropriately scaled to ensure that the target *T-F* units are smaller in magnitude than the masker *T-F* units [note that in other studies (e.g., Brungart *et al.*, 2006), spectral components falling below the SNR threshold are set to zero]. No spectral modifications are done to individual *T-F* units in the masker spectrum if the target *T-F* units happen to be smaller in magnitude than the masker *T-F* units. The two types of scaling done to the masker spectrum ensure that only the prescribed frequency band contains glimpsing information. Following the masker magnitude modification, an inverse FFT is applied to the modified magnitude spectrum to obtain the masker signal in the time domain. The original phase spectrum of the masker is used in the reconstruction. The modified masker signal is finally added in the time domain to the clean speech signal to obtain the desired stimulus with glimpses present in a prescribed frequency band (see Appendix A for more details).

Three different frequency bands were considered: a low-

frequency (LF) band (0–1 kHz), a middle-frequency (MF) band (1–3 kHz), and a high-frequency (HF) band (>3 kHz). These bands were chosen to assess the individual contribution of formant frequencies ($F1$ and $F2$) on glimpsing in noise. The LF band contains primarily $F1$ information and the MF band contains $F2$ information. In addition to the above three bands, we also considered a low-to-mid-frequency (LF+MF) band: 0–3 kHz. This band was included as it contains both $F1$ and $F2$ information critically important for speech recognition. For comparative purposes, we also considered the following two conditions: (1) a condition spanning the full (FF) signal bandwidth, and (2) a condition, termed RF, in which the LF, MF and HF bands were randomly selected in each frame with equal probability.

To assess the effect of number of glimpses (i.e., the number of glimpse opportunities) on speech recognition, we created stimuli with different glimpse window widths (i.e., glimpse window durations). More specifically, we created stimuli with glimpse window widths of 20, 200, 400, and 800 ms spanning the duration of a phoneme to a few words. The glimpse window width is defined here as the total duration of a single glimpse spanning multiple, and neighboring in time, frames of speech. For instance, a single 200-ms glimpse is composed of ten consecutive frames (20 ms each) all containing glimpse information in a prescribed frequency band. Similarly, one 400-ms glimpse is composed of 20 consecutive frames, and one 800-ms glimpse is composed of 40 consecutive frames. The total duration of all glimpses introduced over the whole utterance was fixed to 800 ms. This number was chosen as it corresponds approximately to 33% of the total duration of most sentences in the IEEE database (average duration of sentences in the IEEE corpus was 2.4 s with a standard deviation of 0.3 s). Cooke (2005) observed that speech corrupted by eight talkers contains approximately 30% glimpses (based on a −3-dB SNR threshold). Since the signal processing involved is based on spectrally modifying the masker spectrum on a frame-by-frame basis, which is 20 ms in our experiments, we chose 20 ms as the smallest window width (duration) to be evaluated. Pilot data showed that glimpse window widths between 20 and 200 ms yielded comparable performance. Given that the total duration of all glimpses across the whole utterance was fixed at 800 ms, we created stimuli that had either 40 20-ms window glimpses, four 200-ms window glimpses, two 400-ms window glimpses, or one 800-ms window glimpse. The time location of each glimpse within the utterance was selected randomly. For comparative purposes, we also constructed stimuli in which the glimpses were present throughout the whole duration of each utterance.

In summary, we created stimuli which had low-frequency (LF) glimpse information, middle-frequency (MF) glimpse information, high-frequency (HF) glimpse information, low-to-mid frequency (LF+MF) glimpse information, randomly selected frequency (RF) information, and full-bandwidth (FF) glimpse information. For each of the above spectral regions, the glimpse window width was set to 20, 200, 400, 800 ms, and the whole utterance. To assess the potential gain in intelligibility introduced by glimpsing, we also included as a baseline condition the unmodified noisy sentences (−5-dB SNR). Two lists of sentences (i.e., 20 sentences) were used per condition, and none of the lists were repeated across conditions.

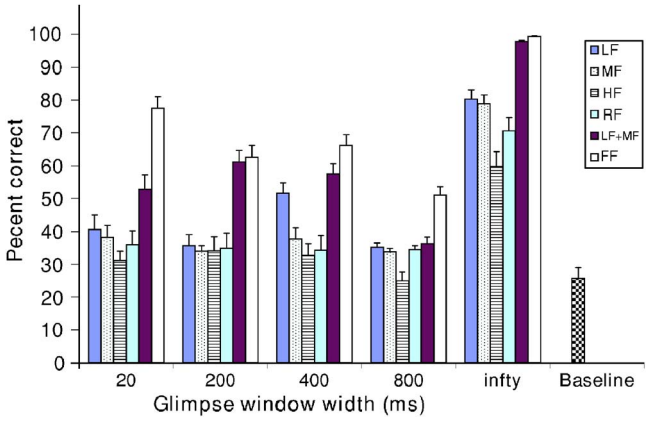


FIG. 3. (Color online) Mean subject recognition performance as a function of glimpse window width (in ms) for different frequency bands. The "infty" condition corresponds to the condition in which the indicated frequency bands were glimpsed throughout the whole utterance. The baseline condition corresponds to the unprocessed stimuli embedded in −5-dB SNR. Error bars indicate standard errors of the mean.

4. Procedure

The experiments were performed in a sound-proof room (Acoustic Systems, Inc) using a PC connected to a Tucker-Davis system 3. Stimuli were played to the listeners monaurally through Sennheiser HD 250 Linear II circumaural headphones at a comfortable listening level. Prior to the test, each subject listened to a set of noisy sentences to get familiar with the testing procedure. During the test, the subjects were asked to write down the words they heard. The order of the test conditions was randomized across subjects.

B. Results and discussion

The mean scores for all conditions are shown in Fig. 3. Performance was measured in terms of percent of words identified correctly (all words were scored). The mean baseline score of the unprocessed stimuli was 25.8% correct (s.d.=9.2%). Two-way ANOVA (repeated measures) indicated a significant effect of glimpse window width ($F[4,12]=193.9$, $p<0.0005$), a significant effect of frequency band location ($F[5,15]=122.9$, $p<0.0005$), and a significant interaction ($F[20,60]=7.75$, $p<0.0005$).

Protected *posthoc* tests (Fisher's LSD) were run to examine whether there were any differences in performance between the various glimpse window widths. This analysis aims to answer the question whether it is more beneficial to have multiple, but short, glimpse opportunities or few, but long, glimpse opportunities. Separate analysis was performed for each frequency band. For the LF band, and considering only glimpse window widths from 20 to 800 ms, performance peaked at 400 ms. That is, performance at 400 ms was significantly ($p<0.05$) higher than performance at 20, 200, or 800 ms. A different pattern emerged for the other frequency bands. For the MF, HF, LF+MF, and RF bands, performance remained relatively flat across all

glimpse window widths (20–800 ms). That is, there was no statistically significant ($p>0.05$) difference in performance between the 20, 200, or 800-ms conditions. When the full bandwidth (FF) was available for glimpsing, performance peaked at 20 ms. This suggests that it is more beneficial to have multiple, but short (20 ms), glimpse opportunities rather than few, but long (400–800 ms), glimpse opportunities. This finding applies only to the full-bandwidth (FF) condition, which does not reflect the realistic scenario of listening in noise. It does, however, have important implications for speech enhancement algorithms. If an enhancement algorithm improves the spectral SNR across the whole signal bandwidth, and does so for at least 33% of the utterance duration (which is the duration used in experiment 1), then there is a good likelihood that the algorithm will significantly improve speech intelligibility. In practice, it is extremely challenging to improve the spectral SNR at all frequencies; hence, it is more practical to look for frequency bands that perform as well (or nearly as well) as when glimpsing the full signal bandwidth (more on this follows).

Next, we examined the effect of frequency band location on glimpsing in noise. We were interested in knowing whether a particular frequency band offers more benefit than others (in terms of intelligibility); hence, we ran protected *posthoc* analysis (Fisher's LSD) on the data for a fixed glimpse-window width. Results indicated the LF+MF band performed significantly ($p<0.05$) better than the other bands (LF, MF, RF) in nearly all conditions. The exception was in the 400 and 800-ms conditions wherein performance with the LF band was not statistically different ($p>0.05$) from the performance obtained with the LF+MF band. Comparison between the performance obtained with the LF+MF band and the full bandwidth (FF) condition indicated that the intelligibility scores did not differ significantly ($p>0.05$) in three of the five conditions tested. More specifically, performance with the LF+MF band in the 200-ms, 400-ms, and whole utterance glimpse conditions was the same as that obtained with the FF band (whole bandwidth), and was significantly ($p<0.05$) lower than the FF condition only in the 20 and 800-ms conditions. The finding that the LF+MF band condition performed the best and attained in nearly all cases the upper bound in performance (i.e., was as good as FF) is not surprising given that the LF+MF band contains $F1$ and $F2$ information critically important for speech recognition. The implications of this finding for speech enhancement and CASA applications is that in order to improve speech intelligibility is it extremely important to improve at the very least the spectral SNR in the region of 0–3 kHz (LF+MF band), which is the region containing $F1$ and $F2$ information.

Finally, we assessed the gain in speech intelligibility introduced by glimpsing in the various frequency bands. This gain is assessed in reference to the baseline noisy condition (−5-dB SNR). Figure 4 plots the difference in score between the scores reported in Fig. 3 and the baseline score (26.8% correct). Protected *posthoc* tests (Fisher's LSD) were run to examine whether there were any significant differences between the scores obtained with and without glimpsing (i.e., baseline score). Asterisks in Fig. 4 indicate the presence of statistically significant differences. Results indicated that introducing glimpses in the LF band produced small (about 10%–5%), but statistically significant ($p<0.05$), improvement in performance. This outcome is consistent with the findings by Anzalone *et al.* (2006), who applied, in one condition, the ideal speech energy detector only to the lower frequencies (70–500 Hz). Significant reductions in SRT were obtained by both normal-hearing and hearing-impaired listeners when the ideal speech detector was applied only to the lower frequencies (Anzalone *et al.*, 2006).

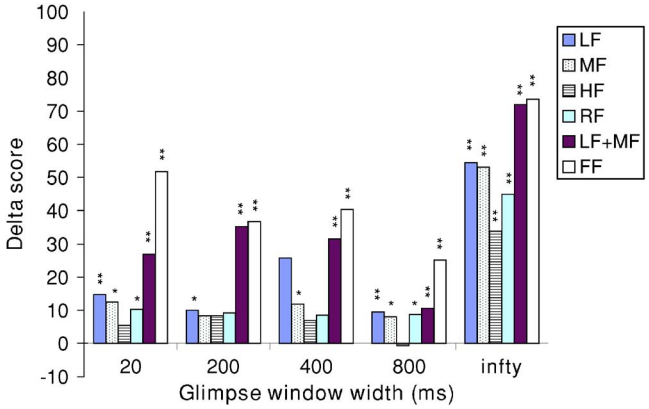


FIG. 4. (Color online) Difference in performance between that reported in Fig. 3 with glimpsed stimuli, and the baseline performance (26.8% correct). Asterisks ($*p<0.05$, $**p<0.005$) indicate statistically significant differences between the performance obtained with glimpsed stimuli and baseline stimuli. The "infty" condition corresponds to the condition in which the indicated frequency bands were glimpsed throughout the whole utterance.

Considerably larger (25%–35%), and significant ($p<0.005$), improvements were obtained in our study when glimpses were introduced in the LF+MF region. No significant ($p>0.05$) gain in intelligibility was observed when the glimpses were introduced in the HF band in any of the conditions (20–800 ms).[1] Also, no significant gain was observed when glimpses were introduced in the MF band (200 ms) or in the RF band (200, 400 ms). As one might expect, large improvements (>50%) were observed when glimpses were introduced in all frames throughout the utterance. Performance in the RF condition was consistently poor in nearly all conditions. This suggests that it is more difficult for listeners to integrate glimpses available in different frequency regions at different times, than to integrate glimpses available in the same region across time. It should be pointed out that the glimpses in the RF condition appeared randomly in time and frequency and differed in this respect to the checkerboard type of noise used in other studies (e.g., Buss *et al.*, 2003; Howard-Jones and Rosen, 1993) which appeared periodically.

The local SNR threshold used for defining the glimpses in the present experiment was fixed at 0 dB, and its value can understandably influence the outcome of the experiment. Interested to know whether a different pattern of results would be obtained with different SNR threshold values, we ran a follow-up experiment in which we varied the SNR threshold from −6 to 12 dB. Five new subjects were recruited for this experiment. The same signal-processing technique described in Sec. II A 3 (see Fig. 1) was adopted to construct stimuli with glimpses available in the LF+MF

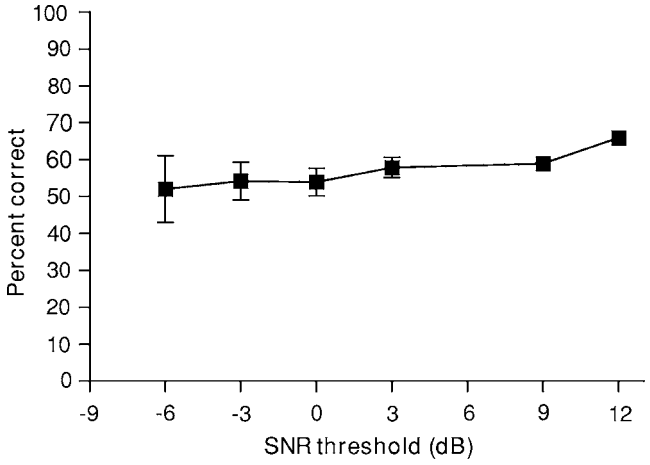


FIG. 5. Mean subject recognition performance as a function of the local SNR threshold for stimuli glimpsed in the LF+MF band. Error bars indicate standard errors of the mean.

band. This band was chosen as it performed nearly as well as the FF condition (full spectrum available). The glimpse window width was set to 20 ms. The procedure outlined in Sec. II A 4 was followed. The results, plotted in terms of percent correct, are shown in Fig. 5 as a function of the local SNR threshold. ANOVA (with repeated measures) indicated a nonsignificant ($F[5,20]=1.78$, $p=0.163$) effect of SNR threshold on performance. Performance increased slightly, but nonsignificantly, as the SNR threshold increased, and remained the same for negative values of the SNR threshold. It is worth noting that the plateau in performance seen in Fig. 5 is partially consistent with that observed by Brungart *et al.* (2006) using the ideal binary mask. The main difference between our study and that of Brungart *et al.* (2006) is that in our case performance remained flat even for positive SNR thresholds, whereas in Brungart *et al.* (2006), performance dropped precipitously for SNR thresholds above 0 dB. This difference is attributed to the fact that in Brungart *et al.* (2006) all *T-F* units falling below the SNR threshold were zeroed out; hence, the number of retained *T-F* units progressively decreased as the SNR threshold increased. In contrast, in our study all *T-F* units falling below the SNR threshold were retained [see Eq. (A5) in the Appendix A].

In summary, the results from the present experiment indicate that the glimpse window width as well as the SNR threshold had only a minor effect on performance. Glimpsing in noise was primarily affected by the location of the frequency band containing glimpses. High gains in intelligibility were obtained when glimpse information was available in the $F1$-$F2$ region (0–3 kHz).

III. EXPERIMENT 2: EFFECT OF TOTAL GLIMPSE DURATION ON SPEECH INTELLIGIBILITY

In the previous experiment, we fixed the total glimpse duration to 800 ms, corresponding roughly to 33% of the total duration for most utterances in the IEEE corpus. As shown in Fig. 3, large improvements in intelligibility were observed when the total glimpsing duration increased from 33% to 100% (compare the "infty" condition against all other conditions). This suggests that the total glimpse duration can have a significant effect on intelligibility. For that reason, we examine next the effect of total glimpse duration on performance.

A. Methods

1. Subjects and material

Nine new normal-hearing listeners participated in this experiment. All subjects were native speakers of American English, and were paid for their participation. Subjects age ranged from 18 to 40 years, with the majority being undergraduate students from the University of Texas at Dallas. The speech material consisted of sentences taken from the IEEE database (IEEE, 1969). As in experiment 1, the sentences were corrupted by a 20-talker babble masker (Auditec CD, St. Louis) at −5-dB SNR.

2. Signal processing

The method used to introduce glimpses in the time-frequency plane was the same as that used in experiment 1 (see Fig. 1). Given the relatively weak effect of glimpse window width on performance, we set the glimpse window width to 20 ms for this experiment. Unlike experiment 1, we varied the total glimpse duration to 20%, 30%, 50%, 60%, 70%, 80%, and 100% of the whole utterance. In the 50% condition, for instance, glimpses were introduced in half of the (20-ms) frames in the utterance. The time placement of the glimpses was random. Glimpses were introduced in two different bands, the LF band (0–1 kHz) and the LF+MF band (0–3 kHz). These two bands were chosen as they were found in experiment 1 to yield significant gains in intelligibility (see Fig. 4). To assess any potential gain in intelligibility introduced by glimpsing, we also included as a baseline condition the unmodified noisy sentences (−5-dB SNR). Two sentence lists were used per condition, and none of the lists were repeated.

3. Procedure

The procedure was identical to that used in experiment 1.

B. Results and discussion

The mean scores for all conditions are shown in Fig. 6. Performance was measured in terms of percent of words identified correctly. Two-way ANOVA (repeated measures) indicated a significant effect of total glimpse duration ($F[6,24]=81.5$, $p<0.0005$), a significant effect of frequency band location ($F[1,4]=269.7$, $p<0.0005$), and a significant interaction ($F[6,24]=16.54$, $p<0.0005$).

As expected, performance improved as more glimpses were introduced in both LF and LF+MF conditions. Protected *posthoc* tests (Fisher's LSD) were run to examine at which point (glimpse duration) performance reached an asymptote. Results indicated that, when the glimpses were introduced in the LF band, performance reached an asymptote at 80% of utterance duration. That is, scores obtained with 80% glimpse duration did not differ significantly ($p=0.981$) from those obtained with 100% duration (i.e., whole

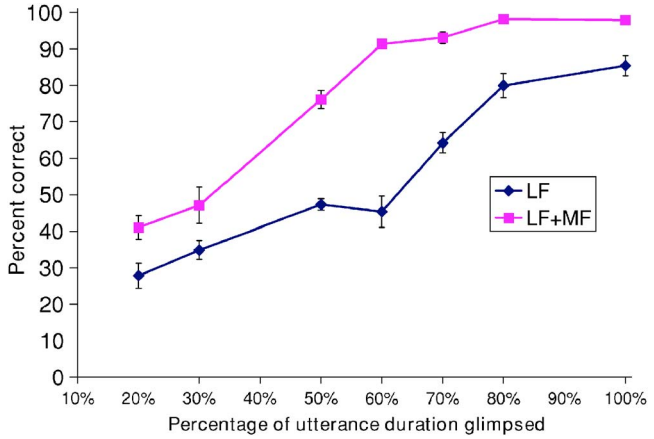


FIG. 6. (Color online) Mean subject recognition performance as a function of the percentage of the utterance glimpsed for two frequency bands. Error bars indicate standard errors of the mean.

utterance) and were significantly ($p<0.05$) higher than all other conditions ($<80\%$). In stark contrast, analysis of the LF+MF scores indicated that the asymptote occurred when glimpsing 60% of the utterance. Performance with glimpsing 100% duration (whole utterance) did not differ significantly ($p=0.093$) from that obtained with glimpsing 60% of the utterance.

The findings of experiment 2 are in close agreement with those of Miller and Licklider (1950). Near-perfect identification was achieved when only 50% of the signal was available for glimpsing during the uninterrupted portions. In their study, the listeners had access to the full clean spectrum of the target signal during the "on" segments of the signal. In our case, listeners had access to the full noisy spectrum but only the LF+MF band was above the SNR threshold and presumably available for glimpsing. For this type of stimuli containing partially masked spectral information, listeners required at least 60% of the total duration of the utterance to obtain high levels of speech understanding.

The results from the present experiment suggest that the extent of the benefit introduced by glimpsing relies heavily on both the total duration of glimpsing and the frequency band glimpsed. This suggests that, in order for CASA and enhancement algorithms to improve speech intelligibility, glimpsing in the LF+MF band needs to occur more than 50% of the time.

IV. CONCLUSIONS

A signal processing technique (Fig. 1) was proposed that can be used as a tool for studying auditory scene analysis and speech segregation in the presence of various types of maskers. Unlike the time-frequency masks used in the previous studies (e.g., Roman *et al.*, 2003; Brungart *et al.*, 2006), the proposed time-frequency mask is not binary but takes real values.

The present study primarily focused on identifying factors that may influence glimpsing speech in noise with the proposed time-frequency mask. Experiment 1 investigated the effect of glimpse window width and frequency location of the glimpse for a fixed duration (33% of utterance) of glimpsing. Experiment 2 investigated the effect of total glimpse duration for two frequency bands. From the results of these two experiments, we can draw the following conclusions:

(1) The frequency location of the glimpses had a significant effect on speech recognition, with the highest performance obtained for the LF+MF band and the lowest for the HF band. Performance with the LF+MF band was found to be as good as performance with the FF band in the majority of the conditions tested.
(2) The glimpse window width and SNR threshold had a relatively minor effect on performance (see Figs. 3 and 5), at least for the range of values considered.
(3) Relative to the unprocessed stimuli (−5-dB SNR), small (10%–15%), but statistically significant, improvements in intelligibility were obtained when the glimpses were available in the LF band, and comparatively larger (20%–30%) improvements were obtained when the glimpses were available in the LF+MF band containing $F1$ and $F2$ information.
(4) Listeners were able to integrate glimpsed information more easily when the glimpses were consistently taken from the same frequency region over time. Performance with the RF band (randomly chosen bands) was significantly lower than performance obtained with the other frequency bands.
(5) The total glimpse duration had the strongest effect in performance. High levels of speech understanding were obtained when more than 60% of the utterance duration was glimpsed in the LF+MF band, at least for the masker (multitalker babble) considered in this study. Relative to the unprocessed sentences (−5-dB SNR), this corresponds to an improvement of 64 percentage points (from 26% to 90%).

The above results have strong implications for speech enhancement and CASA algorithms aiming to improve intelligibility of speech embedded in multitalker babble. For these algorithms to improve speech intelligibility, it is extremely important to improve the spectral SNR in the region of 0–3 kHz (LF+MF band), which is the region containing $F1$ and $F2$ information. Furthermore, it is not necessary to improve the spectral SNR in all frames (i.e., whole utterance), but in at least 60% of the utterance.

ACKNOWLEDGMENTS

This research was supported by Grant No. R01 DC007527 from the National Institute of Deafness and other Communication Disorders, NIH.

APPENDIX A: A TECHNIQUE FOR INTRODUCING GLIMPSES

In this appendix, we describe the signal processing technique used for modifying the masker magnitude spectra to obtain glimpses in specific regions of the spectrum.

We start by expressing the noisy speech spectrum in the frequency domain as follows:

$$Y(\tau,\omega_k) = X(\tau,\omega_k) + N(\tau,\omega_k), \tag{A1}$$

where $Y(\tau,\omega_k)$, $X(\tau,\omega_k)$, $N(\tau,\omega_k)$ are the complex FFT spectra of the noisy speech, clean speech, and masker, respectively, obtained at time (frame) τ and frequency bin ω_k (in our case, multitalker babble was added in experiment 1 to the speech signal at −5-dB SNR). The spectral SNR in time-frequency unit $\{\tau,\omega_k\}$ is given by

$$\xi(\tau,\omega_k) = 10\log_{10}\frac{|X(\tau,\omega_k)|^2}{|N(\tau,\omega_k)|^2}, \tag{A2}$$

where $|\cdot|$ indicates the magnitude spectrum. For all *T-F* units falling within the prescribed frequency region (i.e., the glimpse region), the spectral SNR $\xi(\tau,\omega_k)$ in time-frequency unit $\{\tau,\omega_k\}$ is compared against a threshold, T, and the masker magnitude spectrum is modified accordingly if $\xi(\tau,\omega_k) < T$ or left unaltered if $\xi(\tau,\omega_k) \geq T$. More precisely,

if $\xi(\tau,\omega_k) \geq T$

$Y(\tau,\omega_k) = X(\tau,\omega_k) + N(\tau,\omega_k)$

else

$Y(\tau,\omega_k) = X(\tau,\omega_k) + \hat{N}_M(\tau,\omega_k)$

end, (A3)

where $\hat{N}_M(\tau,\omega_k)$ is the modified masker spectrum, given by

$$\hat{N}_M(\tau,\omega_k) = N(\tau,\omega_k)\cdot 10^{(\xi(\tau,\omega_k)-T)/20}, \tag{A4}$$

and T is the SNR threshold given in decibels. In experiment 1, T was set to 0 dB. The operation described in Eq. (A3) is applied to all *T-F* units falling within the glimpse region. For all *T-F* units falling outside the glimpse region, the following operation is applied to ensure that the spectral SNR $\xi(\tau,\omega_k)$ of the remaining target *T-F* units is below the SNR threshold T:

if $\xi(\tau,\omega_k) < T$

$Y(\tau,\omega_k) = X(\tau,\omega_k) + N(\tau,\omega_k)$

else

$Y(\tau,\omega_k) = X(\tau,\omega_k) + \hat{N}_M(\tau,\omega_k)$

end, (A5)

where $\hat{N}_M(\tau,\omega_k)$ is given by Eq. (A4). The two types of scaling done to the masker spectrum by Eq. (A3) and Eq. (A5) ensure that only the prescribed frequency band contains glimpsing information. After applying Eq. (A3) to all *T-F* units inside the glimpse region and Eq. (A5) for all *T-F* units outside the glimpse region, we reconstruct the noisy speech in frame τ by taking inverse Fourier transform of $Y(\tau,\omega_k)$.

[1]Note that we cannot directly compare the outcome obtained in the HF condition in the present study with that obtained by Anzalone *et al.* (2006). This is because the high-frequency condition tested in the study by Anzalone *et al.* (2006) included all frequencies above 1.5 kHz, whereas in the present study the HF condition included all frequencies above 3 kHz.

Anzalone, M., Calandruccio, L., Doherty, K., and Carney, L. (**2006**). "Determination of the potential benefit of time-frequency gain manipulation," Ear Hear. **27**(5), 480–492.

Brungart, D., Chang, P., Simpson, B., and Wang, D. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**(6), 4007–4018.

Buss, E., Hall, J. W., and Grose, J. H. (**2003**). "Spectral integration of synchronous and asynchronous cues to consonant identification," J. Acoust. Soc. Am. **115**, 2278–2285.

Cooke, M.P., Green, P.D., and Crawford, M.D. (**1994**). "Handling missing data in speech recognition," *Proc. 3rd Int. Conf. Spok. Lang. Proc.*, pp. 1555–1558.

Cooke, M. (**2003**). Glimpsing speech. J. Phonetics **31**, 579–584.

Cooke, M. (**2005**). "Making sense of everyday speech: A glimpsing account," in *Speech Separation by Humans and Machines*, edited by P. Divenyi (Kluwer Academic, Dordrecht), pp. 305–314.

Cooke, M. P., Green, P. D., Josifovski, L., and Vizinho, A. (**2001**). "Robust automatic speech recognition with missing and uncertain acoustic data," Speech Commun. **34**, 267–285.

Cooke, M.P. (**2006**). "A glimpse model of speech perception in noise," J. Acoust. Soc. Am. **119**(3), 1562–1573.

Culling, J., and Darwin, C. (**1994**). "Perceptual and computational separation of simultaneous vowels: Cues arising from low frequency beating," J. Acoust. Soc. Am. **95**, 1559–1569.

Drullman, R. (**1995**). "Speech intelligibility in noise: Relative contribution of speech elements above and below the noise level," J. Acoust. Soc. Am. **98**, 1796–1798.

Festen, J., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Howard-Jones, P.A., and Rosen, S. (**1993**). "Uncomodulated glimpsing in 'checkerboard' noise," J. Acoust. Soc. Am. **93**(5), 2915–2922.

IEEE. (**1969**). "IEEE Recommended Practice for Speech Quality Measurements," IEEE Trans. Audio Electroacoust. **17**(3), 225–246.

Loizou, P. (**2007**). *Speech Enhancement: Theory and Practice* (CRC Press, Taylor Francis Group, Boca Raton, FL).

Miller, G. (**1947**). "The masking of speech," Psychol. Bull. **44**(2), 105–129.

Miller, G.A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**(2), 167–173.

Roman, N., Wang, D., and Brown, G. (**2003**). "Speech segregation based on sound localization," J. Acoust. Soc. Am. **114**, 2236–2252.

Roman, N., and Wang, D. (**2006**). "Pitch-based monaural segregation of reverberant speech," J. Acoust. Soc. Am. **120**, 458–469.

Wang, D. (**2005**). "On ideal binary mask as the computational goal of auditory scene analysis," in *Speech Separation by Humans and Machines*, edited by P. Divenyi (Kluwer Academic, Dordrecht), pp. 181–187.

Oscillation and extinction thresholds of the clarinet: Comparison of analytical results and experiments

Jean-Pierre Dalmont[a] and Cyrille Frappé
Laboratoire d'Acoustique de l'Université du Maine (UMR CNRS 6613), Université du Maine, 72085, Le Mans, France

(Received 23 January 2007; revised 11 May 2007; accepted 13 May 2007)

In the context of a simplified model of the clarinet in which the losses are assumed to be frequency independent the analytic expressions of the various thresholds have been calculated in a previous paper [Dalmont *et al.*, J. Acoust. Soc. Am. **118**, 32.94–3305 (2005)]. The present work is a quantitative comparison between "theoretical" values of the thresholds and their experimental values measured by using an artificial mouth. It is shown that the "Raman" model, providing that nonlinear losses are taken into account, is reliable and able to predict the values of thresholds. [DOI: 10.1121/1.2747197]

PACS number(s): 43.75.Pq [NHF] Pages: 1173–1179

I. INTRODUCTION

This work is an attempt to validate analytical results presented in Ref. 1. In this reference the oscillation and extinction threshold are calculated analytically on the basis of a simple model of the clarinet. This model is the so-called "Raman" model, in which the losses are assumed to be frequency independent. In the present work, the parameters of the model are determined experimentally for some mouthpiece configurations by using an artificial mouth and a procedure described in Ref. 2. Then, the thresholds are measured and compared with the theoretical prediction.

II. THE THEORY OF CLARINET OSCILLATIONS

A. The Raman model

The main characteristic of Raman's model is to consider a perfectly cylindrical resonator in which the dispersion is ignored and the losses are assumed to be frequency independent. These assumptions, as well as for the lossless model,[3] lead to pressure and velocity signals which are perfectly square signals but the introduction of losses avoids some nonphysical results. In particular, the prediction of the lossless model that the amplitude of the oscillation can increase indefinitely when the mouth pressure increases, disappears when losses are included in the model.

Considering frequency independent losses and no dispersion leads to perfectly harmonic resonance frequencies, and the prediction that input impedance values at all resonance frequencies are the same and given by

$$Z = \frac{Z_c}{\tanh \alpha L}, \tag{1}$$

where Z_c is the characteristic impedance, α the damping factor (assumed to be constant in the context of Raman's model), and L the effective length of the clarinet, that is $L = c/4f$ with c the speed of sound and f the playing frequency. The value of α is taken equal to the theoretical value (considering visco-thermal losses only) at the playing frequency.

The reed is considered to be an ideal spring without any damping and inertia and the reed is supposed to close the reed channel suddenly for a given pressure p_M. The reed channel opening is assumed to be rectangular and equal to $S = wH$ with w the constant width and H the variable opening height. The opening height is related to the pressure difference Δp between the two sides of the reed by

$$\begin{aligned} &H = H_0 - \Delta p/K \quad \text{for } \Delta p \leqslant p_M, \\ &H = 0 \quad \text{for } \Delta p \geqslant p_M, \end{aligned} \tag{2}$$

where H_0 is the opening height at rest and K the stiffness (pressure/displacement) of the reed. It can be verified that

$$p_M = KH_0. \tag{3}$$

The overpressure in the mouth is the cause of a jet directed toward the pipe which kinetic energy is assumed to be completely dissipated into turbulence which leads[4] to

$$\Delta p = \tfrac{1}{2}\rho \frac{v^2}{L}, \tag{4}$$

where ρ is the density of the air and v the velocity of the jet. The pipe cross section is assumed to be equal to the reed channel opening and the positive volume flow u at the input of the pipe is then given by

$$u = wH\sqrt{\frac{2\Delta p}{\rho}}. \tag{5}$$

Combining Eqs. (2), (3), and (5) leads to the nonlinear characteristics of the embouchure:

$$\begin{aligned} &u = u_A\left(1 - \frac{\Delta p}{p_M}\right)\sqrt{\frac{\Delta p}{p_M}} \quad \text{for } \Delta p \leqslant p_M, \\ &u = 0 \quad \text{for } \Delta p \geqslant p_M, \end{aligned} \tag{6}$$

where

[a]Author to whom correspondence should be addressed: Electronic mail: jean-pierre.dalmont@univ-lemans.fr

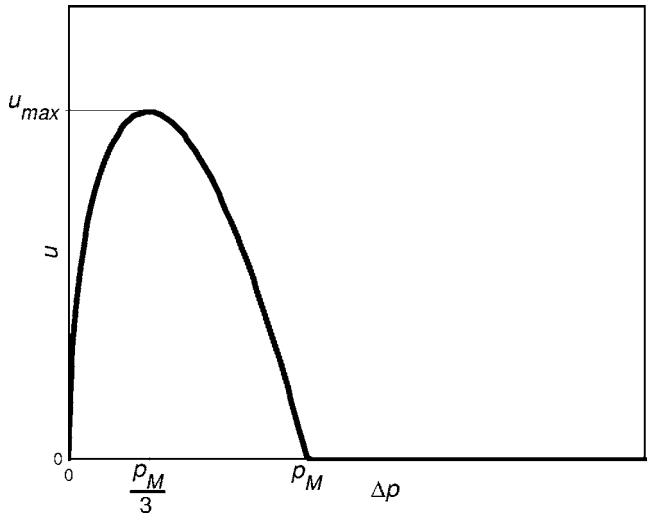


FIG. 1. Theoretical nonlinear characteristic (volume flow vs pressure difference [Eq. (6)]) with specific points.

$$u_A = w\sqrt{\frac{2KH_0^3}{\rho}}. \tag{7}$$

Volume velocity is a function of the three parameters Δp, u_A, and p_M. It is plotted versus Δp in Fig. 1 for given values of u_A, and p_M. It depends on two parameters: the beating pressure p_M and a volume flow coefficient u_A. It can be noticed that the maximum of this function is obtained for $\Delta p = p_M/3$ and

$$u_{\max} = \frac{2}{3\sqrt{3}}u_A. \tag{8}$$

B. The solutions

On the basis of the Raman model, described in Sec. II A, solutions are square signals whose amplitudes can be calculated analytically.[1] A theoretical bifurcation diagram can be obtained that is the amplitude of the pressure in the mouthpiece p as a function of the pressure in the mouth p_m (Fig. 2). On this bifurcation diagram some specific points, named thresholds, can be marked out. The theoretical values of these thresholds, given in Ref. 1, are recalled in the following.

1. Oscillation threshold

The oscillation threshold corresponds to the point for which oscillation starts. It is given by

$$p_{m\text{th}} = \frac{1}{9}(\beta_1 + \sqrt{3+\beta_1^2})^2 p_M, \tag{9}$$

where β_1 is a nondimensional parameter equal to

$$\beta_1 = \frac{1}{Z}\frac{p_M}{u_A}. \tag{10}$$

In the absence of losses, $\beta_1 = 0$ and the oscillation threshold is given by $p_{m\text{th}} = p_M/3$. When β_1 increases, the oscillation threshold tends to p_M. If $\beta_1 > 1$, no oscillations occur for any mouth pressure: the instrument is unplayable. This occurs if losses are too large, the reed too stiff, or the opening too small.

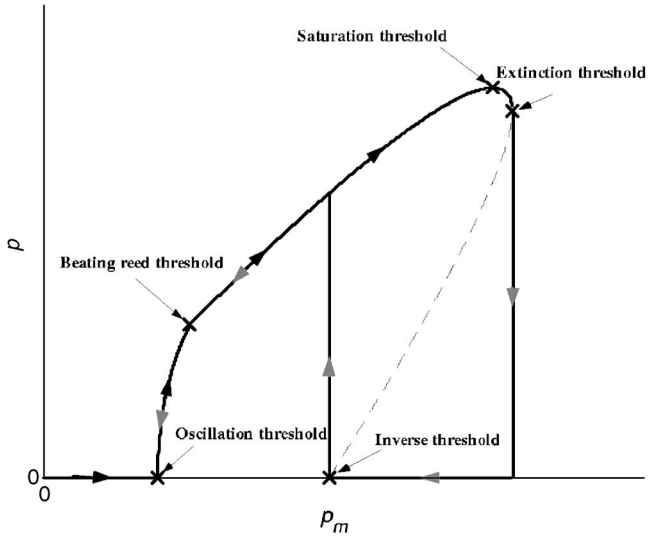


FIG. 2. Typical bifurcation diagram (pressure in the mouthpiece vs mouth pressure) with indication of the various thresholds and hysteretic cycle. Dotted line is an unstable branch.

2. beating threshold

The beating reed threshold marks the transition between the nonbeating reed regime and the beating reed regime. The beating reed regime corresponds to an oscillation during which the reed channel is closed during half of the time period. The beating reed threshold is given by

$$p_{m\text{b}} = \left(\tfrac{1}{2} + \tfrac{1}{2}[\beta\beta_1 + \beta_1^2(1-\beta\beta_1)]\right)p_M, \tag{11}$$

with

$$\beta = \frac{u_A}{p_M} Z_c \tanh \alpha L. \tag{12}$$

In the absence of losses $p_{mb} = p_M/2$. When β_1 increases the oscillation threshold $p_{m\text{th}}$ and the beating reed threshold p_{mb} both tend to p_M.

3. Saturation threshold

The saturation threshold corresponds to the pressure for which the amplitude of the oscillation is maximum. It is given by

$$p_{m\text{sat}} = \left[1 + \frac{2}{\sqrt{3}\beta_3}\right]\frac{p_M}{3}, \tag{13}$$

where

$$\beta_2 = \frac{2\beta_1}{1+\beta\beta_1} = \frac{p_M}{Z_c u_A}\tanh 2\alpha L. \tag{14}$$

In the absence of losses this threshold tends to infinity. It can be noticed that the extinction threshold can be larger that the beating pressure p_M which leads to a hysteretic bifurcation diagram (see Fig. 2). When β_1 increases this threshold tends to be equal to the beating reed threshold.

4. Extinction threshold

Up to the saturation threshold, the pressure in the mouthpiece decreases despite the pressure in the mouth increasing. When the mouth pressure reaches the extinction threshold, oscillations stop. This threshold is given by

$$p_{m\text{ext}} = \left(\frac{1}{9} + \frac{2}{27\beta_2}(3+\beta_2^2)[\beta_2 + \sqrt{3+\beta_2^2}]\right)p_M. \quad (15)$$

Equation (10) is only valid for $\beta_2 \leqslant 1$. If $\beta_2 \geqslant 1$, $\Delta p_{\text{ext}} = p_{m\text{ext}} = p_M$ and the hysteresis cycle in the bifurcation diagram disappears.

5. Inverse threshold

The inverse threshold corresponds to the point for which oscillations start when pressure is decreased after the reed has been blocked on the lay. It is equal to the beating pressure:

$$p_{m\text{inv}} = p_M. \quad (16)$$

This threshold is an inverse oscillation threshold for decreasing pressure if $\beta_2 \leqslant 1$ and a direct oscillation threshold if $\beta_2 \geqslant 1$.

In practice parameter β is small compared to unity and can be neglected. Then $\beta_2 \approx 2\beta_1$ and all the threshold values depend on p_M and β_1. This parameter can be seen as a parameter characterizing the good matching of the embouchure, characterized by p_M/u_A, with the resonator characterized by its impedance at resonances Z. If this parameter is larger than unity no oscillation is possible, that is the instrument is unplayable.

III. EXPERIMENTS

In order to evaluate the validity of the previous results, a series of experiments has been realized by using an artificial mouth allowing the measurement of both the pressure in the mouth p_m and the pressure in the mouthpiece p. A force sensor, made with strain gauges stuck on the support of the lip, measures the force of the artificial lip on the reed in order to have a control of the reed opening. For a given embouchure (that is a given set of reed parameters) the nonlinear characteristic is measured using the procedure described in Ref. 2. Then, for the same embouchure, the bifurcation diagram is recorded and the thresholds are determined. From the nonlinear characteristic, the identification of the embouchure parameters, p_M and u_A, allows the calculation of the thresholds, which values can be compared to the experimental values. This experiment is repeated for different values of the force exerted by the artificial lip on the reed.

A. Nonlinear characteristics measurement

The nonlinear characteristic is obtained by using the method described by Dalmont *et al.*[2] in which the pipe is replaced by an orifice allowing the determination of the volume velocity. The measurement is quasistatic; that is, the pressure is slowly varied and no oscillation occurs. To avoid oscillations a mass of paste is fixed on the tip of the reed. It is verified that this mass does not modify the measured characteristic. This method has also been used by Almeida *et al.*[5] The mouthpiece is a RV40 by Vandoren and the reed a plasticover by Rico. The characteristic is obtained first for an increasing pressure beyond the beating pressure. Then, the pressure is decreased and a second characteristic is obtained. A typical result is plotted in Fig. 3 for increasing pressure. As was already observed in Refs. 2 and 5, due to the viscoelastic behavior of the reed, the two characteristics, for increasing and decreasing pressure, are different. As discussed in the following it seems that in dynamic situations the effective nonlinear characteristic is the characteristic for decreasing pressure. This experiment is repeated for various forces on the reed measured with a force sensor. The various forces result in various initial reed opening values H_0. Results are plotted in Fig. 4 for increasing pressure. It is observed that beyond the beating pressure, the closing of the reed channel is not perfect, some residual volume flow being observed. It is noticeable that this residual flow does not depend on the

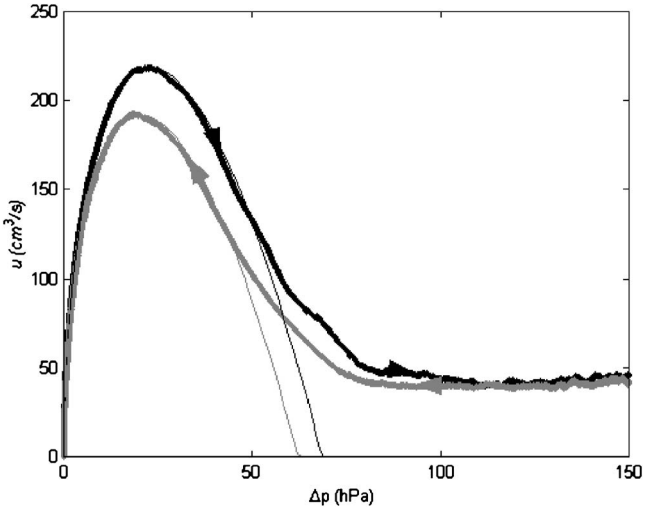


FIG. 3. Typical nonlinear characteristic (volume flow vs pressure difference) for increasing and decreasing mouth pressure (force on the reed is 2.24 N). Thick black line: increasing pressure; Thick grey line: decreasing pressure; and Thin line: model [Eq. (6)].

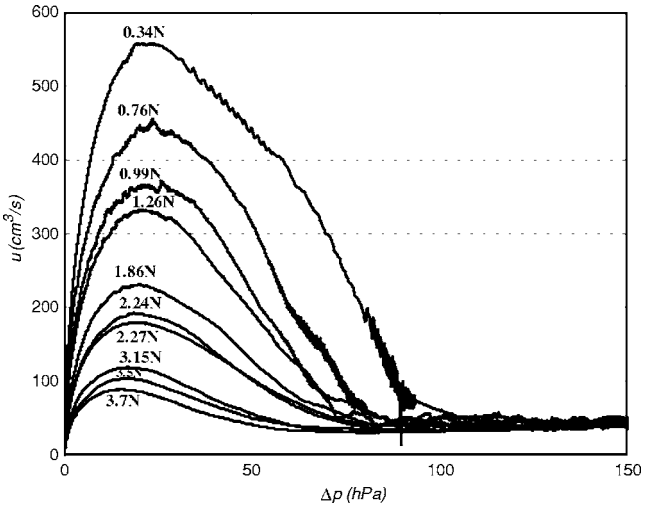


FIG. 4. Experimental nonlinear characteristics (volume flow vs pressure difference) for various forces on the reed (10 values) for a decreasing mouth pressure (from 150 hPa to 0).

TABLE I. Embouchure and reed parameters deduced from the nonlinear characteristics measurements.

Force on the reed (in N)	$p_M = kH_0$ (hPa)		u_{max} (en cm^3/s)		H_0 (mm)		Stiffness K (hPa/mm)	
	Increasing pressure	Decreasing pressure	Increasing pressure	Decreasing pressure	Increasing pressure	Decreasing pressure	Increasing pressure	Decreasing pressure
0.34	86.1	73.8	634	562	1.14	1.10	75	67
0.76	79.3	74.4	479	443	0.90	0.86	88	86
0.99	75.7	71.0	394	366	0.76	0.73	100	97
1.26	72.9	66.7	363	329	0.71	0.68	102	99
1.86	64.0	61.1	255	229	0.54	0.49	120	125
2.24	66.6	60.3	219	190	0.45	0.41	148	147
2.27	66.5	58.6	207	187	0.43	0.41	156	143
3.15	57.0	51.7	136	119	0.30	0.28	188	186
3.51	54.5	50.6	120	103	0.27	0.24	200	208
3.71	50.8	48.3	103	90	0.24	0.22	210	223

initial reed opening. As the coordinates of the maximum of the theoretical characteristic are $(p_M/3, 2/(3\sqrt{3})u_A)$, the values of parameters p_M and u_A could be deduced from the determination of the maximum of the measured characteristic. However, it was found more accurate to obtain these parameters by fitting the theoretical curve on the experimental one using a minimization process (see Fig. 3). Results for increasing and decreasing pressure are given in Table I.

B. Bifurcation diagram measurements

In this experiment the mass of paste on the tip of the reed is removed and the orifice is replaced by a cylindrical tube of diameter 16 mm and length 50 cm. The procedure is then similar to the previous one: starting from 0 the mouthpressure p_m is increased until oscillations start, increased again until oscillations stop, the reed being blocked on the lay. The mouthpressure is then decreased until oscillation starts and decreased again until oscillation stops. During the experiment, the mouthpressure p_m and the pressure in the mouthpiece p are recorded, and the rms value of the pressure in the mouthpiece is plotted as a function of the mouthpressure. This experiment is repeated for each embouchure for which the nonlinear characteristics have been measured. Results are plotted in Figs. 5(a) and 5(b). It is noticeable that the curves for increasing pressure and decreasing pressure are almost superimposed below the inverse threshold. This suggests that the viscoelastic deformation observed on the characteristics is always present when the reed oscillates. So, we hypothesize that the reed parameters needed for the calculation of the various thresholds are that obtained from the characteristic with a decreasing pressure. As the measurement is not strictly stationary, some transitory regimes can be observed. This especially the case for the oscillation threshold for which the threshold value has to be extrapolated, leading to a rather large uncertainty. For some embouchures, it can be observed that after the extinction threshold oscillations do not stop but bifurcate to another regime which can no longer be assimilated to a square signal (see Appendix B). Due to this, the determination of the extinction threshold is less accurate than that of the saturation threshold which corresponds to the maximum amplitude of the pressure in the mouthpiece. The beating reed threshold has not been determined because it does not correspond to a well-defined point of the experimental bifurcation diagram. Results for the various embouchures are given in Table II.

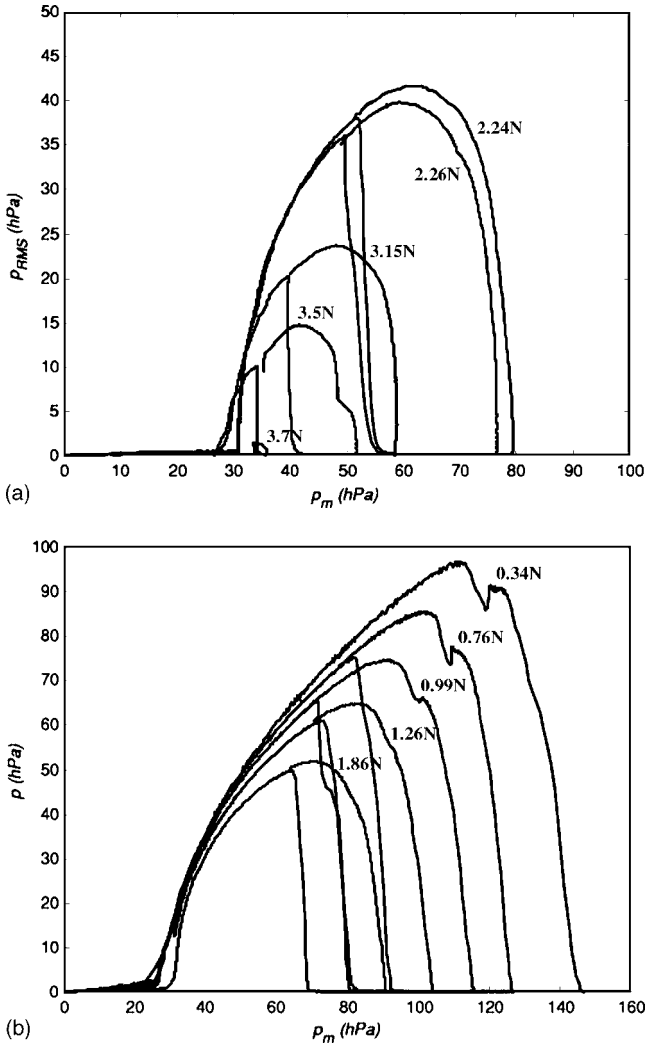


FIG. 5. (a), (b) Experimental bifurcation diagrams (rms value of the pressure in the mouthpiece vs mouth pressure) for various forces on the reed (10 values) for increasing and decreasing mouth pressure.

TABLE II. Thresholds deduced from experimental bifurcation diagrams.

Force (N)	Oscillation threshold (hPa)	Inverse threshold (hPa)	Saturation threshold (hPa)	Extinction threshold (hPa)
0.34	28-30	82	111	116
0.76	29-30	82	101	107
0.99	27	70	90	95
1.26	29	70	83	92
1.86	32	65	71	76
2.24	30,5	52	62	66
2.27	31-32	50	59	69
3.15	29	40	48	52
3.51	29	34	43	48
3.71	NA[a]	NA[a]	(34)	36

[a]Not available.

IV. COMPARISON OF THEORETICAL AND EXPERIMENTAL THRESHOLDS

In this section, "theoretical" threshold values calculated by using the values of parameters obtained from the experimental characteristics for decreasing pressure (Table I) are compared to the values obtained from the experimental bifurcation diagrams (Table II). The thresholds values are given in Figs. 6 and 7 versus the effective reed channel opening height H_0 which is not measured directly but deduced from p_M and u_A ($H_0=(u_A/w)\sqrt{(\rho/2p_M)}$, with $w=12$ mm).

A. Oscillation threshold

This threshold is close to 30 hPa for any embouchure (see Figs. 4 and 6). This is explained by the fact that the beating pressure $p_M=KH_0$ varies slightly with the embouchure. Indeed, when the force on the reed is increased, the reed opening decrease is partly compensated by a reed stiffness increase. Theory expects a slight decrease of the oscillation thresholds when the reed opening decreases which is not observed in the experiments. This difference might be explained by the fact that Raman's model is not well adapted near the oscillation threshold, in the vicinity of which quasisinusoidal pressure signals are observed.

B. Inverse oscillation threshold

This threshold is close to the beating reed pressure p_M for large opening as expected from theory (Fig. 6). However, a discrepancy is observed for small openings. This is an illustration of the limits of the model: in the vicinity of the beating pressure p_M the theoretical characteristic differs significantly from the experimental one (see Fig. 4). For large force values the inverse threshold tends to the oscillations threshold and for larger force values ($F>3.7$ N) no oscillation occurs.

C. Saturation and extinction threshold

Saturation thresholds calculated with the model are somewhat higher than those measured with the artificial blowing especially for large reed openings. This discrepancy is interpreted as a consequence of vortex shedding at the end of the pipe. This phenomenon was first mentioned by Bouasse[6] and Benade and Keefe designed an unplayable clarinet with holes of small diameter leading to important nonlinear losses by vortex shedding.[7] As shown by Atig *et al.*,[8] it is necessary to take into account nonlinear losses at the end of the pipe, as they reduce significantly the values of the saturation and extinction thresholds (Fig. 7). This effect is especially noticeable for large reed openings for which high acoustic velocities at the output are encountered. It is possible to include simply nonlinear losses in Raman's model by assuming these losses to be proportional to the squared velocity amplitude at the end of the pipe (see Appendix A). This leads to Eq. (A10) in which the nonlinear losses are proportional to a parameter c_d depending on the pipe end geometry. In Atig *et al.*,[8] some experimental values of c_d are given. For the pipe used in the experiments with a small radius of curvature (less than 0.5 mm) a value of c_d

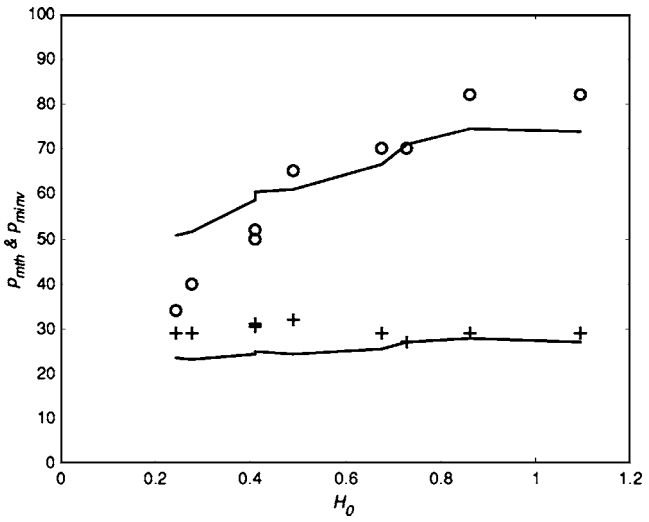


FIG. 6. Oscillation threshold and inverse threshold vs effective reed opening H_0. Upper line: "theoretical" inverse threshold [Eq. (16)]. Open circle: experimental inverse threshold. Lower line: theoretical oscillation threshold [Eq. (9)]. Plus: experimental oscillation threshold.

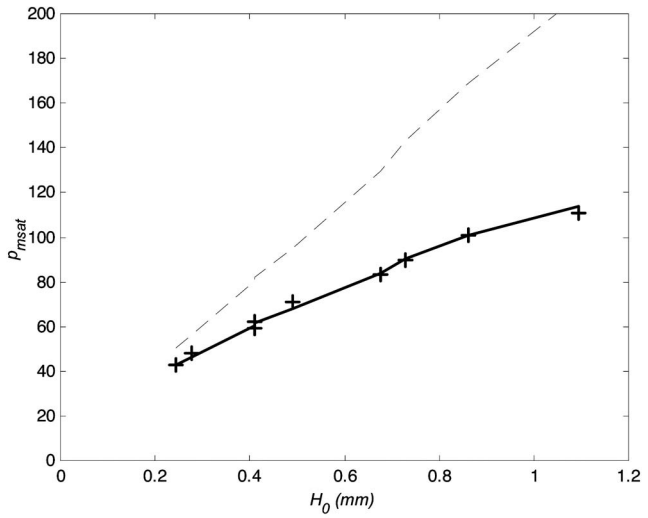


FIG. 7. Saturation threshold vs effective reed opening H_0. Dashed line: Theory with linear losses only [Eq. (13)]. Continuous line: theory including nonlinear losses [Eq. (A10)]. Asterisk: experiments.

=1.7 is plausible. However, the best fit with experiment is obtained for a significantly larger value c_d=2.8 (Fig. 7). Two explanations can be given: first, the value obtained by Atig *et al.*[8] is for a sinusoidal excitation signal different from that observed in a clarinet; second, in the Raman model, the losses for higher harmonics are underestimated. A result very similar to that of Fig. 7 is obtained with linear losses increased by 20% and c_d=2.3.

The extinction threshold is close to the saturation threshold as could be expected. It must be pointed out that the extinction threshold for a quasisquare signal is different from the effective extinction threshold because of a bifurcation to another regime (see Appendix B).

V. CONCLUSION

Present experiments emphasize the high reliability of the Raman model providing that nonlinear losses at the end of the pipe are included in the model. It may seem surprising that so simple a model, based on strong simplifications (square signals), is able to predict the values of the thresholds with so good an accuracy. This suggests that parameters missing in the model, such as the reed damping and inertia, have a minor influence on the auto-oscillation process and the playing dynamics. On the other hand, the present work suggests that any realistic model of the clarinet should include nonlinear losses in the side holes.

ACKNOWLEDGMENTS

The authors acknowledge Murray Campbell, Joël Gilbert, Jean Kergomard, Kees Nederveen, and Sébastien Ollivier for fruitful discussions and corrections.

APPENDIX A: SATURATION THRESHOLD WITH NONLINEAR LOSSES

Nonlinear losses at the end of the pipe can be included in Raman's model by replacing the input impedance Z, given by Eq. (1), by an impedance taking account a resistive impedance Z_t at the end of the pipe:

$$Z = \frac{Z_c}{\tanh(\alpha L + Z_t/Z_c)} \tag{A1}$$

assuming $\alpha L + Z_t/Z_c \ll 1$ this leads to

$$Z \approx \frac{Z_c}{\alpha L + Z_t/Z_c}. \tag{A2}$$

Atig *et al.*[8] (see also Ref. [9]) have shown that the nonlinear losses at the end of the pipe can be approximated by the following formula:

$$Z_t = \frac{c_d}{4}\frac{|v_t|}{c} Z_c, \tag{A3}$$

where v_t is the amplitude of the acoustic velocity at the end of the pipe, c the speed of sound, and c_d a nondimensional coefficient depending on the output geometry. The velocity amplitude v_t at the end of the pipe is related to the pressure amplitude p at the input (i.e., in the mouthpiece) by

$$v_t = p/\rho c. \tag{A4}$$

So, the termination impedance can be written as a function of the mouthpiece pressure amplitude p:

$$Z_t = \frac{c_d}{4}\frac{p}{\rho c^2} Z_c. \tag{A5}$$

Equation (A2) can consequently be written:

$$Z = \frac{Z_c}{\alpha L + c_d p/(4\rho c^2)}. \tag{A6}$$

At the saturation threshold the volume velocity u_1 at the input, when the reed channel is open, is equal to the maximum volume velocity of the nonlinear characteristic $u_{\max}$ given by Eq. (8). The volume velocity amplitude being half of the peak-peak value, the input pressure amplitude p_{sat} at the saturation threshold is then given by

$$p_{\text{sat}} = Z u_{\max}/2. \tag{A7}$$

Combining Eqs. (A6) and (A7) leads to

$$p_{\text{sat}} = \frac{Z_c u_{\max}}{2[\alpha L + c_d p_{\text{sat}}/(4\rho c^2)]}. \tag{A8}$$

The resolution of this second-order equation gives p_{sat}:

$$p_{\text{sat}} = \frac{2\rho c^2}{c_d}\left[\sqrt{(\alpha L)^2 + \frac{c_d}{2\rho c^2} Z_c u_{\max}} - \alpha L\right]. \tag{A9}$$

At the saturation threshold, when the reed channel is open $\Delta p = p_M/3$, since $p_m = \Delta p + p$ the saturation threshold is finally deduced from Eq. (A9) and given by

$$p_{m\text{sat}} = \frac{p_M}{3}\left(1 + 2A\left[\sqrt{1 + \frac{2}{\sqrt{3}\beta_2 A}} - 1\right]\right), \tag{A10}$$

where $A = 3\rho c^2/p_M c_d \alpha L$. It can be verified that when c_d tends toward zero, Eq. (A10) leads to Eq. (13).

APPENDIX B: EVOLUTION OF THE PRESSURE SIGNAL NEAR THE EXTINCTION THRESHOLD

In this appendix it is shown that the shape of the pressure signal in the mouthpiece can be modified before extinction. The corresponding signals can no longer be assimilated to a quasisquare signal. In our experiments, some bifurcations to other regimes can be observed. An example of such bifurcation is given in Figs. 8 and 9. This corresponds to the

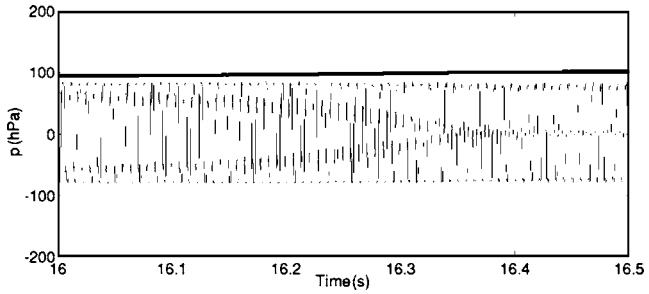


FIG. 8. Bifurcation from a quasisquare signal to another shaped signal (experiment). Thin line: mouthpiece pressure signal. Thick line: mouth pressure signal.

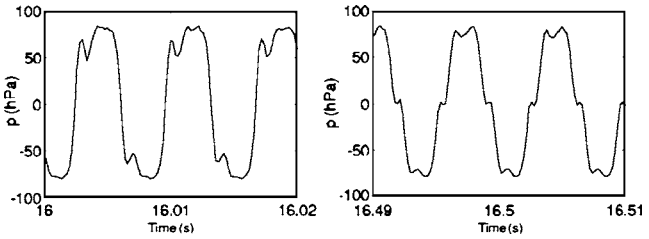


FIG. 9. Signals before and after bifurcation (experiment).

case with a force on the reed of 0.99N [third curve of Fig. 5(b)] at a mouth pressure varying between 94 and 101 hPa, that is after the saturation threshold. This bifurcation is clearly visible in Fig. 5(b): there is a kink after the saturation threshold and before the effective extinction. It is hypothesized that this bifurcation corresponds to the extinction threshold of the square signal. It must be specified that such a bifurcation is not systematic and other shapes of nonsquare signals can in some cases be observed.

[1]J. P. Dalmont, J. Gilbert, J. Kergomard, and S. Ollivier, "An analytical prediction of the oscillation and extinction thresholds of a clarinet," J. Acoust. Soc. Am. **118**, 3294–3305 (2005).

[2]J.-P. Dalmont, J. Gilbert, and S. Ollivier, "Non-linear characteristics of single reed instruments: Quasi-static volume flow and reed opening measurements," J. Acoust. Soc. Am. **114**, 2253–2262 (2003).

[3]J. Kergomard, "Elementary considerations on reed-instrument oscillations," in *Mechanics of Musical Instruments*, CISM Course and Lectures No. 355 (Springer, New York, 1995).

[4]T. A. Wilson and G. S. Beavers, "Operating modes of the clarinet," J. Acoust. Soc. Am. **56**, 653–658 (1974).

[5]A. Almeida, C. Vergez, and R. Caussé, "Quasi-static nonlinear characteristics of double-reed instruments," J. Acoust. Soc. Am. **121**, 536–546 (2007).

[6]H. Bouasse, *Instruments à Vent* (Librairie Delagrave, Paris 1929; reprint Blanchard, Paris 1985), Vols. **I** and **II**.

[7]D. H. Keefe, "Acoustic streaming, dimensional analysis of nonlinearities, and tone hole mutual interactions in woodwinds," J. Acoust. Soc. Am. **73**, 676–687 (1982).

[8]M. Atig, J.-P. Dalmont, and J. Gilbert, "Saturation mechanism in clarinet-like instruments, the effect of the localized non-linear losses," Appl. Acoust. **65**, 1133–1154 (2004).

[9]J. H. M. Disselhorst and L. Van Wijngaarden, "Flow in the exit of open pipes during acoustic resonance," J. Fluid Mech. **99**, 293–319 (1980).

Characterization of dense bovine cancellous bone tissue microstructure by ultrasonic backscattering using weak scattering models

D. D. Deligianni[a] and K. N. Apostolopoulos
Biomedical Engineering Laboratory, Department of Mechanical Engineering and Aeronautics, University of Patras, Rion 26500, Greece

(Received 1 November 2006; revised 22 May 2007; accepted 23 May 2007)

A weak scattering model was proposed for the ultrasonic frequency-dependent backscatter in dense bovine cancellous bone, using two autocorrelation functions to describe the medium: one with discrete homogeneities (spherical distribution of equal spheres) and another, which considers tissue as an inhomogeneous continuum (densely populated medium). The inverse problem to estimate trabecular thickness of bone tissue has been addressed. A combination of the two autocorrelation functions was required to closely approximate the backscatter from bovine bone with various microarchitecture, given that the shape of trabeculae ranges from a rodlike to a platelike shape. Because of the large variation in trabecular thickness, both at an intraspecimen and an interspecimen level, thickness distributions for individual trabeculae for each bone specimen were obtained, and dominant trabecular sizes were determined. Comparison of backscatter measurements to theoretical predictions indicated that there were more than one dominant trabecular sizes that scatter sound for most specimens. Linear regression, performed between dominant trabecular thickness and estimated correlation length, showed significant linear correlation ($R^2=0.81$). Attenuation due to scattering by a continuous distribution of scatterers was predicted to be linear over a frequency range from 0.3 to 0.9 MHz, suggesting a possibility that scattering may be a significant source of attenuation.
 [DOI: 10.1121/1.2749461]

PACS number(s): 43.80.Cs, 43.20.Fn, 43.80.Ev, 43.80.Jz [CCC] Pages: 1180–1190

I. INTRODUCTION

Ultrasound is currently being assessed as an alternative method of evaluating bone quality, following reports that it provides information about structure in addition to density. Despite the diagnostic utility, the fundamental mechanisms underlying the interaction between ultrasound and cancellous bone are not well understood presently.

In general, and particularly in soft tissues, ultrasonic backscatter is known to provide information regarding size, shape, number density, and elastic properties of scatterers (Insana, 1995; Lizzi *et al.*, 1986; Oelze and O'Brien, 2002). Ultrasonic backscattering has shown its potential to characterize bone microarchitecture (Chaffaï *et al.*, 2002; Insana *et al.*, 1990; Wear, 2003). In cancellous bone applications, trabeculae or marrow (Luppé *et al.*, 2002) can be regarded as possible scattering sites due to high contrast in acoustic properties between mineralized trabeculae and marrow. Measurements of ultrasonic backscatter coefficient from human cancellous bone have been reported in vitro and in vivo (Chaffaï *et al.*, 2000; Hakulinen *et al.*, 2006; Wear, 1999; Wear *et al.*, 2000). These results suggest that ultrasonic backscattering may contain substantial information not already contained in ultrasound properties obtained by transmission measurements.

Two different approaches have been proposed to model backscattering from cancellous bone. A first approach, proposed by Wear (1999, 2003, 2004) uses the analytical model of Faran, which provides an exact solution of the backscatter cross section by a spherical or a cylindrical solid elastic object, to describe the interaction of trabecular bone with the ultrasonic wave. Trabeculae were modeled as long, thin cylinders. Backscatter measurements were in good agreement with theoretical predictions at lower frequencies. Chaffaï *et al.* (2000) also found good agreement of experimental data with cylindrical and spherical Faran models.

The second approach (Chernov, 1960; Morse and Ingard, 1968) is based on Chernov's proof that scattering is proportional to the product of the mean compressibility fluctuations and the autocorrelation function integrated over volume. It considers the medium as a fluid random continuum and has been successfully employed for soft tissue characterization. The scatterers are described as source terms that perturb the homogeneous wave equation in an ambient fluid. This approach assumes weak scattering (Born approximation). The fluctuations of the medium acoustical properties (density and compressibility) are described by their second order statistical properties, such as the autocorrelation function. Autocorrelation functions used in the literature to model scattering in soft tissues are the spherical distribution, Gaussian, and exponential functions (Insana, 1995; Lizzi *et al.*, 1986; Oelze *et al.*, 2002). A general assumption for these models is that the scatterers are weak, randomly and sparsely distributed.

Strelitzki *et al.* (1998) and Nicholson *et al.* (2000) presented a scattering model using velocity fluctuations in a

[a]Author to whom correspondence should be addressed. Electronic mail: deligian@mech.upatras.gr

binary mixture based on continuum approach and an exponential autocorrelation function to describe the statistical properties of the medium. However, no comparison with experimental data was performed.

Jenson *et al.* (2003) used various autocorrelation functions (Gaussian, exponential and densely populated medium) to compute backscatter coefficient and found good agreement between experimental data and theoretical predictions. The best prediction was achieved with the Gaussian autocorrelation function, although large differences among these functions were not observed. A significant relationship was found between the estimated correlation length from these models and the mean trabecular thickness (Chaffaï *et al.*, 2002; Jenson *et al.*, 2003; Wear, 2003), but a moderate correlation at the individual level (Padilla *et al.*, 2006).

Ultrasound backscatter is one of the two components of attenuation of the ultrasound wave traveling through trabecular bone, representing the conserved and redistributed energy in the wave. The other component is the absorption process in which some of the wave energy is converted into internal energy of the structure. Attenuation is widely reported to vary approximately as frequency to the first power (Glüer, 1997; Langton *et al.*, 1984; Tavakoli and Evans, 1991). Backscatter is found to vary as frequency to the third power. Consistency of these two findings leads to the conclusion that absorption may be a greater component of attenuation than scattering in human calcaneous bone (Chaffaï *et al.*, 2000; Wear, 1999), although there are indications that scattering may be a significant source of attenuation (Kaufman *et al.*, 2003; Nicholson and Bouxsein, 2002).

All the above works have studied backscatter from human trabecular bone and, in particular, from osteoporotic calcaneous bone obtained from older subjects with very low density and thin trabeculae. This is justified by the fact that it provides information for diagnostic purposes in humans (osteoporosis). Healthy or younger human bone, and even more bovine bone, which are denser and are comprised of thicker trabeculae, may display a different scattering behavior.

Cancellous bone is a heterogeneous material composed of a solid matrix (mineralized collagen) of interconnected trabeculae filled with a fluid-like medium (marrow). Between human and bovine cancellous bone there are structural and compositional differences. The trabecular diameter of bovine bone ranges between 90 and 400 μm or, sometimes, it is even larger (Simmons and Hipp, 1997). In human bone the respective range is between 50 and 230 μm (Trebacz and Natali, 1999; Thomsen *et al.*, 2002). Because of the large mean trabecular thickness of bovine bone and its large variation, it is a suitable model to study ultrasound scattering and test scattering theories, toward a better understanding of the ultrasound interaction with bone.

The objective of the current work is to study the frequency-dependent ultrasonic attenuation and backscatter coefficient of dense bovine cancellous bone with large variation in trabecular thickness, both at an intraspecimen and an interspecimen level. A weak scattering theoretical model for ultrasonic backscattering is presented and tested experimentally. A combination of two autocorrelation functions is proposed to describe the medium: one with discrete homogeneities (spherical distribution of equal spheres) and another, which considers tissue as an inhomogeneous continuum (densely populated medium). The potential of these models to predict ultrasonic backscattering from cancellous bone and estimate trabecular thickness is investigated.

Dominant trabecular sizes were determined from thickness distributions of trabeculae for each specimen instead of a mean thickness (of all bone specimens) and the applied theoretical model predicted these (more than one) dominant sizes that scatter sound.

Although the results of the study of bovine bone ultrasonic behavior are not directly applicable for diagnostic purposes, they are useful for the understanding of the ultrasonic behavior of bone in a wide range of densities, toward a better understanding of the relationship of frequency-dependent backscatter and bone microarchitecture. Study of ultrasound backscatter can elucidate mechanisms, responsible for attenuation of the ultrasound wave traveling through cancellous bone, and the respective roles of absorption and scattering phenomena. The determined in this study attenuation due to scattering suggested a possible significant role of scattering in total attenuation.

II. THEORETICAL BACKGROUND

A. Backscatter coefficient

Cancellous bone can be modeled as randomly positioned impedance fluctuations contained within an acoustically uniform material. For statistically homogeneous and isotropic random media, with size of inhomogeneities much smaller than the size of the scattering volume V, the backscatter coefficient, σ_b ($\mathrm{cm}^{-1}\,\mathrm{sr}^{-1}$), which represents the fraction of the incident intensity scattered at 180° to the propagation direction for incoherent scattering, is expressed as (Insana and Brown, 1993)

$$\sigma_b = \frac{k^3\langle\gamma^2\rangle}{8\pi}\int_0^\infty b_\gamma(\Delta r)\sin(2k\Delta r)\cdot\Delta r\cdot d(\Delta r), \tag{1}$$

where k is the wave number, $\langle\gamma^2\rangle$ is the mean square fluctuation in medium properties, and $b_\gamma(\Delta r)$ is the autocorrelation function for the scattering medium as a function of Δr, the vector connecting two points in the medium.

A model that has been used in the literature considers tissue to comprise a spatial distribution of identical fluid spheres of radius a randomly distributed within another fluid of different density and compressibility, with relatively low concentration. In this case, the autocorrelation function is determined by the three-dimensional (3D) autocorrelation function of a single sphere and the backscatter coefficient is (Lizzi *et al.*, 1986)

$$\sigma_{b1} = \frac{k^4\langle\gamma^2\rangle\alpha^3}{12\pi}\left(\frac{3}{2k\alpha}J_1(2k\alpha)\right)^2, \tag{2}$$

where J_1 is the first order spherical Bessel function of the first kind.

In bovine cancellous bone, the scattering particles (trabeculae) are not sparsely distributed. Consequently, the mean values of density and compressibility of bovine cancellous

bone are not those of the surrounding the scatterers' homogeneous fluid, but the scatterers' properties contribute to these mean values. A densely populated medium containing weak scatterers will be assumed for bovine cancellous bone. Morse and Ingard (1968) have suggested a modified Gaussian correlation model to describe isotropic, dense random media. The corresponding backscatter coefficient is (Insana and Brown, 1993)

$$\sigma_{b2} = \frac{2}{3}\left(\frac{e}{2}\right)^{3/2} \langle \gamma^2 \rangle k^6 \alpha^5 e^{-2k^2\alpha^2}, \tag{3}$$

where α is the correlation length.

B. Attenuation due to scattering

Assuming weak scattering within a continuum model, the attenuation coefficient due to scattering A_{sc} (dB cm^{-1}), in a medium composed of identical scatterers, is (Sehgal and Greenleaf, 1984)

$$A_{sc} = \frac{k^2\langle \gamma^2 \rangle}{2} \int_0^\infty b_\gamma(\Delta r)(\cos(k \cdot \Delta r) - \cos(3k \cdot \Delta r))d(\Delta r) \tag{4}$$

and, with the modified Gaussian correlation model, the attenuation coefficient due to scattering for identical scatterers will be

$$A_{sc} = 0.6\langle \gamma^2 \rangle \alpha k^2 \left(1 + \frac{1}{2}\alpha^2 k^2\right) e^{-\frac{1}{2}\alpha^2 k^2} \tag{5}$$

in Np cm^{-1}, where $k = 2\pi/\lambda$ is the wave number and a the scatterer size (or the correlation length).

Bovine cancellous bone possesses a distribution of trabecular thicknesses and the trabeculae may exhibit the same thickness over smaller regions. Thus, it can be regarded to be composed of a large number of small regions defined by their correlation lengths. If A_{sc}^i is the attenuation due to scattering coefficient from volume V_i, which represents the volume of tissue that has correlation length between a_i and $a_i + \Delta a_i$, then the total attenuation coefficient due to scattering will be

$$A_{sc}^T = \frac{1}{V}\sum_i A_{sc}^i V_i. \tag{6}$$

In the limit $\Delta V_i \to dV$, and since V_i is proportional to $(4/3)\pi(\alpha_i/2)^3$, the summation is replaced by an integral

$$A_{sc}^T = 3\int_a \frac{A_{sc} d\alpha}{a}. \tag{7}$$

Substituting A_{sc} in Eq. (7) by Eq. (5), and solving numerically the resulting integral between the limits α_n to α_m, where α_n and α_m are the minimum and maximum values of correlation lengths, the attenuation coefficient due to scattering, A_{sc}^T, for a continuous distribution of scatterers is obtained.

III. EXPERIMENTAL PROCEDURE

A. Bone specimens

Samples of cancellous bone were obtained from ten bovine tibiae of different animals. They were prepared using a band saw to make transverse cuts oriented parallel to the articular surface. These sections were subdivided into approximately cubic specimens with a side length of 22–25 mm. The surfaces of each specimen were finished with a slow speed diamond wafering saw. The resulting specimens had parallel faces with side lengths ranging between 20 and 22 mm. A total of 45 specimens (four or five specimens from each tibia) were prepared in this way.

Bone marrow was removed in order to determine apparent density, bone volume fraction and perform image analysis by alternately water jetting and immersing the samples in trichloro-ethylene solution in an ultrasound bath. The apparent density (the ratio of the dehydrated, defatted tissue mass to the total specimen volume) of the samples was determined from separate measurements of mass and volume. Mass was measured with a balance, after centrifuging the specimen for 10 min to remove excess water. Volume was determined by measuring the exterior dimensions of the cubes with calipers.

B. Ultrasonic measurements

Ultrasonic measurements were performed in a water bath at room temperature without removing marrow, because of the fact that some samples were too dense and extremely difficult to be defatted completely. In addition, it was shown that the influence of bone marrow on ultrasound properties is larger with decreasing density (Nicholson and Bouxsein, 2002b).

A number of 25 samples were interrogated in all three perpendicular directions [proximal-distal or axial (X), mediolateral (Y), and anteroposterior (Z)]. Focused (Panametrics, $V303$, $d=0.5$ in., center frequency 1 MHz, spherical focus $F=20$ mm) and unfocused (Panametrics, $V303$, $d=0.5$ in., center frequency 1 MHz) immersion transducers have been used in this study. They were connected to an ultrasonic pulser receiver (USD 10NF, Krautkraemer, Germany) that could be operated in a through-transmission or pulse-echo mode. Radio frequency (RF) signals were digitized at 35 MHz. The −20 dB frequency bandwidths were 0.24–1.24 MHz and 0.38–1.18 MHz, for unfocused and focused transducers, respectively.

A standard through-transmission substitution method was used to measure ultrasonic speed of sound and attenuation. Using two opposite, coaxially aligned unfocused transducers, the pulse transit time and pulse amplitude spectrum were calculated in the presence and in the absence of the bone sample in the acoustic path [Fig. 1(b)]. Attenuation coefficient was estimated using a log spectral difference technique. Diffraction related errors were corrected according to the method of Xu and Kaufman (1993). Transmission losses at the interface between the sample and the water were neglected. Attenuation versus frequency was least-squares fit to a linear function over the range from 200 to 900 kHz. The function was characterized by the slope of the resulting line (Broadband Ultrasound Attenuation, BUA). Speed of sound

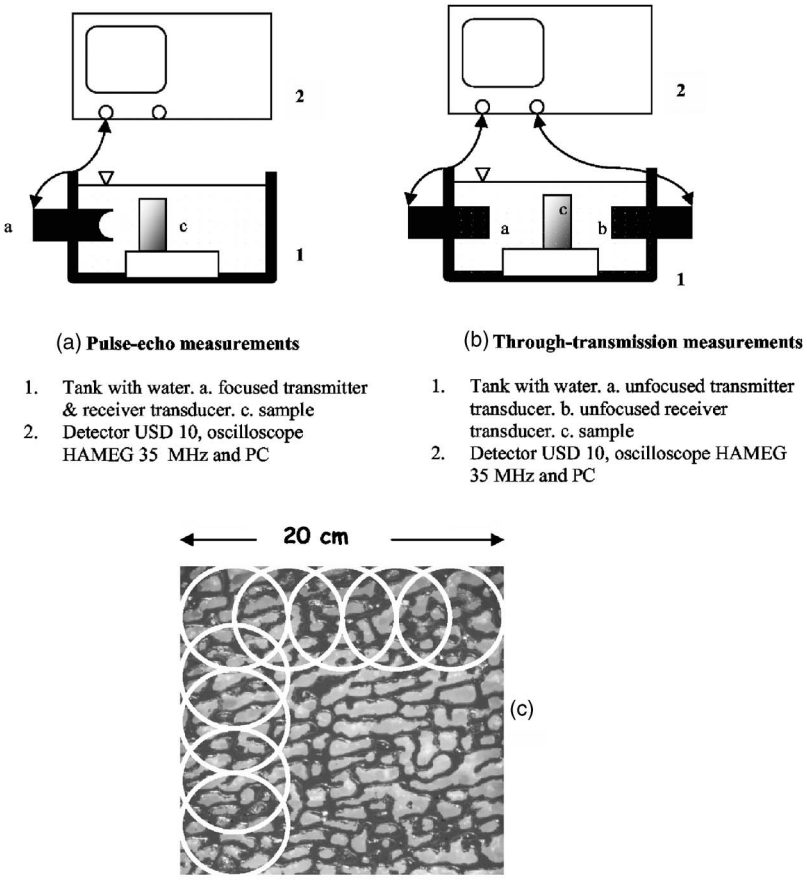


FIG. 1. Diagram of the experimental setup for the backscatter measurements (pulse echo) (a) and for the through transmission ultrasonic measurements (b). The number of independent measurements available for each bone sample is limited by the ratio of the bone specimen surface to the beam cross section (white circles). Due to the overlap between two adjacent measurements, 25 measurements with less than 50% overlap were obtained (c).

(SOS) was estimated using a time-of-flight method described by Njeh *et al.* (1997). To minimize the effect of frequency-dependent attenuation on measured SOS, the pulse transit time was calculated at the first zero crossing of the signal (Haiat *et al.*, 2006).

For backscatter measurements, the same system was used in pulse-echo mode using a single focused transducer [Fig. 1(a)]. The frequency-dependent backscatter coefficient was derived following the method of Roberjot *et al.* (1996) and Chaffaï *et al.* (2002), using a substitution technique. Backscattered RF signals from the bone specimens were acquired along two-dimensional (2D) scans in steps of 1 mm. The number of independent measurements available from each bone sample was limited by the ratio of the sample surface area to the beam cross section. In the middle of the usable bandwidth (0.8 MHz) an estimate of the beam width is $2.44\lambda F/d=6.75$ mm (Wear, 2001). Due to the overlap between two adjacent measurements, 25 measurements with less than 50% overlap were obtained from each sample surface [Fig. 1(c)]. The transducer output was gated appropriately (Hamming window) to permit analysis of a 7-mm-long region. The power spectra of the signals obtained from each windowed region were averaged and divided by a calibration spectrum (derived from the front surface of a flat steel plate, positioned at a distance to the transducer equal to that of the center of the gated region of the specimen under study) to remove various frequency dependent transfer functions associated with electronic units and the transducer. The backscatter measurements were compensated for three sources of error: attenuation (O'Donell and Miller, 1981), diffraction (Xu and Kaufman, 1993), and the effect of the gating function (Roberjot *et al.*, 1996).

The precision of backscatter measurements was assessed by measuring a group of five specimens three times with intermediate repositioning of the specimen. The precision for each specimen was expressed by the coefficient of variation, which is the standard deviation divided by the mean value of the three readings for this specimen.

Bone volume fraction was obtained experimentally using Archimedes' principle. Because of the fact that we had not at our disposal a microCT scanner to image the samples three dimensionally, we scanned a limited number of samples (five samples) and correlated trabecular thickness (Tb.Th, according to the standardized notation for bone histomorphometry) and bone volume fraction, obtained by microCT, with those obtained by our methods (Müller *et al.*, 1998).

Individual measurements of trabecular thickness were performed on each specimen. A volume of interest $20\times20\times7$ mm^3, centrally located within each specimen, was analyzed. Successive slices, perpendicular to the ultrasonic propagation, of 1 mm thickness, were produced by a microtome. Images of slices of each sample were taken by a stereoscope. The images were captured as color images of a resolution of 2560×2048 pixels. They were converted to black and white images through a threshold filtering.

The trabecular thickness was measured from these images, which are two-dimensional projections of the 3D tra-

becular structure, by an in-house code, based on a method developed by Kothari *et al.* (1999). Estimates of the thickness dimension were made in directions perpendicular to the trabecular orientation. They were measured at five equispaced points along the length of the trabecula. The used code was validated on simulated images. Thickness distributions for individual trabeculae were obtained for each bone specimen.

Trabecular thickness was compared to estimates of the correlation length from the measurements of backscatter coefficient. Experimental data and theoretical model were compared using the densely populated medium autocorrelation function, the spherical distribution, or a combination of the two functions. The best fit between experimental data and predictions, obtained by computing a least-square regression, yielded the value of an estimate of the correlation length a.

The calculation of the mean square fluctuation in medium properties $\langle \gamma^2 \rangle$ is based on the treatment of materials as immiscible mixtures. For a two component mixture, $\langle \gamma^2 \rangle$ is defined as (Sehgal, 1993))

$$\langle \gamma^2 \rangle = \langle \mu^2 \rangle + \langle \delta^2 \rangle \sin^4\left(\frac{\alpha}{2}\right)$$

$$\langle \mu^2 \rangle = \phi(1-\phi)\left(1 - \phi + \phi\left(\frac{c_b}{c_w}\right)^2\right)\frac{(c_b - c_w)^2}{c_b^2}$$

$$\langle \delta^2 \rangle = \phi(1-\phi)\left(1 - \phi + \phi\left(\frac{\rho_b}{\rho_w}\right)^2\right)\frac{(\rho_b - \rho_w)^2}{\rho_b^2},$$

where $\langle \mu^2 \rangle$ is the mean square of velocity fluctuation over scattering volume, $\langle \delta^2 \rangle$ is the mean square of density fluctuation, c_b, c_w, ρ_b, ρ_w are the velocities and densities in solid bone and water, respectively, and φ is the porosity of the specimen. The numerical value of $\langle \gamma^2 \rangle$ has been estimated by taking $\rho_b = 1800$ kg/m^3, $\rho_w = 1000$ kg/m^3, $c_b = 3300$ m/s, and $c_w = 1470$ m/s (Jenson *et al.*, 2003; Nicholson *et al.*, 2000) constant throughout a given specimen as well as for different individuals.

IV. RESULTS

Attenuation slope in the X direction (corresponding to the long axis of the femur) was 25.27 ± 3.52 dB/cm MHz compared with 26.36 ± 2.43 dB/cm MHz in the mediolateral direction (Y) and 27.02 ± 2.25 dB/cm MHz in the anterior-posterior direction (Z) (not significantly different) (Fig. 2). The difference between X, Y, and Z attenuation slopes for individual samples had a mean of 1.02, 1.76, and 1.86 dB/cm MHz correspondingly (X-Y, Y-Z, X-Z) and a standard error of 2.83, 2.28, and 3.22 dB/cm MHz. A t test revealed that these differences were not significantly different from zero.

The mean values of speed of sound were 2073 ± 253, 2010 ± 209, and 1882 ± 155 m/s along X, Y, and Z directions, respectively. Student paired t test showed statistically significant differences in speed of sound between anteroposterior (lower value) and the other two directions, which did not differ statistically.

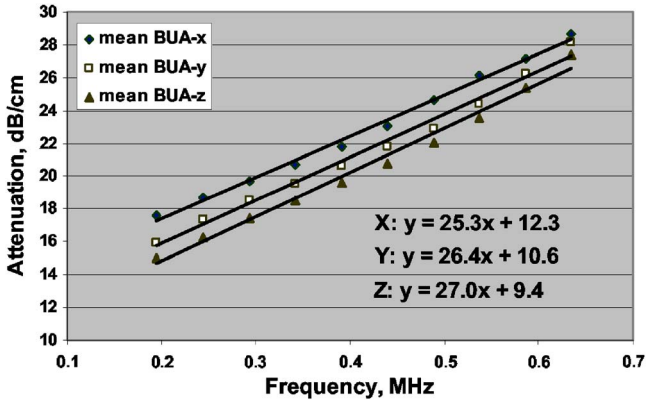


FIG. 2. (Color online) The group averaged attenuation, over 25 bone specimens, is plotted as a function of frequency. Directions are: proximal-distal (X), anteroposterior (Y), and mediolateral (Z).

Mean backscatter coefficients, as functions of frequency, were plotted for the 25 cubic specimens and the three orientations (Fig. 3). The backscatter coefficient at 800 kHz in the X direction was 0.210 ± 0.092 Sr^{-1} cm^{-1} compared with 0.235 ± 0.041 Sr^{-1} cm^{-1} in the Y direction and 0.102 ± 0.024 Sr^{-1} cm^{-1} in the Z direction (no significant differences). Thus, bovine cancellous bone can be regarded as isotropic material in respect to ultrasonic properties or its homogeneities are distributed isotropically in space.

The precision of estimate of the backscatter coefficient, assessed in the frequency range 0.4–1.2 MHz, resulted in a mean coefficient of variation of 2.8%.

The mean *Tb.Th*, obtained from the 3D reconstructed microarchitecture and averaged over six specimens of bovine bone, is 193 ± 98 μm. Because of the large variation of trabecular size, the mean *Tb.Th* was not used in this study. Instead, thickness distributions for individual trabeculae for each bone specimen were obtained. The thickness distribution graphs revealed the presence of one, two or, rarely, three dominant trabecular sizes for each specimen (Fig. 4).

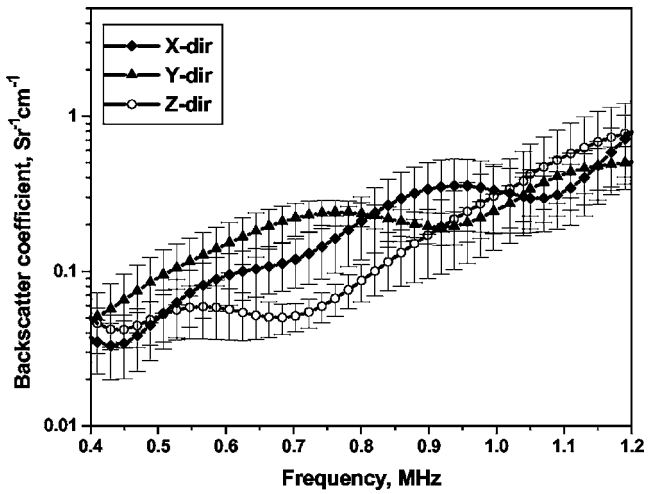


FIG. 3. The group averaged backscatter coefficient over 25 bone specimens is plotted as a function of frequency. Directions are: proximal-distal (X), anteroposterior (Y), and mediolateral (Z). Error bars denote standard errors of the mean.

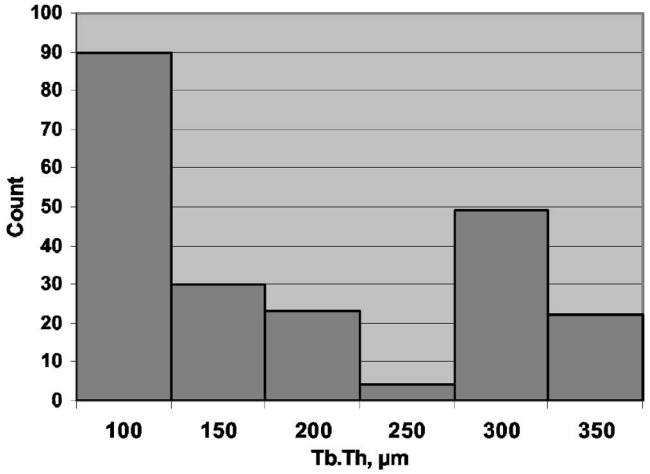


FIG. 4. Thickness distribution of individual trabeculae for a representative bovine bone specimen.

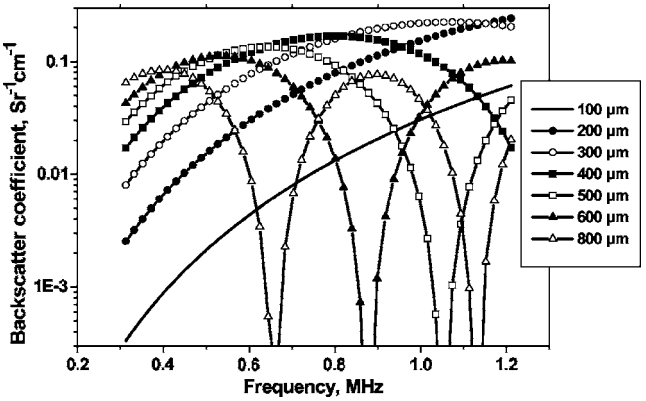


FIG. 6. Predicted backscatter coefficient for bovine bone with porosity of 70% for different scatterer sizes. The autocorrelation function for bovine bone tissue description was the spherical distribution.

Figures 5 and 6 represent the theoretical backscatter coefficients for the densely populated medium (DPM) and the spherical distribution (ES) autocorrelation functions, respectively, as functions of frequency for correlation lengths from 100 to 800 μm. The sharply defined resonances obtained for the spherical distribution model occur because we have assumed that only one value of diameter exists; in practical cases a range of sphere sizes would be present and the strong spectral scalloping expected for large scatterers is not observed.

When the correlation length is up to 150 μm, the backscatter coefficient is approximately proportional to 3.5 power of frequency (between the values of the exponent of the Faran cylindrical (f^3) and spherical (f^4) models for the frequency dependence). At larger correlation lengths, the values of backscatter coefficient vary more rapidly with increasing frequency and, in the range of frequencies examined, undulated features of the spectra are produced. For each correlation length, the maximum value of the backscatter coefficient appears at a certain frequency, which decreases with increasing correlation length. The frequencies that correspond to a certain scatterer size are different for the two autocorrelation functions, but they are linearly correlated ($R^2=0.996$, $p<0.0001$) (Fig. 7).

In general, the spectral features in Figs. 5 and 6 resembled those determined experimentally. The theoretically predicted backscatter coefficient magnitude from the densely populated model was of the order of those obtained experimentally from bovine cancellous bone. On the contrary, the predicted backscatter coefficient from the spherical distribution model was very low in comparison to the experimentally determined. Experimentally measured backscatter coefficient showed substantial between-sample variation.

Figure 8 depicts the experimental backscatter coefficient of one representative specimen and the data fitting with the theoretical model. The experimental data and the theoretical results fitted to within 5 dB, that is, the maximum accepted offset of the theoretical curve to the experimental was ±5 dB (Lizzi *et al.*, 1986). For most specimens, neither of the two models, based on the examined autocorrelation functions, was able to approximate accurately the shape of the experimental backscatter coefficient by itself. Thus, a combination of the two models was used to describe the experimental

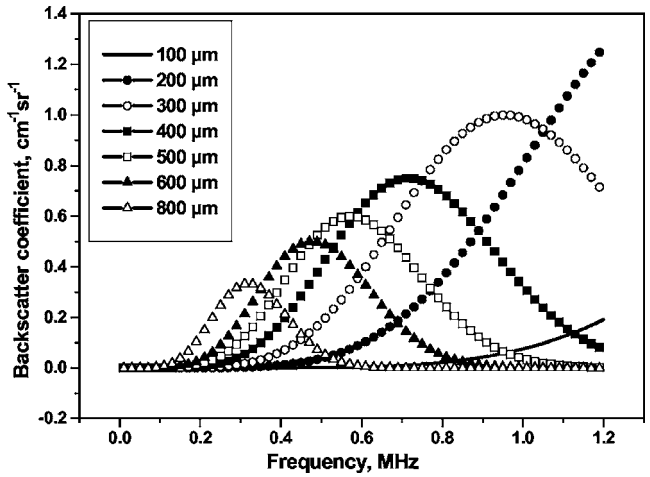


FIG. 5. Predicted backscatter coefficient for bovine bone with porosity of 70% for different scatterer sizes. The autocorrelation function for bovine bone tissue description was the densely populated medium model.

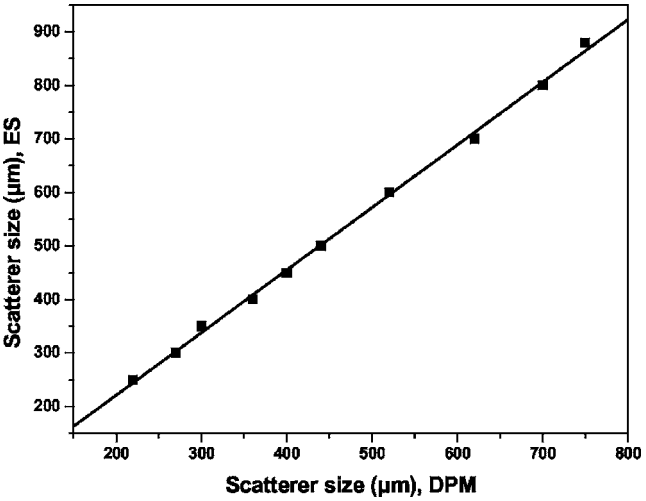


FIG. 7. The frequencies which correspond to a certain scatterer size are different for the two autocorrelation functions, but they are linearly correlated (DPM, densely populated medium model; ES, spherical distribution).

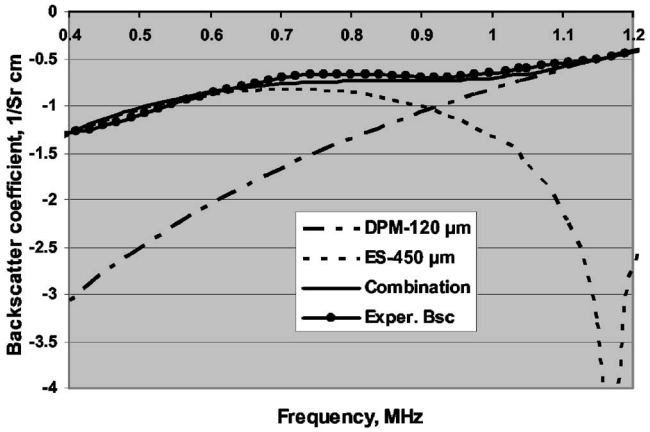


FIG. 8. Approximation of the experimental data by a combination of autocorrelation models. *Y*-axis scale is logarithmic. The densely populated medium model (DPM) predicts correlation length of 120 μm and the spherical distribution (ES) 450 μm.

data. The data were approximated within 5 dB by a two-component model: a densely populated model and a spherical distribution one. From the latter component, scatterers with a mean characteristic dimension of 450 μm were estimated. From the densely populated model component, scatterers with a mean characteristic dimension of 120 μm were estimated. The stereomicroscopically measured trabecular thickness values of the representative specimen were 130 ± 15 and 380 ± 28 μm. The predicted value of 450 mm by the spherical distribution model was very high in comparison to the measured one. However, a linear relationship was found between the correlation lengths predicted by the two models. The value of 450 mm predicted by the spherical distribution corresponded to 400 mm, predicted by the densely populated model. The estimated by the densely populated model correlation length closely corresponded to the larger scatterer sizes, measured by the stereomicroscope for the representative sample. Thus, a good prediction of the thickness of the two dominant groups of trabeculae was obtained. Comparison of backscatter ultrasonic measurements to theoretical predictions indicated that there was more than one dominant trabecular size that scattered sound for most specimens.

Figure 9 represents the measured trabecular thickness

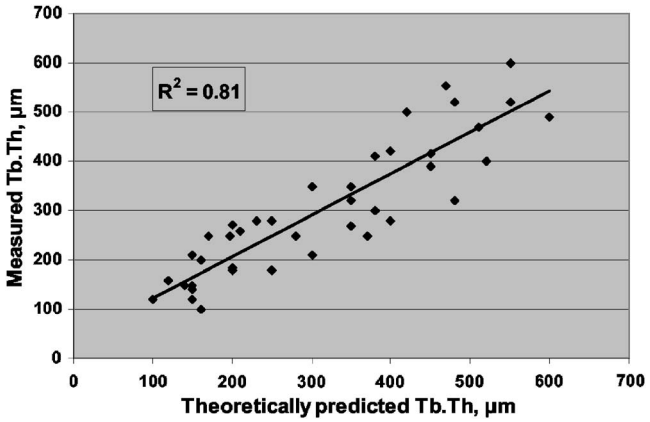


FIG. 9. Predicted correlation length versus measured trabecular thickness.

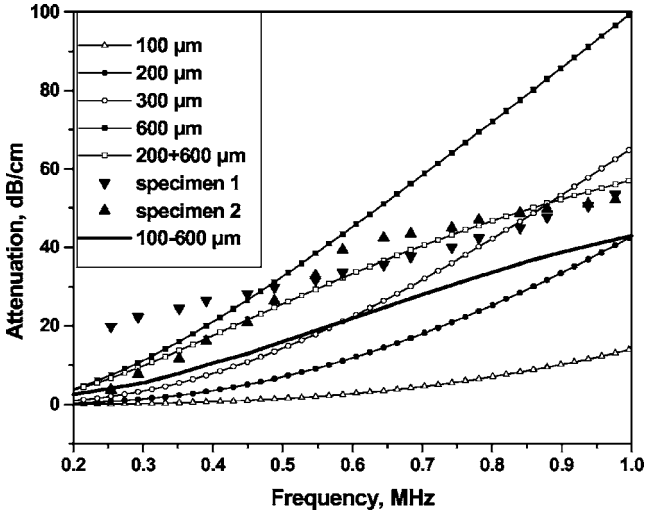


FIG. 10. Predicted attenuation due to scattering, with densely populated medium model, for a range of identical scatterers sizes from 100 to 600 μm, for two dominant groups of scatterers sizes, 200 and 600 μm, and for a continuous distribution of scatterers sizes 100–600 μm. ▲, ▼: experimental data, which correspond to the attenuation of two representative specimens, one displaying linear behavior with frequency in the examined frequency range and the other two distinct slopes of attenuation in the same frequency range.

(corresponding to one or more dominant scatterer sizes of trabeculae for each specimen) as a function of the estimated by the densely populated model correlation length. Linear regression, performed between trabecular thickness values and the estimated correlation lengths, showed significant linear correlation ($R^2=0.81$, $p<0.0001$).

Figure 10 represents the predicted from the densely populated model attenuation due to scattering by a continuous distribution of scatterers, ranging from 100 to 600 μm, versus frequency as well as the predicted attenuation due to scattering by identical scatterers of the same range. Attenuation due to scattering by a distribution of scatterers is linear in the range of 0.3–0.9 MHz. On the contrary, attenuation by identical scatterers is not linear or it is linear in a very limited frequency range, depending on the scatterer size. If two sizes of scatterers were present in the scattering volume, the attenuation versus frequency displays two different slopes. Experimentally, a number of specimens presented two distinct slopes of attenuation function versus frequency (Fig. 10). Linear dependence of ultrasonic attenuation from bovine cancellous bone specimens with high density was observed in this frequency range. BUA due to scattering by a distribution of scatterers was predicted to be 45 dB cm^{-1} MHz^{-1}. This value is of the order of the measured experimentally total attenuation for bovine cancellous bone. Predicted BUA due to scattering by identical scatterers increased with scatterers' size.

V. DISCUSSION

Isotropic weak scattering theory of random inhomogeneities was used successfully to provide a framework for the description of bovine cancellous bone tissue ultrasonic behavior. The analytical model described here has proven useful in improving our understanding of how tissue features are

related to measured ultrasonic characteristics. The random continuum approach has been used previously (Chaffaï *et al.*, 2000; Padilla *et al.*, 2003), obtaining good agreement with experimental data.

Two simple scattering autocorrelation functions have been utilized to describe the microscopic structure of the random inhomogeneous medium: A modified Gaussian model has been considered to account for a large volume fraction of scatterers in the medium, and the spherical distribution model because a number of larger plate-like scatterers with ellipsoid or roughly spherical shape may contribute to a portion of the scattering. The frequency dependence of the backscatter coefficient from bovine trabecular bone could not be described accurately either by scattering from a collection of randomly distributed identical scatterers or by a continuous model with a single dominant correlation length. This could be due to the tremendous variation in microarchitecture encountered in bovine trabecular bone. A combination of models was required to approximate within 5 dB the experimentally determined backscatter from bovine bone specimens with various microarchitecture, especially given the fact that the shape of trabeculae ranges from a platelike to a rodlike shape.

The backscatter model proposed in this study provided quantitative measurements of microstructural features of bone. Overall, a good qualitative agreement was found between predicted values and experimental results for both the magnitude and the frequency dependence of the backscatter coefficient. Absolute magnitude of backscatter coefficient was of the order of that predicted by the theoretical model.

The present model assumed the scattering medium to be isotropic, because in dense bovine bone trabeculae are randomly oriented. Unlike human trabecular bone, from calcaneous or from vertebrae, which displays anisotropy of ultrasonic backscatter and attenuation (Nicholson *et al.*, 1994; Wear *et al.*, 2000), bovine bone was found to be isotropic in respect to these ultrasonic properties. Contradictory results on bovine bone BUA anisotropy have been reported (Hoffmeister *et al.*, 2000; Wu *et al.*, 1998). It seems that bovine bone displays more or less anisotropic behavior in respect to ultrasonic properties depending on anatomic site and density. The measurements of the current study displayed differences in BUA and backscatter coefficient between axial and the other directions, however not statistically significant. On the contrary, the mean values of speed of sound displayed similar anisotropy like other researchers' results (Hans *et al.*, 1999; Njeh *et al.*, 1996). Axial and mediolateral directions were equivalent and differed significantly from the anteroposterior direction which displayed the lowest value.

The estimates of correlation lengths from the combination of the two models were compared to trabecular thickness. Mean values of trabecular thickness display very large variation at both intraspecimen and interspecimen levels and cannot be used as estimates of the correlation length a. A significant contribution of this work is that the interpretation of the correlation length was the dominant trabecular thickness, obtained from distribution graphs for each specimen. Significant linear correlation was found between predicted correlation length and measured trabecular thickness ($R^2=0.81$). Lower correlation coefficients ($R^2=0.44-0.53$) have been found for human cancellous bone (Jenson *et al.*, 2003; Padilla *et al.*, 2006). The better prediction of trabecular thickness for bovine bone may be due to various reasons. One reason may be the difference in the range of scatterers sizes, present in the two kinds of trabecular bone. In bovine bone scatterer sizes up to 600 and 700 μm were found (plate-like scatterers or interconnections of trabeculae), whereas the range of trabecular thickness in human bone from the elderly is limited to 50–200 μm. The distribution of scatterers sizes of bovine bone in a broader range results in a higher correlation coefficient between predicted and measured trabecular thickness. Another reason for the moderate correlation between measured and ultrasonically estimated trabecular thickness at an individual level for human bone might be the consideration of a single dominant structure, the mean value of trabecular thickness. The scalloping, observed in some experimental curves of backscatter from human trabecular bone, has been attributed to statistical fluctuations, which result from random interference noise (speckle) between the wavelets scattered by the randomly distributed trabeculae. There is a possibility that the scalloping is due to the presence of larger scatterers that produce undulated spectra. These scatterers might be interconnections of two trabeculae in a direction perpendicular to the wave propagation.

The estimated correlation lengths from the frequency dependence of backscatter coefficient, calculated by the densely populated model, were highly correlated with the measured scatterer sizes. The precision of the estimate of frequency dependence has been found to be much better than that of the magnitude of backscatter coefficient in human trabecular bone (Wear, 2001). Wear (2001) speculated that spatial variations in scatterer concentration would be expected to contribute additional variance in the magnitude of backscatter but not in the frequency dependence. The frequency dependence of backscatter depends on scatterer size but not on concentration provided that multiple scattering and coherent scattering are not big effects.

Precision limitations in trabecular thickness prediction arise from the fact that all models, either random continua or with discrete scatterers, deal with a single dominant scatterer size. In this work, although two or even more scatterer sizes have been considered, the dominant scatterer thicknesses are still distributed over a certain range. Thus, there is a variance of backscatter coefficient σ_b because of the thickness distribution. If the ultrasonic property is a slowly varying monotonic function of a and the spread of the size distribution is small, an approximate expression of the variance $\sigma^2(\sigma_b)$ is (Ishimaru, 1997)

$$\sigma^2(\sigma_b) \approx (\partial\sigma_b/\partial a)^2_{a_o}\sigma_a^2,$$

where $a_o=\langle a\rangle$ is the average scatterer size and σ_a^2 is the variance of the size distribution. Figure 11 depicts the variance of backscatter coefficient over the examined frequency range, for variation of ± 25 μm around the mean value of trabecular thickness.

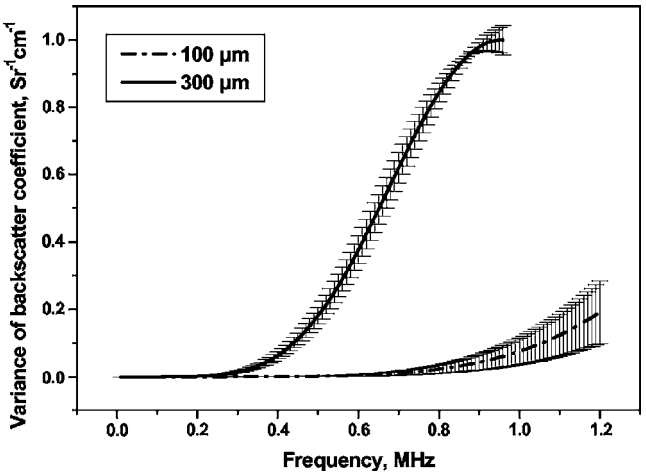


FIG. 11. Mean value ± variance of backscatter coefficient, calculated for the densely populated medium model, over the examined frequency range, for variation of ±25 μm around the mean value of trabecular thickness.

Attenuation due to scattering was predicted from the densely populated distribution model to be strong. The predicted attenuation due to scattering for a distribution of scatterers from 100 to 600 μm at 0.8 MHz was 70–80% of the measured total attenuation (due to both scattering and absorption mechanisms) for bovine cancellous bone (Fig. 10). Sehgal and Greenleaf (1984), using an exponential autocorrelation function and assuming a continuous distribution of correlation lengths between 0 and a_m, showed that attenuation due to scattering will yield an almost linear dependence with frequency for large values of ka_m. In this study, it was found that the frequency dependence of attenuation due to scattering is linear for a densely populated distribution of scatterers from 100 to 600 μm, which have been measured in bovine cancellous bone, and in the range of frequencies 0.3–0.9 MHz. This model predicted that commonly observed linear attenuation of bovine trabecular bone, in the diagnostic range of frequencies, could be explained by scattering.

The experimental results of this study suggest that scattering may be a major component of ultrasound attenuation through bovine trabecular bone. Indications that scattering may be a significant source of attenuation in human trabecular bone have been reported. Nicholson and Bouxsein (2002a) reported measurements of ultrasonic attenuation in human trabecular bone to be proportional to mean trabecular thickness to the 3.2 power. This value is very close to the approximate cubic dependence of backscattering on trabecular thickness in human bone, reported by Wear (1999) and Chaffaï *et al.* (2000). If scattering is the primary source of attenuation, these two results are very compatible. Kaufman *et al.* (2003), with computer simulations of ultrasonic propagations through 2D bone slices obtained from 3D data sets from human calcaneous, concluded that most of the frequency-dependent attenuation observed in trabecular bone is due to scattering of the ultrasound wave and not from absorption losses. Much more theoretical analysis and experimentation are necessary to determine the relative roles of scattering and absorption in attenuation of ultrasound through cancellous bone.

The main limitations of the model of random heterogeneous continuum for the description of low frequency ultrasound scattering from bone microarchitecture are the following:

- The assumption of statistical homogeneity over one sample. The accuracy was limited by the degree of inhomogeneity of the volume under investigation. Bovine bone is very strongly heterogeneous.
- First order multiple scattering approximation, which takes into account the attenuation of the incident wave due to the randomness of the medium, is considered in this study. A single scattering assumption is often used with good success for materials with weak scattering for early times or for experiments involving focused transducers (Turner and Weaver, 1995). Multiple scattering must be taken into account for media whose typical structures have sizes comparable to the wavelength (Tourin *et al.*, 2000; Page *et al.*, 1996). The issue of multiple scattering from trabecular bone has been addressed by Jenson *et al.* (2003) and Wear (1999).
- Weak scattering, that is, small fluctuations in velocity and density, have been assumed in this study. The weak fluctuation theory is valid only when the logarithm of the amplitude variance along the propagation path is small compared with unity and no more than about 0.2–0.5 (Ishimaru, 1997). The amplitude variance σ_χ^2 in a homogeneous random medium, calculated with the densely populated medium model (Appendix), is given by

$$\sigma_\chi^2 = \langle \gamma^2 \rangle \frac{2}{3\sqrt{2}} ak^2 L \sqrt{\pi},$$

where L is the propagation distance (in these experiments $L=7$ mm). The amplitude variance is a function of ak^2, when L (propagation distance) and $\langle \gamma^2 \rangle$ are constant. According to this relationship, for the maximum value of the refractive index ($\langle \gamma^2 \rangle$) fluctuation, which is 0.4 and occurs for porosities of about 65%, the amplitude fluctuation can be regarded as weak for frequencies up to 600 kHz when the scatterer size is 300 μm. For smaller scatterer sizes and lower value of $\langle \gamma^2 \rangle$, the weak scattering assumption is valid for higher frequencies. Thus, the experimental data of the current study fall in the limits of the weak fluctuation theory.
- An additional limitation of the autocorrelation models is that they do not account for shear waves in the scatterers. Possible longitudinal/transversal mode conversions at the solid-fluid interfaces are neglected.

Trabecular thickness is an important determinant of osteoporotic fracture risk. The current study showed that by ultrasound backscatter can predict the dominant trabecular sizes that scatter sound instead of a mean thickness and may add in the direction of noninvasively monitor microarchitectural changes associated with osteoporosis. However, the higher density of bovine cancellous bone may affect ultrasonic scattering in a fundamentally different manner than

human bone. Extrapolation of these results to diagnostic purposes involving less dense human bone should be done carefully.

ACKNOWLEDGMENTS

This project was financially supported by the project "K. Karatheodori" of the Research Committee of the University of Patras. The authors are grateful to Professor R. Muller, Institute for Biomedical Engineering ETH and University, Zurich, Switzerland, for μCT measurements, V. Cotsopoulos, University of Patras, for SEM images, and E. Chotra, MD, ASKLIPIOS, for BMD (QCT) measurements.

APPENDIX

When a wave is incident upon a random medium which may not be uniform, the amplitude and phase of the wave experience fluctuations due to the fluctuation of the refractive index of the medium along the path of propagation L. In regions that $L \gg \alpha^2/\lambda$, (λ is the wavelength), the variance of amplitude fluctuation is given by (Ishimaru, 1997, p. 356)

$$\sigma_\chi^2 = 2\pi^2 K^2 L \int_0^\infty k J_0(K \cdot \Delta r)\Phi_n(K) dK, \tag{A1}$$

where L is the propagation distance, $K=2k$, Δr is the correlation distance, J_0 is the first kind spherical Bessel function of zero order, and

$$\Phi_n(k) = \frac{1}{2\pi^2 k}\int_0^\infty b_\gamma(\Delta r)\sin(2k \cdot \Delta r) \cdot \Delta r \cdot d(\Delta r). \tag{A2}$$

In this work the densely populated autocorrelation function has been used as the correlation function of the refractive index fluctuation

$$b_\gamma(\Delta r) = \left(1 - \frac{\Delta r^2}{3\alpha^2}\right) e^{-\Delta r^2/2a^2}. \tag{A3}$$

Substituting Eqs. (A2) and (A3) in (A1), the amplitude variance is

$$\sigma_\chi^2 = \langle \gamma^2 \rangle \frac{2}{3\sqrt{2}} a k^2 L \sqrt{\pi}.$$

Chaffaï, S., Roberjot, V., Peyrin, F., Berger, G., and Laugier, P. (**2000**). "Frequency dependence of ultrasonic backscattering in cancellous bone: Autocorrelation model and experimental results," J. Acoust. Soc. Am. **108**, 2403–2411.

Chaffaï, S., Peyrin, F., Nuzzo, S., Porcher, R., Berger, G., and Laugier, P. (**2002**). "Ultrasonic characterization of human cancellous bone using transmission and backscatter measurements: Relationships to density and microstructure," Bone (N.Y.) **30**, 229–237.

Chernov, L. V. (**1960**). *Wave Propagation in Random Medium* (Dover, New York).

Glüer, C. C. (**1997**). "Quantitative ultrasound techniques for the assessment of osteoporosis: Expert agreement on current status," J. Bone Miner. Res. **12**, 1280–1288.

Haiat, G., Padilla, F., Cleveland, R. O., and Laugier, P. (**2006**). "Effects of frequency-dependent attenuation and velocity dispersion on in vitro ultrasound velocity measurements in intact human femur specimens," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 39–51.

Hakulinen, M. A., Day, J. S., Toyras, J., Weinans, H., and Jurvelin, J. S. (**2006**). "Ultrasonic characterization of human trabecular bone microstructure," Phys. Med. Biol. **51**, 1633–1648.

Hans, D., Wu, C., Njeh, C. F., Zhao, S., Augat, P., Newitt, D., Link, T., Lu, Y., Majumdar, S., and Genant, H. K. (**1999**). "Ultrasound velocity of trabecular cubes reflects mainly bone density and elasticity," Calcif. Tissue Int. **64**, 18–23.

Hoffmeister, B. K., Whitten, S. A., and Rho, J. Y. (**2000**). "Low-megahertz ultrasonic properties of bovine cancellous bone," Bone (N.Y.) **26**, 635–642.

Insana, M. (**1995**). "Modeling acoustic backscatter from kidney microstructure using an anisotropic correlation function," J. Acoust. Soc. Am. **97**, 649–655.

Insana, M. F., Wagner, R. F., Brown, D. G., and Hall, T. J. (**1990**). "Describing small structure in random media using pulse-echo ultrasound," J. Acoust. Soc. Am. **87**(1), 179–192.

Insana, M., and Brown, D. (**1993**). *Acoustic Scattering Theory Applied to Soft Biological Tissues* (CCR Press, London).

Ishimaru, A. (**1997**). *Wave Propagation and Scattering in Random Media* (Academic, Reissued: IEEE Press and Oxford University Press, UK).

Jenson, F., Padilla, F., and Laugier, P. (**2003**). "Prediction of frequency-dependent ultrasonic backscatter in cancellous bone using statistical weak scattering model," Ultrasound Med. Biol. **29**, 455–464.

Kaufman, J. J., Luo, G., and Siffert, R. S. (**2003**). "On the relative contributions of absorption and scattering to ultrasound attenuation in trabecular bone: A simulation study," *IEEE Ultrasonics Symposium*, 1519–1523.

Kothari, M., Keaveny, T. M., Lin, J. C., Newitt, D. C., and Majumdar, S. (**1999**). "Measurement of intraspecimen variations in vertebral cancellous bone architecture," Bone (N.Y.) **25**, 245–250.

Langton, C. M., Palmer, S. B., and Porter, R. W. (**1984**). "The measurement of broadband ultrasonic attenuation in cancellous bone," Eng. Med. **13**, 89–91.

Lizzi, F. L., Ostromogilsky, M., Feleppa, E. J., Rorke, M. C., and Yaremko, M. M. (**1986**). "Relationship of ultrasonic spectral parameters to features of tissue microstructure," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **33**, 319–329.

Luppé, F., Conoir, J.-M., and Franklin, H. (**2002**). "Scattering by a fluid cylinder in a porous medium: Application to trabecular bone." J. Acoust. Soc. Am. **111**, 2573–2582.

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (Princeton University Press, Princeton).

Müller, R., Van Campenhout, H., Van Damme, B., Van der Perre, G., Dequeker, J., Hildebrand, T., and Rüegsegger, P. (**1998**). "Morphometric analysis of human bone biopsies: A quantitative structural comparison of histological sections and micro-computed tomography," Bone (N.Y.) **23**, 59–66.

Njeh, C. F., Hodgskinson, R., Currey, J. D., and Langton, C. M. (**1996**). "Orthogonal relationships between ultrasonic velocity and material properties of bovine cancellous bone," Med. Eng. Phys. **18**, 373–381.

Njeh, C. F., Boivin, C. M., and Langton, C. M. (**1997**). "The role of ultrasound in the assessment of osteoporosis: A review," Osteoporosis Int. **7**, 7–22.

Nicholson, P. H. F., Haddaway, M. J., and Davie, M. W. J. (**1994**). "The dependence of ultrasonic properties on orientation in human vertebral bone," Phys. Med. Biol. **39**, 1013–1024.

Nicholson, P. H. F., Strelitzki, R., Cleveland, R. O., and Bouxsein, M. L. (**2000**). "Scattering of ultrasound in cancellous bone: Predictions from a theoretical model," J. Biomech. **33**, 503–506.

Nicholson, P., and Bouxsein, M. (**2002**). "On the relationship of ultrasonic properties to density and architecture in trabecular bone," J. Acoust. Soc. Am. **111**, 2413.

Nicholson, P. H. F., and Bouxsein, M. L. (**2002**). "Bone marrow influences quantitative ultrasound measurement in human cancellous bone," Ultrasound Med. Biol. **28**, 369–375.

O'Donell, M., and Miller, J. G. (**1981**). "Quantitative broadband ultrasonic backscatter: An approach to nondestructive evaluation in acoustically inhomogeneous materials," J. Appl. Phys. **52**, 1056–1065.

Oelze, M. L., Zachary, J. F., and O'Brien, W. D., Jr. (**2002**). "Characterization of tissue microstructure using ultrasonic backscatter: Theory and technique for optimization using a Gaussian form factor," J. Acoust. Soc. Am. **112**, 1202–1211.

Oelze, M. L., and O'Brien, W. D., Jr. (**2002**). "Frequency-dependent attenuation-compensation functions for ultrasonic signals backscattered from random media," J. Acoust. Soc. Am. **111**, 2308–2319.

Page, J. H., Sheng, P., Schriemer, H. P., Jones, I., Jing, X., and Weitz, D. A. (**1996**). "Group velocity in strongly scattering media," Science **271**, 634–637.

Padilla, F., Peyrin, F., and Laugier, P. (**2003**). "Prediction of backscatter coefficient in trabecular bones using a numerical model of three-dimensional microstructure," J. Acoust. Soc. Am. **113**, 1122–1129.
Padilla, F., Jenson, F., and Laugier, P. (**2006**). "Influence of the precision of spectral backscatter measurements in the estimation of scatterers size in cancellous bone," Ultrasonics **44**, Suppl. 1, e57–60.
Roberjot, V., Bridal, S. L., Laugier, P., and Berger, G. (**1996**). "Absolute backscatter coefficient over a wide range of frequencies in a tissue-mimicking phantom containing two populations of scatterers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 970–978.
Sehgal, C. M. (**1993**). "Quantitative relationship between tissue composition and scattering of ultrasound," J. Acoust. Soc. Am. **94**, 1944–1952.
Sehgal, C. M., and Greenleaf, J. F. (**1984**). "Scattering of ultrasound by tissues," Ultrason. Imaging **6**, 60–80.
Simmons, C. A., and Hipp, J. A. (**1997**). "Method based differences in the automated analysis of the three-dimensional morphology of trabecular bone," J. Bone Miner. Res. **12**, 942–947.
Strelitzki, R., Nicholson, P. H. F., and Paech, V. (**1998**). "A model for ultrasonic scattering in cancellous bone based on velocity fluctuations in a binary mixture," Physiol. Meas **19**, 189–196.
Tavakoli, M. B., and Evans, J. A. (**1991**). "Dependence of the velocity and attenuation of ultrasound in bone on the mineral content," Phys. Med. Biol. **36**, 1529–1537.
Thomsen, J. S., Ebbesen, E. N., and Mosekilde, L I. (**2002**). "Static histomorphometry of human iliac crest and vertebral trabecular bone," Bone (N.Y.) **30**, 267–274.
Tourin, A., Derode, A., Peyre, A., and Fink, M. (**2000**). "Transport parameters for an ultrasonic pulsed wave propagating in a multiple scattering medium," J. Acoust. Soc. Am. **108**, 503–512.
Trebacz, H., and Natali, A. (**1999**). "Ultrasound velocity and attenuation in cancellous bone samples from human vertebra and calcaneous," Osteoporosis Int. **9**, 99–105.
Turner, J. A., and Weaver, R. L. (**1995**). "Time dependence of multiply scattered diffuse ultrasound in polycrystalline media," J. Acoust. Soc. Am. **97**, 2639–2644.
Wear, K. A., and Garra, B. S. (**1998**). "Assessment of bone density using ultrasonic backscatter," Ultrasound Med. Biol. **24**, 689–695.
Wear, K. A. (**1999**). "Frequency dependence of ultrasonic backscatter from human trabecular bone: Theory and experiment," J. Acoust. Soc. Am. **106**, 3659–3664.
Wear, K. A., Stuber, A. P., and Reynolds, J. C. (**2000**). "Relationships of ultrasonic backscatter with ultrasonic attenuation, sound speed and bone mineral density in human calcaneous," Ultrasound Med. Biol. **26**, 1311–1316.
Wear, K. A. (**2001**). "Fundamental precision limitations for measurements of frequency dependence of backscatter: Applications in tissue-mimicking phantoms and trabecular bone," J. Acoust. Soc. Am. **110**, 3275–3282.
Wear, K. A. (**2003**). "The dependence of ultrasonic backscatter on trabecular thickness in human calcaneous: Theoretical and experimental results," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 979–986.
Wear, K. A. (**2004**). "Measurement of dependence of backscatter coefficient from cylinders on frequency and diameter using focused transducers—with applications in trabecular bone," J. Acoust. Soc. Am. **115**, 66–72.
Wu, C., Gluer, C., Lu, V., Fuerst, T., Hans, D., and Genant, H. K. (**1998**). "Ultrasound characterization of bone demineralization," Calcif. Tissue Int. **62**, 133–139.
Xu, W., and Kaufman, J. J. (**1993**). "Diffraction correction methods for insertion ultrasound attenuation estimation," IEEE Trans. Biomed. Eng. **40**, 563–569.

Direct observations of ultrasound microbubble contrast agent interaction with the microvessel wall

Charles F. Caskey
Biomedical Engineering, 451 East Health Sciences Drive, University of California, Davis, Davis, California 95616

Susanne M. Stieger
Surgical and Radiological Sciences, School of Veterinary Medicine, University of California, Davis, Davis, California 95616

Shengping Qin, Paul A. Dayton, and Katherine W. Ferrara[a)]
Biomedical Engineering, 451 East Health Sciences Drive, University of California, Davis, Davis, California 95616

(Received 15 May 2007; revised 16 May 2007)

Many thousands of contrast ultrasound studies have been conducted in clinics around the world. In addition, the microbubbles employed in these examinations are being widely investigated to deliver drugs and genes. Here, for the first time, the oscillation of these microbubbles in small vessels is directly observed and shown to be substantially different than that predicted by previous models and imaged within large fluid volumes. Using pulsed ultrasound with a center frequency of 1 MHz and peak rarefactional pressure of 0.8 or 2.0 MPa, microbubble expansion was significantly reduced when microbubbles were constrained within small vessels in the rat cecum ($p < 0.05$). A model for microbubble oscillation within compliant vessels is presented that accurately predicts oscillation and vessel displacement within small vessels. As a result of the decreased oscillation in small vessels, a large resting microbubble diameter resulting from agent fusion or a high mechanical index was required to bring the agent shell into contact with the endothelium. Also, contact with the endothelium was observed during asymmetrical collapse, not during expansion. These results will be used to improve the design of drug delivery techniques using microbubbles. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2747204]

PACS number(s): 43.80.Cs, 43.80.Gx [FD] Pages: 1191–1200

I. INTRODUCTION

Safe and effective local gene therapy and delivery of therapeutic agents across the blood-brain barrier are two crucial goals of current medical research. Recent studies have demonstrated promising early results using ultrasound and microbubble agents to enhance vascular or cellular permeability to drugs or genes [Greenleaf *et al.*, 1998; Kost *et al.*, 2000; Unger *et al.*, 2001; Taniyama *et al.*, 2002; Prausnitz *et al.*, 2004; Mitragotri, 2005]. Increased vascular permeability caused by microbubble contrast agents was first shown in rat muscle but was associated with petechial hemorrhage and vascular inflammation [Miller and Gies, 1998; Skyba *et al.*, 1998; Chappell *et al.*, 2005]. Subsequent studies have shown that increased permeability without excessive leakage of red blood cells or endothelial cell death is possible within the brain and other tissues [Kinoshita *et al.*, 2006; Stieger *et al.*, 2006]. Recently, Bekeredjian *et al.* have found that, with attachment of plasmid DNA to a microbubble shell, ultrasound-targeted microbubble destruction enhances transfection efficiency, with an organ specificity far greater than can be achieved using viral vectors [Bekeredjian *et al.*, 2003]. However, the mechanism and efficiency of transfer of a drug or gene from a microbubble shell to the endothelium have not been fully characterized. A precise understanding is crucial for optimization of these techniques, especially as some preclinical studies employing microbubbles as carrier vehicles have used microbubble doses ten to one hundred times the accepted safe dose for humans [Bekeredjian *et al.*, 2003; Chen *et al.*, 2006]. The goal of this work is to better understand the behavior of microbubbles within microvessels, in order to design effective and repeatable drug delivery techniques without unwanted biological effects.

The oscillation of bubbles in response to an acoustic field has been studied for more than a century [Bjerknes, 1906]. Similarly, the mechanical effects of ultrasound on human tissue have received great attention for more than 30 years [Dyson *et al.*, 1968]. Recent acute interest in these two subjects has resulted from the introduction of gas-filled microbubble contrast agents, with extensive efforts devoted to understanding their biological effects and resulting biomedical applications. When these agents are injected intravenously, they respond strongly to ultrasound imaging pulses, allowing clinicians to create high-contrast images of the blood pool with spatial resolution on the order of hundreds of microns. Augmented forms of the Rayleigh-Plesset equation have been developed to describe the oscillation of shelled microbubbles in an infinite fluid, and the predicted oscillation is highly correlated with experimental observa-

[a)]Electronic mail: kwferrar@ucdavis.edu

tions within cellulose or acrylic tubes (mimetic vessels) with diameters greater than 200 μm [Morgan *et al.*, 2000; Allen *et al.*, 2002].

Microbubble contrast agents oscillate and collapse nonlinearly in response to ultrasound pulses of moderate intensity. When this bubble collapse is driven by inertial forces, substantial kinetic energy is transferred to the contracting bubble, and a wideband pressure wave is generated [Crum, 1979; Dear *et al.*, 1988]. This condition is generally termed "inertial cavitation." Inertial cavitation near a rigid boundary can result in shock-wave formation and pitting of the surface by a liquid jet. In the absence of exogenous microbubble contrast agents, the presence of the wideband sound wave correlates with biological effects. To avoid vascular damage, ultrasound transmission from commercial imaging instruments is maintained below a threshold for spontaneous bubble formation: the mechanical index (MI, pressure in megaPascals divided by the square root of frequency in megahertz) must be less than 1.9 [Deng *et al.*, 1996; Barnett *et al.*, 2000]. However, in the presence of microbubble contrast agents, a wideband scattered echo occurs even with a very low mechanical index [Kruse and Ferrara, 2005] and the mechanical index does not directly correlate with microbubble destruction [Chomas *et al.*, 2001a; Forsberg *et al.*, 2006], or biological effects [Stieger *et al.*, 2006]. A relative expansion (maximum radius divided by the resting radius) of 3.46 has also been associated with cavitation [Apfel, 1986].

Our group was the first to conduct high-speed optical studies to characterize the oscillation and destruction of ultrasound contrast agents *in vitro*, and we and others have characterized oscillation in mimetic vessels with diameters ranging from microns to centimeters [Dayton *et al.*, 1997; Klibanov *et al.*, 1998; Morgan *et al.*, 1998; Dayton *et al.*, 1999b; Morgan *et al.*, 2000]. These experimental observations have been compared with increasingly refined theoretical models [Allen and Roy, 2000; de Jong *et al.*, 2000; Marmottant and Hilgenfeldt, 2003; Sassaroli and Hynynen, 2005; Caskey *et al.*, 2006; Marmottant *et al.*, 2006; Qin and Ferrara, 2006]. Factors such as vessel diameter, proximity of the microbubble to the vessel wall, liquid viscosity, and vessel compliance have all been shown theoretically and with *in vitro* experiments to have a dramatic effect on microbubble oscillation and the mechanisms associated with contrast agent bioeffects. At least two possible mechanisms have been suggested for ultrasound-mediated drug delivery using microbubble contrast agents. One possibility is that microbubbles expand and collapse violently against the endothelial wall [Postema *et al.*, 2004]. A second possible mechanism is that during the bubbles' oscillation the agents expand against the wall, facilitating exchange between bubble shell and cell wall components [Zhong *et al.*, 2001]. However, neither of these mechanisms has been directly observed in a vessel network. We have recently developed the ability to conduct high-speed optical studies *within a vascular bed*. This paper presents the first high-speed optical observations of microbubbles in *ex vivo* vessels.

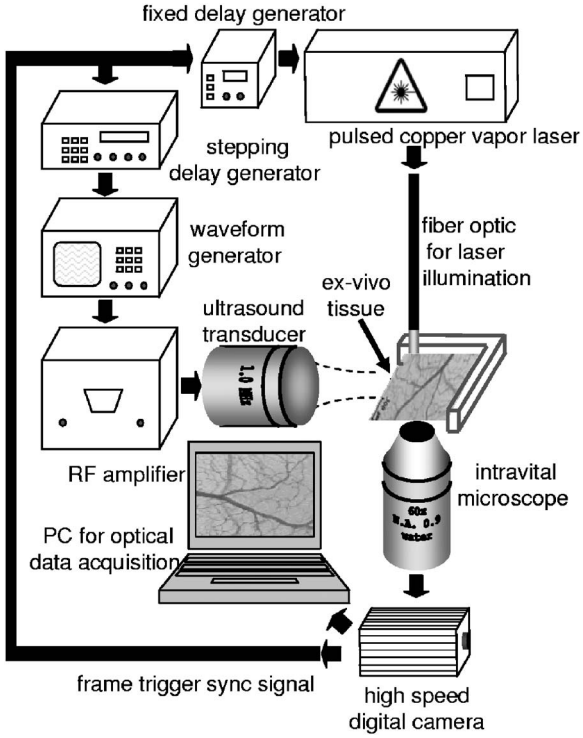


FIG. 1. Diagram of the experimental system.

II. METHODS

A. Experimental methods

1. Experimental system

We have designed an imaging system to capture high-resolution images through tissue (Fig. 1). In our system, illumination is achieved by using a 20-W pulsed copper vapor laser (LS-10-10, Oxford Lasers; Shirley, MA) with a 30-nanosecond pulse width and wavelength of 510 and 578 nm. The optical system consists of a custom-designed intravital microscope (IV500L Mikron Instruments; San Diego, CA). Digital video capture was provided by a high-speed camera system (APX-RS, Photron; Tucson, AZ) which acquired 10-bit, 1-megapixel frames at 1000 frames per second. Both the laser and a 250-W halogen light source (Techniquip; Danville, CA) were coupled to the microscope through fiber optics for high-speed ($>$1000 frames per second) and low-speed ($<$120 frames per second) imaging, respectively.

The acoustical system consists of a 1-MHz, 1.91-cm-diameter element ultrasound transducer (IL0106HP; Valpey Fisher, Hopkinton, MA) spherically focused at a depth of 5.08 cm, with a −6-dB lateral beam width of 3.6 mm. The transducer was perpendicularly and confocally focused with a 60×, 0.9-N.A. water immersion microscope objective (Olympus; Melville, NY). The ultrasound transducer was energized by an arbitrary waveform generator (AWG2021; Tektronix, Irvine, CA) and a 55-dB radiofrequency amplifier (3200L, ENI; Rochester, NY). Acoustic pressure measurements and confocal alignment were performed with a 400-μm-diameter calibrated needle hydrophone (HNZ-0400; Onda Corp, Sunnyvale, CA), where the calibration was provided by Onda Corp. Ultrasound exposure was provided with

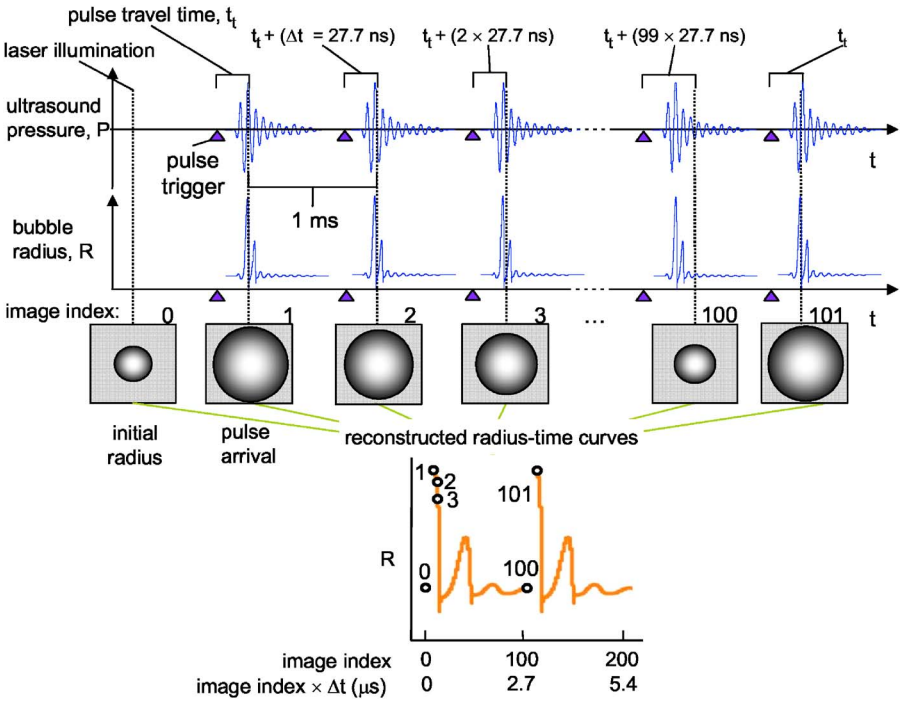


FIG. 2. (Color online) Schematic of the system operation. A train of ultrasound pulses with pressure, P, was transmitted shifting each transmission by a multiple (denoted image index) of 27.7 ns relative to the image acquisition. Over the pulse train, images were acquired with an interimage delay of 1 ms and an illumination of 30 ns duration. P is shown as measured by hydrophone. Typical oscillations, R, are shown for a center frequency of 1 MHz, PRP of 0.8 MPa, and initial radius of 2 μm.

2-cycle sinusoidal waveforms at either of two transmission pressures: a peak rarefactional pressure of 0.8 MPa (MI =0.8), or 2.0 MPa (MI=2).

Data were captured as a sequence of images that reflects both a microbubble's oscillation in response to a single pulse, and the evolution of the microbubble over a series of pulses. A first image was acquired prior to insonation and the second image acquired at a time estimated to correspond with peak expansion. The sequence was summarized in an effective radius-time curve, displaying one point for each pulse (Fig. 2). The composite of images was acquired and indexed according to the frame number, mapping out the expansion and compression of microbubbles, with 100 pulses and optical images required to map out one radius-time replica. After 100 images, the system had fully sampled the radial oscillation and the timing generator was reset to again sample the oscillation, repeating such sampling for 50 consecutive replicas.

2. Timing generation

A 1-kHz signal from the camera served as the system clock, which was then delayed by a delay generator (9650A, EG&G; Fremont, CA) by approximately 35 microseconds and used to trigger the laser. This time delay, which accounted for the propagation delay for the ultrasound pulse to travel from the transducer to the tissue, was measured prior to each experiment via the hydrophone which was placed in the optical sample volume. The system clock was also delayed by an adjustable stepping delay generator (AVX-DD, Avtech; Ogdensburg, NY) before triggering the waveform generator. The stepping delay generator was set to capture the oscillation maximum on the first ultrasonic pulse and subsequently decrement the delay time of the ultrasonic pulse relative to the optical acquisition by 27.7 ns per pulse for 100 pulses. A laptop PC (Latitude 400, Dell; Round Rock, TX) interfaced to the digital camera records image and timing data.

3. Animal model

All studies involving animals were approved and accepted by the Animal Care and Use Committee at the University of California, Davis. The cecum was extracted as described below and maintained in a heated saline bath. A solution of microbubbles and PBS was slowly perfused into the cecum. The perfusion was halted briefly during ultrasound application to reduce motion artifacts in the image. A total of 20 procedures was conducted. Representative data are presented here from the subset of studies in which a small number of microbubbles (as described below) were isolated within blood vessels with a diameter of 30 μm or less.

4. Cecum preparation and observation

Male, Zucker lean rats (Harlan Sprague Dawley; Indianapolis, IN) were humanely euthanized following approved euthanasia protocols and the abdominal cavity was opened at the linea alba. In the next step, the ilecolic vein was canalized and PE10 tubing (Intramedic, BD; Franklin Lakes, NJ) was inserted into the vessel lumen. After ligation of all surrounding mesenteric vessels, the cecum was carefully removed from the abdominal cavity. Next, for *ex vivo* observation, the cecum with the PE tube in place was opened, cleaned, and placed on a tissue holder. The opened cecum wall was attached to the tissue holder and the tissue was placed in a heated saline bath (37 °C) for optical observation. The translucent epithelial layer of the cecum allows for visualization of blood vessels with a microscope. Both the objective and the ultrasound transducer were immersed in a phosphate-buffered saline (PBS) bath, which served both as

an acoustic coupling medium and as a perfusion bath for the tissue. A custom polycarbonate holder attached to a 2-axis positioning stage secured the tissue at the focus of the objective for optical observation, and allows movement of the tissue in the horizontal plane. The microscope's focusing mechanism was used for focusing in the z plane, while maintaining confocal optical and acoustical alignment. The contrast agent was introduced and flow was then stopped momentarily with the microbubbles within the field of view. After the images were acquired, flow was restarted and a new microbubble solution pumped into the optical and acoustical field of view.

5. Image processing for diameter and radius-time curves

Images were analyzed using IMAGEJ (NIH, http://rsb.info.nih.gov/ij/) by measuring the largest diameter of the microbubble normal to the direction of the vessel. A line through the center of the bubble and its surroundings can be compiled for each image in the set, producing an effective radius-time image. The pixel size in the digital image is approximately $(200\ \text{nm})^2$ and only bubbles with clearly defined boundaries were measured.

6. Displacement measurement methods

Displacement measurements were made using a block-matching algorithm implemented in MATLAB. The algorithm makes measurements on two adjacent images from a video sequence, I_n and I_{n-1}, referred to here as the current and previous images, respectively. A small area in the previous image, called the kernel, is centered at a given location and used for comparison to an area of the same size in the current image. The normalized correlation coefficient between the kernel and a sub-block of the same size in the current image is calculated for sub-blocks centered at each location in the search area, creating a normalized cross-correlation function.

Maximization of the cross-correlation function yields 2D displacement estimation. Subpixel estimation is achieved by fitting a polynomial to the row or column of the search area containing the maximum correlation and solving for the maximum. A kernel size of 21×21 pixels ($4.1\times4.1\ \mu\text{m}^2$) was experimentally determined to give the best displacement map for the noise level in the microscope images. A search area of 15×15 pixels $(2.9\ \mu\text{m})^2$ was used based on the observation that interframe movement of any pixel is less than 15 pixels (2.9 μm) in any direction. Accumulation of interframe data was computed by a path integration of the 2D displacement across displacement maps for a sequence of images.

7. Microbubble size distribution

The experimental agent is a lipid-shelled, perfluoropropane-filled ultrasound contrast agent, containing microbubbles of $\sim1.8\pm1.5\ \mu$m in diameter, where the creation of this agent was described in Borden *et al.*, 2005. The initial formulation of the agent contains $\sim1\times10^{10}$ microbubbles/ml. During studies of microbubble expansion, the agent is diluted such that fewer than 10 agents are visible within each $(200\text{-}\mu\text{m})^2$ optical field. During studies of microbubble fusion, a slightly higher concentration was used, producing $\sim$20 microbubbles per field.

8. Statistical methods

The effect of resting diameter on expansion and oscillatory lifetime was analyzed. A two-sided Student's t-test with alpha of 0.05 and assuming unequal variance was performed to compare the expansion and the oscillatory lifetime of bubbles. A p value of 0.05 indicated a significant difference.

B. Model for microbubble oscillation in small compliant vessels

The model, which was proposed for bubble oscillation in small compliant vessels, was used to evaluate microbubble expansion and vessel wall deflection, using the recorded transmitted pressure waveform as the excitation. A liquid column (20 μm in length) was added at each end of the vessel and the liquid viscosity within these two columns set to 10 poise in order to achieve boundary conditions at the ends of the vessel of a liquid velocity near zero and pressure approximately equal to the acoustic pressure. In order to match the measured oscillatory period ($<1\ \mu$s), the nondimensional rigidity index, k, is set to zero. We assume a vessel wall composed of one layer of endothelial cells with a thickness of 1.3 μm, surrounding interstitial thickness of $\sim140\ \mu$m, surface tension of 0.35 N/m, and effective dynamic thickness of the surrounding tissue, h_t, of 20 μm. For a 25-μm vessel, the ratio of vessel radius to wall thickness, γ, is then calculated to be 10.5 and the nondimensional surrounding connective tissue thickness coefficient, ξ, is calculated to be 10. For a 14-μm vessel, γ is 6 and ξ is 20. All other parameters were held constant with values as described in Qin and Ferrara, 2006.

III. RESULTS

This study employs low-frequency (1-MHz) contrast ultrasound with parameters previously associated with successful drug and gene delivery [Chen *et al.*, 2006], but also with inflammation (0.8 MPa) or significant endothelial disruption (2.0 MPa) [Stieger *et al.*, 2006]. We used a high-speed camera system with 30-ns laser illumination to visualize the oscillation and destruction of microbubbles within a vascular bed (Fig. 1, Fig. 2). We therefore were able to directly observe the microbubble-endothelial interactions.

All measurements described below were made in vessels with a diameter of 30 μm or less and a mean vessel diameter of $18\pm7\ \mu$m. Prior to the application of ultrasound, microbubbles were observed to be distributed evenly throughout the vessel volume. As described in Sec. II, ten or fewer bubbles were visible within $(200\text{-}\mu\text{m})^2$ optical fields.

A. Oscillation amplitude

With a peak rarefactional pressure (PRP) of either 0.8 or 2.0 MPa, microbubble expansion within the microvasculature was significantly constrained, as compared with expansion within a 200-μm tube ($p=0.01$) [Fig. 3(a)]. With 0.8-

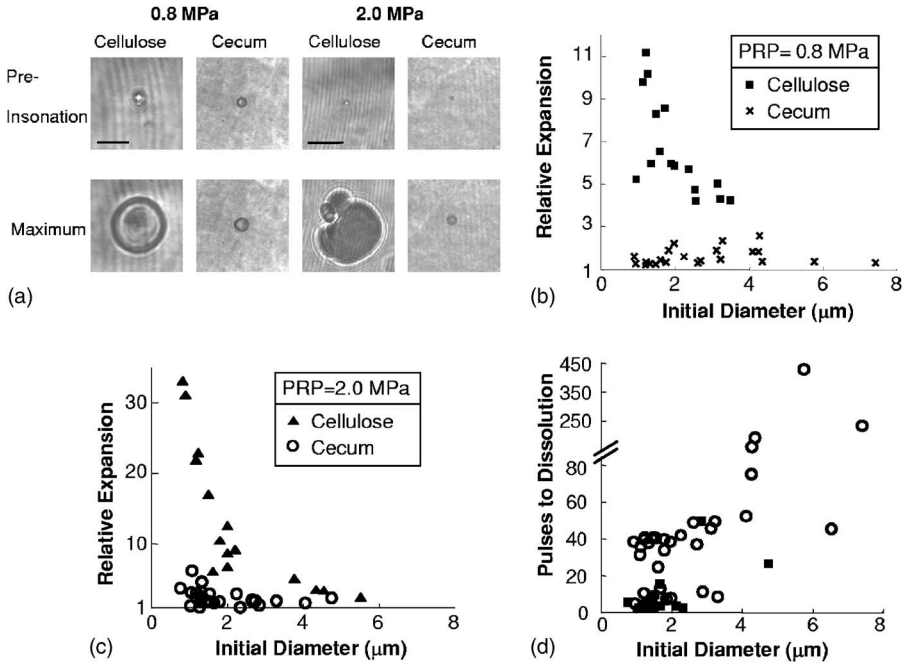


FIG. 3. Summary plots for microbubble expansion and lifetime. (a) Sample images of microbubble expansion within a 200-μm diameter cellulose tube and 14-μm blood vessel obtained before and during the peak rarefaction. Scale bar indicates 10 μm. (b), (c) Relative expansion vs initial microbubble diameter for experimental measurements (points) within 200-μm cellulose tubing and cecum microvessels with (b) PRP of 0.8 MPa; (c) PRP of 2.0 MPa. (d) Number of pulses prior to disappearance (oscillatory lifetime) vs initial microbubble diameter for insonation with center frequency of 1 MHz, PRP of 0.8, and 2.0 MPa. Under identical conditions within 200-μm cellulose tubing, oscillatory lifetime is less than two pulses.

MPa PRP ultrasound exposure, the observed relative expansion was in all cases less than 2.7 ($n=22$) [Fig. 3(b)]. For microbubbles with a diameter between 1 and 3 μm (mean diameter 1.5 ± 0.4 μm), the average relative expansion was 1.5 ± 0.3. This expansion in the cecum microvasculature was substantially smaller than that observed in 200-μm tubing, where relative expansion ranged from 4 to 11.5 under the same experimental conditions [Fig. 3(b)]. With a higher transmission pressure of 2.0 MPa [Fig. 3(c)] and a similar size group (diameter 1.7 ± 0.8 μm), the relative expansion was significantly larger (2.5 ± 1.2, $p=0.01$), yet still far below that observed in larger mimetic vessels (relative expansion up to 33) ($p<0.001$). No significant relationship between microvessel diameter and microbubble expansion was observed in the range of vessels studied, although it is difficult to achieve an identical distribution of and diameter of microbubbles within different vessels.

Microbubble expansion was compared with our model of microbubble oscillation in small vessels. With a 1-MHz, 0.8-MPa PRP driving pulse, the relative expansion of 4- and 7-μm diameter bubbles was estimated within an infinite fluid to be 7.5 and 4.8, respectively, while within a 14-μm vessel, the predicted relative expansion was 2.9 and 1.8. The experimentally observed relative expansion for these parameters was 2.6 and 1.3.

For microbubbles with an initial diameter smaller than 8 μm, the minimum diameter observed during the ultrasonic compression ranged from a value less than the system resolution ($\sim$300 nm) to $\sim$1 μm. The ratio of maximum to minimum diameter exceeded 8 for all microbubbles with an initial diameter greater than 3 μm and less than 8 μm. Thus, although expansion was constrained, microbubble compression/collapse was observed.

B. Oscillatory lifetime and destruction

With 0.8-MPa PRP insonation, individual microbubbles were observed to be intact for 5 or more pulses in all cases [Fig. 3(d)], increasing to a duration of hundreds of pulses for large microbubbles. Larger microbubbles persisted longer than smaller microbubbles, with bubbles of a resting diameter less than 4 μm visible for 31 ± 14 pulses on average and bubbles with a diameter greater than 4 μm visible for 176 ± 139 pulses on average ($p=0.03$). The diameter during the second rarefaction was $94\pm28\%$ of the peak diameter during the first rarefaction ($n=17$).

With 2.0-MPa insonation, all microbubbles were intact for at least 2 entire pulses. The average oscillatory lifetime decreased relative to 0.8-MPa ultrasound exposure, from 30 pulses ($n=22$) to 8 pulses ($n=26$) ($p=0.006$). However, oscillatory lifetime in the cecum remained significantly greater than microbubbles within 200-μm mimetic tubing, where all microbubbles were destroyed by a single ultrasound pulse ($p<0.001$). The extended oscillation of small microbubbles within these small blood vessels provides additional evidence of constrained oscillation. We hypothesize that the extended oscillation lifetime may result from the decreased oscillation amplitude and therefore reduced instability.

In a previous study with similar ultrasound exposure parameters in a 200-μm tube, large microbubbles were reduced in diameter by 1% to 30% with each ultrasonic pulse (depending on resting diameter), eventually reaching a diameter of 4.4 μm, where fragmentation occurred [Chomas *et al.*, 2001b]. This acoustically driven diffusion mechanism was also observed in the present study, with the rate of decrease in diameter consistently less than 0.5% per pulse for microbubbles smaller than 4.4 μm in initial diameter, and less than 0.3% per pulse for larger microbubbles (>4.4 μm).

C. Sample images and radius-time curves

A typical small microbubble (diameter less than 2 μm), insonified at 0.8 MPa, doubled in diameter during the first rarefaction [Fig. 4(a)], while expansion in the second rarefaction (frame 35) was 38% of that in the first. No evidence of

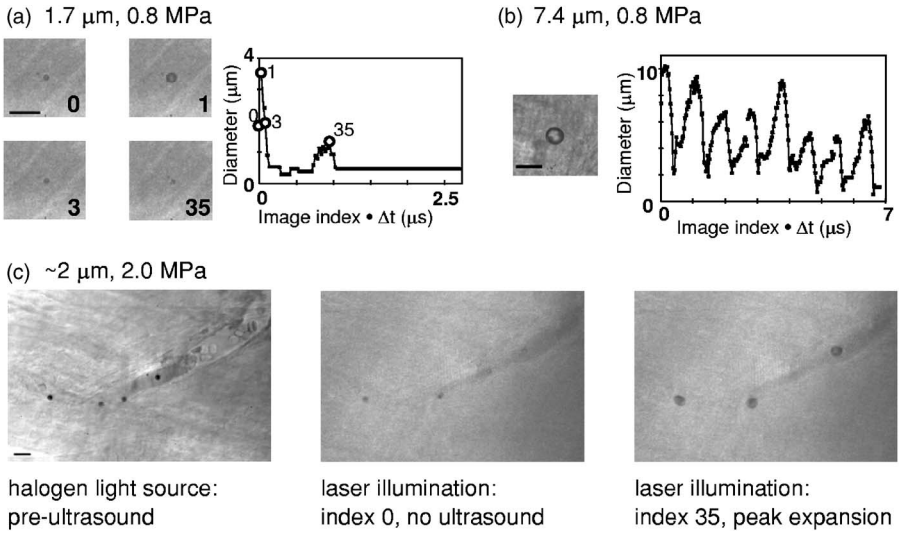


FIG. 4. Sample radius-time curves and images indexed according to description in Fig. 2. Scale bar indicates 10 μm. (a) Images and corresponding radius-time curve obtained from a microbubble with an initial diameter of 1.7 μm, in a vessel with a 13-μm diameter for a PRP of 0.8 MPa. (b) Initial image obtained from microbubble with an initial diameter of 7.4 μm, in a vessel 13 μm in diameter, with ultrasound exposure at 0.8 MPa and corresponding radius-time curve. (c) Images obtained before and after ultrasound exposure at 2.0 MPa from microbubbles with initial diameters of ~2 μm, in a vessel 13 μm in diameter. Left and middle: initial images before ultrasound exposure, acquired with the halogen light source and laser illumination, respectively. Right: Image acquired during peak expansion.

asymmetry, shift in resting position, fluid jet, or interaction with the endothelium was observed. A typical microbubble with a larger resting diameter (7.4 μm), insonified with a PRP of 0.8 MPa, persisted substantially longer than small microbubbles [Fig. 4(b)]. In this case, oscillation was observed for nearly 700 pulses, and asymmetric oscillation was observed during much of this time. As shown by the radius-time curve, the maximum expansion decreased during the observation period, with the radius during the final expansion approximately 62% of that during the first. Figure 4(c) shows a typical microbubble insonified with a PRP of 2.0 MPa that was visible for more than 50 pulses, frequently in contact with the endothelium. At this higher transmission pressure, even small microbubbles were observed to oscillate asymmetrically and to interact with the endothelium.

As visualized on the sample radius-time curves [Figs. 4(a) and 4(b)], for all microbubbles with a diameter less than 8 μm, when an entire oscillatory cycle was observed, the oscillation period was less than 1 μs (920.2±111 ns, n=21). Thus, the oscillation period was nearly equal to the period of the transmitted pulse, indicating that the microvascular constraint has not greatly decreased the microbubble oscillation frequency.

D. Fusion

Closely spaced microbubbles experience a secondary, mutually attractive radiation force when exposed to ultrasound: *in vitro* and *in vivo* observations have demonstrated an approach velocity on the order of meters/second for a transmission pressure on the order of hundreds of kPa [Dayton *et al.*, 1999a]. In these experiments, when an aggregate of microbubbles was separated by a distance of tens of microns, secondary radiation force attracted the agents to one another and fusion occurred within a few pulses (Fig. 5). Although few red blood cells were present in these images, secondary radiation force in the presence of red blood cells was demonstrated in Dayton *et al.*, 1999a.

E. Vessel wall interaction

Asymmetrical collapse during compression and/or vessel wall displacement during expansion was observed for all microbubbles with a resting diameter greater than 4 μm, as shown in Fig. 6. For example, in Figs. 6(a) and 6(b), a 9-μm bubble, formed by the fusion of smaller microbubbles, crosses the endothelium during collapse, stably oscillating partially within and partially outside the endothelium for tens of pulses [Fig. 6(b)]. Eventually, the microbubble center reenters the vessel lumen and asymmetrical oscillation continues.

Collapse into the endothelium was observed without regard to the direction of ultrasound propagation: for example in Fig. 6(c), collapse was parallel to the image plane; in Fig. 6(d), it was perpendicular. Between frames 1 and 181 in Fig. 6(c), the microbubble moved a total of 9.3 μm along and toward the vessel boundary, in a direction consistent with a

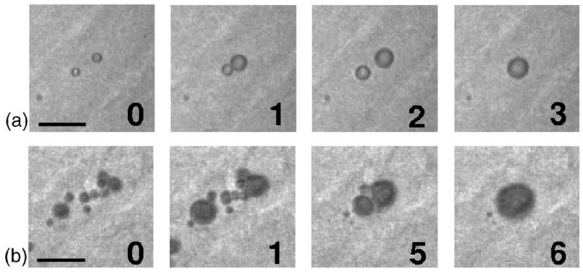


FIG. 5. Microbubble fusion. Two sequences of images (indexed according to description in Fig. 2) obtained with a transmission center frequency of 1 MHz and PRP of 0.8 MPa as multiple microbubbles approach and fuse. The vessel diameter in both sequences is approximately 12 μm. Scale bar indicates 10 μm.

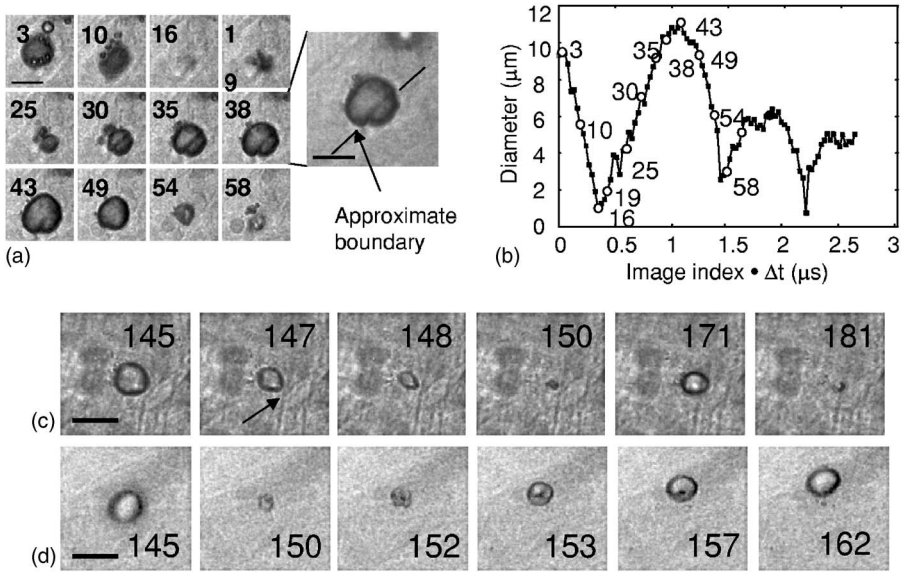


FIG. 6. Collapse of microbubbles into the vessel wall. Data obtained with a transmission center frequency of 1 MHz and PRP of 0.8 MPa. (a) Images documenting microbubble interaction with the endothelium indexed relative to microbubble fusion at index 1. Within frame 19, the microbubble crosses the vessel boundary; during frames 30–49 the microbubble expands, partially within and partially outside the lumen. Subsequently, the microbubble completely re-enters the lumen. Scale bar indicates 10 μm. (b) Radius-time curve corresponding to (a). (c)–(d). Two sequences of images (indexed as described in Fig. 2) in which a microbubble oscillates in contact with a vessel wall. Scale bar indicates 10 μm. In (c), microbubble cross section is elongated during collapse (frames 148 and 150). In frames 171 and 181, the bubble wall remains in contact with the vessel boundary during collapse. (d) When microbubble is viewed from above, a toroidal shape is observed in frames 152 and 153.

secondary attractive force between the bubble and boundary [Crum, 1975], finally coming into contact with the boundary. When viewed from the side, the microbubble cross section was typically elongated during collapse (frames 148 and 150), consistent with previous *in vitro* images of bubble collapse near a boundary [Lindau and Lauterborn, 2003]. Within the last few oscillatory cycles (see frames 171 and 181), the bubble wall remained in contact with the boundary during collapse and the far bubble wall oscillated rapidly. When viewed from above [Fig. 6(d)], a toroidal shape was observed (particularly in frames 152 and 153), again consistent with *in vitro* images of bubble collapse near a boundary [Lindau and Lauterborn, 2003].

Expansion of large microbubbles (not in contact with the vessel wall at rest) displaced the endothelium [Figs. 7(a)–7(c)] but did not bring the microbubble shell into contact with the endothelium. Figures 7(a)–7(c) show the largest wall deflection observed in our data set, visualized as a function of time in Fig. 7(c), where changes in the vessel and microbubble boundary can be directly observed. Across all observations, vessel wall deflection increased with increasing initial microbubble and decreasing vessel diameter. Using a correlation analysis detailed in Sec. II, the local microvessel wall displacement was estimated to be 2.3 μm during peak expansion [Fig. 7(d)], with an inward displacement of 0.8 μm during compression. Traveling into the tissue, the deflection of the vessel wall can be detected over a distance of $\sim$7 μm perpendicular to the vessel wall, where the spatially decreasing displacement is shown in Fig. 7(d). Traveling along the wall surface, one half of the peak displacement ($\sim$1.1 μm) was detected over a distance of $\sim$10 μm from the bubble center [Fig. 7(e)].

The predicted vessel wall displacement, based on Qin and Ferrara, 2006, was overlaid on the experimental observations, demonstrating a similar magnitude and spatial variation. Also based on Qin and Ferrara, 2006, the circumferential wall stress was calculated to be 56.4 kPa (where a rupture threshold for large arteries is 0.8 MPa [Di Martino *et al.*, 2006]). Both the model for microbubble oscillation and experimental results indicate that the vessel wall displace-

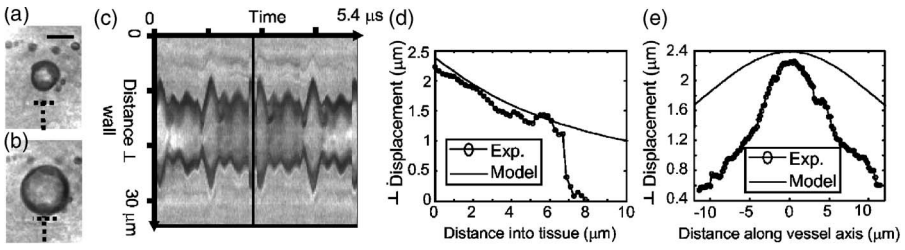


FIG. 7. Deflection of the vessel wall produced by microbubble expansion. (a), (b) Images of a microbubble within a $\sim$23-μm diameter vessel acquired before (a), and during (b), 1-MHz, 0.8-MPa ultrasound exposure. Scale bar indicates 10 μm. Dotted lines indicate locations for displacement profile plots in (d) and (e). (c) M-mode image of center line through bubble imaged in (a) and (b), showing motion of vessel and bubble wall over time. During peak expansion, the microbubble displaces the vessel wall. (d) Displacement of the vessel wall estimated and predicted as a function of the perpendicular distance from the wall. (e) Perpendicular displacement of the vessel wall estimated and predicted as a function of distance along the wall.

ment was ~0.5 μm for a 1-MHz, 0.8-MPa insonation of a microbubble with a diameter of ~4 μm within a 14-μm vessel. The corresponding circumferential stress was estimated to be 6.2 kPa. With increasing PRP, the displacement and circumferential stress were predicted to increase very rapidly. However, experimentally, with a 2.0-MPa PRP, we did not encounter bubbles of as large of a diameter as that shown in Fig. 7 (likely due to a greater propensity for microbubble destruction) and therefore did not observe as large of a vessel wall displacement.

IV. DISCUSSION

In this paper we present the first direct optical observations of ultrasound contrast agents within a capillary bed and demonstrate that the oscillation of microbubbles is constrained within small blood vessels as compared with larger (200-μm) mimetic vessels. Our results have several significant implications for the design of microbubble-based drug delivery techniques.

A. Microbubble oscillation

Many of our observations support the results and predictions of previous studies. Rigid mimetic vessels with a diameter on the order of a capillary have been observed to restrict the expansion of a microbubble [Caskey *et al.*, 2006]. Here, we show that microbubble expansion in an *ex vivo* capillary network was also restricted despite the larger compliance of the vessel wall. This altered oscillation within the smallest vessels could change the scattered echo. Second and third harmonic production has been associated with blood brain barrier opening for the transmission of 260-kHz ultrasound [McDannold *et al.*, 2006], and the detection of wideband echoes (with a "passive cavitation detector") is the established method for noninvasive estimation of the potential for bioeffects in preclinical studies. In the smallest vessels, alternative methods of detecting microbubble-endothelial interactions may be needed.

In most cases, the reduced expansion of bubbles in microvessels was associated with a greatly increased oscillation lifetime (tens to hundreds of pulses rather than just one) and a change in the destruction mechanism from fragmentation to acoustically driven diffusion. A relative expansion of 3.46 has been associated with cavitation (collapse) in previous studies [Apfel, 1986]. With a PRP of 0.8 MPa, relative expansion for all observed microbubbles was smaller than 2.7, which is below the cavitation threshold. If delivery vehicle fragmentation is desired for local delivery, recognition of this limitation in the smallest vessels may be important and require a higher ultrasonic pressure.

The model for microbubble expansion reported previously [Qin and Ferrara, 2006] has been validated for the oscillation of 4–7-μm bubbles within small vessels. Predictions of both the radial oscillation and scattered echo of microbubbles as a function of microbubble and vessel diameter are possible with these techniques. The model and tracking algorithms allow us to demonstrate that a wall displacement of 2.3 μm is possible for a large microbubble that nearly fills the vessel volume during expansion, and these tools will be used to evaluate the effect of circumferential stress in the future.

B. Mechanisms for vascular effects

Holes within and between endothelial cells and vascular inflammation have previously been shown by electron microscopy to be produced by the insonation of microbubbles using the pressure and frequency parameters employed here [Stieger *et al.*, 2006]. Two major mechanisms have been proposed for such microbubble-related bioeffects: asymmetric collapse and jet formation causing damage to the endothelial wall, and bubble expansion pushing against the endothelial lining. We have confirmed that both of these effects do occur in a capillary network. While only microbubble collapse was observed to bring the microbubble and vessel wall into contact (and therefore could transfer a drug), the expansion of the vessel wall may produce an inflammatory response or otherwise change the vessel wall.

C. Effect of microbubble diameter and concentration

We did not expect to observe such a strong dependence of microbubble size on interaction with the endothelium with 1-MHz, 0.8-MPa ultrasound. Microbubbles with a diameter less than 2 μm were not observed to interact with the endothelium. However, large microbubbles (>4 μm) were substantially longer-lived and more likely to expand the vessel wall and to undergo asymmetrical collapse into it. Since bubbles moved slowly across the vessel lumen over tens to hundreds of pulses before contact occurred, large long-lived bubbles may be required. Commercially available agents vary in their initial diameter and the effect of their size distribution should be considered when interpreting reports of bubble-endothelial interactions.

The results described here hold several implications for the design of microbubbles as carriers of drugs or genes. They suggest a possible nonlinear dependence of bioeffects on microbubble diameter and concentration. Recent studies of gene delivery from microbubble-based vehicles have employed high doses of contrast agents, or have used arterial injection with or without venous ligation [Christiansen *et al.*, 2003; Koike *et al.*, 2005; Chen *et al.*, 2006]. Therapeutic gene delivery from a lipid-microbubble shell following venous administration has been demonstrated with ~5.2 $\times 10^9$ microbubbles injected into a blood volume of ~12 ml, or 4×10^5 microbubbles/μl, as compared with a typical human bolus dose of 10^{10} microbubbles in 5 l or 2 $\times 10^3$ microbubbles/μl [Chen *et al.*, 2006]. Secondary radiation force has been shown to result in an attraction and fusion of microbubbles over a distance of ~40 μm. Within the preclinical studies described above, a 40-μm-length vessel with a diameter of 40 μm (containing a blood volume of 5×10^{-5} μl) would be expected to contain 20 microbubbles at steady state and a greater number as the initial bolus circulates; within a typical human study such a volume would not contain multiple bubbles. Thus, preclinical studies which have demonstrated functional gene delivery have involved concentrations that can produce large and long-lived mi-

crobubbles through fusion. Furthermore, microbubble fusion may not produce repeatable and predictable drug or gene delivery, as the local concentration of microbubbles may vary with flow conditions and individual vascular anatomy. High-intensity ultrasound (1 MHz, 2.0 MPa) could produce such effects at a lower concentration without requiring agent fusion. However, hemorrhage and cell death have been associated with the use of such a high pressure in contrast ultrasound.

Future studies of gene delivery could employ monodisperse microbubbles created with a larger resting diameter (≥ 4 μm) or those whose shells include targeting ligands, in order to facilitate microbubble/endothelial interactions and minimize undesired effects such as excessive inflammation or hemorrhage. Bubbles with a larger initial diameter may interact with the endothelium at a lower PRP (0.8 MPa) and lower initial concentration. Targeted agents, once bound, have been shown to oscillate asymmetrically with evidence of a jet when using a low PRP [Zhao *et al.*, 2006]. Otherwise, if smaller (1–2-μm diameter) microbubbles are used, successful delivery may require a greater local concentration (on the order of 10 microbubbles within a 40-μm length of a 40-μm diameter vessel) and therefore require either a large injected dose ($\sim 10^5$ microbubbles/μl) or intra-arterial injection.

D. Limitations of the study

The smallest vessels observed in the rat cecum had a diameter of approximately 10 μm, and the effect of microbubble oscillation may be greater in smaller vessels. However, previous studies [Stieger *et al.*, 2006] have indicated that the greatest volume of drug extravasation occurs in vessels with a diameter from 15 to 55 μm, and therefore the range of diameters studied here corresponds to those most likely to contribute to drug and gene delivery.

The temporal sampling method employed here acquires one sample per ultrasound pulse and relies upon repeated pulsing to insure that the microbubble oscillation is adequately sampled. Events that are not repeated with each oscillatory cycle could be missed. For periodic phenomena, a direct comparison of the oscillation of microbubbles within the cecum and larger mimetic vessels confirmed that our experimental system is capable of imaging large (or small) microbubble expansion and stable asymmetric features.

The spatial resolution of our optical system was on the order of 300 nm, and therefore measurement of the minimum diameter is inaccurate for bubbles with resting diameters below 3 μm. Finally, the analytical model used as a basis of comparison provides a good approximation for oscillation amplitude for large ($>$4-μm) bubbles within a small vessel. However, the current model must be modified in the future to predict the limited expansion of microbubbles with a diameter of 1–2 μm, as the model does not yet incorporate the shell inertial and viscous effects.

V. CONCLUSION

Here, we present the first direct observations of oscillating microbubbles within a capillary network. We find that microbubble expansion is decreased by two- to ten-fold and oscillatory lifetime is substantially increased compared with oscillation in large vessels. Microbubble fragmentation (destruction following a single ultrasound pulse) is not observed. At a high microbubble concentration, similar to that of published preclinical studies of gene therapy, microbubbles fuse, forming agents with a diameter greater than 4 μm. Microbubbles larger than 4 μm were observed to interact with the endothelium, where the mechanisms for microbubble-endothelial interactions demonstrated here include asymmetrical collapse into the endothelium and repeated expansion of the microbubble beyond the vessel limits. Our results suggest that the efficacy of ultrasound-mediated drug delivery using microbubble agents will depend strongly on agent size distribution and on *in vivo* concentration, as well as the acoustic parameters.

ACKNOWLEDGMENTS

The support of NIH CA 103828 and 112356 and the assistance of Susannah Bloch and Mark Borden in the preparation of this manuscript are gratefully acknowledged.

Allen, J. S., and Roy, R. A. (**2000**). "Dynamics of gas bubbles in viscoelastic fluids. I. Linear viscoelasticity," J. Acoust. Soc. Am. **107**, 3167–3178.

Allen, J. S., May, D. J., and Ferrara, K. W. (**2002**). "Dynamics of therapeutic ultrasound contrast agents," Ultrasound Med. Biol. **28**, 805–816.

Apfel, R. E. (**1986**). "Possibility of microcavitation from diagnostic ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **33**, 139–142.

Barnett, S. B., Ter Haar, G. R., Ziskin, M. C., Rott, H. D., Duck, F. A., and Maeda, K. (**2000**). "International recommendations and guidelines for the safe use of diagnostic ultrasound in medicine," Ultrasound Med. Biol. **26**, 355–366.

Bekeredjian, R., Chen, S. Y., Frenkel, P. A., Grayburn, P. A., and Shohet, R. V. (**2003**). "Ultrasound-targeted microbubble destruction can repeatedly direct highly specific plasmid expression to the heart," Circulation **108**, 1022–1026.

Bjerknes, V. F. K. (**1906**). *Fields of Force* (Columbia University Press, New York).

Borden, M. A., Kruse, D. E., Caskey, C. F., Zhao, S. K., Dayton, P. A., and Ferrara, K. W. (**2005**). "Influence of lipid shell physicochemical properties on ultrasound-induced microbubble destruction," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 1992–2002.

Caskey, C. F., Kruse, D. E., Dayton, P. A., Kitano, T. K., and Ferrara, K. W. (**2006**). "Microbubble oscillation in tubes with diameters of 12, 25, and 195 microns," Appl. Phys. Lett. **88**, 033902.

Chappell, J. C., Klibanov, A. L., and Price, R. J. (**2005**). "Ultrasound-microbubble-induced neovascularization in mouse skeletal muscle," Ultrasound Med. Biol. **31**, 1411–1422.

Chen, S. Y., Ding, J. H., Bekeredjian, R., Yang, B. Z., Shohet, R. V., Johnston, S. A., Hohmeier, H. E., Newgard, C. B., and Grayburn, P. A. (**2006**). "Efficient gene delivery to pancreatic islets with ultrasonic microbubble destruction technology," Proc. Natl. Acad. Sci. U.S.A. **103**, 8469–8474.

Chomas, J. E., Dayton, P., May, D., and Ferrara, K. (**2001a**). "Threshold of fragmentation for ultrasonic contrast agents," J. Biomed. Opt. **6**, 141–150.

Chomas, J. E., Dayton, P., Allen, J., Morgan, K., and Ferrara, K. W. (**2001b**). "Mechanisms of contrast agent destruction," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 232–248.

Christiansen, J. P., French, B. A., Klibanov, A. L., Kaul, S., and Lindner, J. R. (**2003**). "Targeted tissue transfection with ultrasound destruction of plasmid-bearing cationic microbubbles," Ultrasound Med. Biol. **29**, 1759–1767.

Crum, L. A. (**1975**). "Bjerknes forces on bubbles in a stationary sound field," J. Acoust. Soc. Am. **57**, 1363–1370.

Crum, L. A. (**1979**). "Surface oscillations and jet development in pulsating bubbles," J. Phys. (Paris) **41**, 285–288.

Dayton, P., Klibanov, A., Brandenburger, G., and Ferrara, K. (**1999a**). "Acoustic radiation force *in vivo*: A mechanism to assist targeting of mi-

crobubbles," Ultrasound Med. Biol. **25**, 1195–1201.
Dayton, P. A., Morgan, K. E., Klibanov, A. L., Brandenburger, G. H., and Ferrara, K. W. (**1999b**). "Optical and acoustical observations of the effects of ultrasound on contrast agents," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 220–232.
Dayton, P. A., Morgan, K. E., Klibanov, A. L. S., Brandenburger, G., Nightingale, K. R., and Ferrara, K. W. (**1997**). "A preliminary evaluation of the effects of primary and secondary radiation forces on acoustic contrast agents," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1264–1277.
de Jong, N., Frinking, P. J. A., Bouakaz, A., Goorden, M., Schourmans, T., Xu, J. P., and Mastik, F. (**2000**). "Optical imaging of contrast agent microbubbles in an ultrasound field with a 100-MHz camera," Ultrasound Med. Biol. **26**, 487–492.
Dear, J. P., Field, J. E., and Walton, A. J. (**1988**). "Gas-compression and jet formation in cavities collapsed by a shock wave," Nature (London) **332**, 505–508.
Deng, C. X., Xu, Q. H., Apfel, R. E., and Holland, C. K. (**1996**). "*In vitro* measurements of inertial cavitation thresholds in human blood," Ultrasound Med. Biol. **22**, 939–948.
Di Martino, E. S., Bohra, A., Vande Geest, J. P., Gupta, N., Makaroun, M. S., and Vorp, D. A. (**2006**). "Biomechanical properties of ruptured versus electively repaired abdominal aortic aneurysm wall tissue," J. Vasc. Surg. **43**, 570–576.
Dyson, M., Pond, J. B., Joseph, J., and Warwick, R. (**1968**). "Stimulation of tissue regeneration by means of ultrasound," Clin. Sci. **35**, 273–&.
Forsberg, F., Merton, D. A., and Goldberg, B. B. (**2006**). "*In vivo* destruction of ultrasound contrast microbubbles is independent of the mechanical index," J. Ultrasound Med. **25**, 143–144.
Greenleaf, W. J., Bolander, M. E., Sarkar, G., Goldring, M. B., and Greenleaf, J. F. (**1998**). "Artificial cavitation nuclei significantly enhance acoustically induced cell transfection," Ultrasound Med. Biol. **24**, 587–595.
Kinoshita, M., McDannold, N., Jolesz, F. A., and Hynynen, K. (**2006**). "Noninvasive localized delivery of Herceptin to the mouse brain by MRI-guided focused ultrasound-induced blood-brain barrier disruption," Proc. Natl. Acad. Sci. U.S.A. **103**, 11719–11723.
Klibanov, A. L., Ferrara, K. W., Hughes, M. S., Wible, J. H., Wojdyla, J. K., Dayton, P. A., Morgan, K. E., and Brandenburger, G. H. (**1998**). "Direct video microscopic observation of the dynamic effects of medical ultrasound on ultrasound contrast microspheres," Invest. Radiol. **33**, 863–870.
Koike, H., Tomita, N., Azuma, H., Taniyama, Y., Yamasaki, K., Kunugiza, Y., Tachibana, K., Ogihara, T., and Morishita, R. (**2005**). "An efficient gene transfer method mediated by ultrasound and microbubbles into the kidney," Journal of Gene Medicine **7**, 108–116.
Kost, J., Mitragotri, S., Gabbay, R. A., Pishko, M., and Langer, R. (**2000**). "Transdermal monitoring of glucose and other analytes using ultrasound," Nat. Med. **6**, 347–350.
Kruse, D. E., and Ferrara, K. W. (**2005**). "A new imaging strategy using wideband transient response of ultrasound contrast agents," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 1320–1329.
Lindau, O., and Lauterborn, W. (**2003**). "Cinematographic observation of the collapse and rebound of a laser-produced cavitation bubble near a wall," J. Fluid Mech. **479**, 327–348.
Marmottant, P., and Hilgenfeldt, S. (**2003**). "Controlled vesicle deformation and lysis by single oscillating bubbles," Nature (London) **423**, 153–156.
Marmottant, P., Versluis, M., de Jong, N., Hilgenfeldt, S., and Lohse, D. (**2006**). "High-speed imaging of an ultrasound-driven bubble in contact with a wall: Narcissus effect and resolved acoustic streaming," Exp. Fluids **41**, 147–153.
McDannold, N., Vykhodtseva, N., and Hynynen, K. (**2006**). "Targeted disruption of the blood-brain barrier with focused ultrasound: Association with cavitation activity," Phys. Med. Biol. **51**, 793–807.
Miller, D. L., and Gies, R. A. (**1998**). "The interaction of ultrasonic heating and cavitation in vascular bioeffects on mouse intestine," Ultrasound Med. Biol. **24**, 123–128.
Mitragotri, S. (**2005**). "Innovation—Healing sound: The use of ultrasound in drug delivery and other therapeutic applications," Nat. Rev. Drug Discovery **4**, 255–260.
Morgan, K. E., Allen, J. S., Dayton, P. A., Chomas, J. E., Klibanov, A. L., and Ferrara, K. W. (**2000**). "Experimental and theoretical evaluation of microbubble behavior: Effect of transmitted phase and bubble size," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1494–1509.
Morgan, K. E., Dayton, P. A., Kruse, D. E., Klibanov, A. L., Brandenburger, G. H., and Ferrara, K. W. (**1998**). "Changes in the echoes from ultrasonic contrast agents with imaging parameters," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 1537–1548.
Postema, M., Van Wamel, A., Lancee, C. T., and De Jong, N. (**2004**). "Ultrasound-induced encapsulated microbubble phenomena," Ultrasound Med. Biol. **30**, 827–840.
Prausnitz, M. R., Mitragotri, S., and Langer, R. (**2004**). "Current status and future potential of transdermal drug delivery," Nat. Rev. Drug Discovery **3**, 115–124.
Qin, S. P., and Ferrara, K. W. (**2006**). "Acoustic response of compliable microvessels containing ultrasound contrast agents," Phys. Med. Biol. **51**, 5065–5088.
Sassaroli, E., and Hynynen, K. (**2005**). "Resonance frequency of microbubbles in small blood vessels: A numerical study," Phys. Med. Biol. **50**, 5293–5305.
Skyba, D. M., Price, R. J., Linka, A. Z., Skalak, T. C., and Kaul, S. (**1998**). "Direct *in vivo* visualization of intravascular destruction of microbubbles by ultrasound and its local effects on tissue," Circulation **98**, 290–293.
Stieger, S. M., Caskey, C. F., Adamson, R. H., Qin, S. P., Curry, F. R. E., Wisner, E. R., and Ferrara, K. W. (**2006**). "Enhancement of vascular permeability with low frequency contrast ultrasound using the chorioallantoic membrane model," Radiology **243**, 112–121.
Taniyama, Y., Tachibana, K., Hiraoka, K., Namba, T., Yamasaki, K., Hashiya, N., Aoki, M., Ogihara, T., Yasufumi, K., and Morishita, R. (**2002**). "Local delivery of plasmid DNA into rat carotid artery using ultrasound," Circulation **105**, 1233–1239.
Unger, E. C., Hersh, E., Vannan, M., and McCreery, T. (**2001**). "Gene delivery using ultrasound contrast agents," Echocardiography (Mount Kisco, N.Y.) **18**, 355–361.
Zhao, S. K., Kruse, D. E., Ferrara, K. W., and Dayton, P. A. (**2006**). "Acoustic response from adherent targeted contrast agents," J. Acoust. Soc. Am. **120**, EL63–EL69.
Zhong, P., Zhou, Y. F., and Zhu, S. L. (**2001**). "Dynamics of bubble oscillation in constrained media and mechanisms of vessel rupture in SWL," Ultrasound Med. Biol. **27**, 119–134.

Automatic classification of killer whale vocalizations using dynamic time warping

Judith C. Brown[a)]
Physics Department, Wellesley College, Wellesley, Massachusetts 02481 and Media Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Patrick J. O. Miller[b)]
Sea Mammal Research Unit, University of St. Andrews, St. Andrews, Fife KY16 9QQ, United Kingdom

(Received 15 November 2006; revised 30 April 2007; accepted 14 May 2007)

A set of killer whale sounds from Marineland were recently classified automatically [Brown *et al.*, J. Acoust. Soc. Am. **119**, EL34–EL40 (2006)] into call types using dynamic time warping (DTW), multidimensional scaling, and kmeans clustering to give near-perfect agreement with a perceptual classification. Here the effectiveness of four DTW algorithms on a larger and much more challenging set of calls by Northern Resident whales will be examined, with each call consisting of two independently modulated pitch contours and having considerable overlap in contours for several of the perceptual call types. Classification results are given for each of the four algorithms for the low frequency contour (LFC), the high frequency contour (HFC), their derivatives, and weighted sums of the distances corresponding to LFC with HFC, LFC with its derivative, and HFC with its derivative. The best agreement with the perceptual classification was 90% attained by the Sakoe-Chiba algorithm for the low frequency contours alone. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2747198]

PACS number(s): 43.80.Ka [WWA] Pages: 1201–1207

I. INTRODUCTION

Marine mammals produce a wide range of vocalizations, and an improved ability to classify recorded sounds could aid in species identification as well as in tracking movements of animal groups. For the most part, the sounds produced by killer whales have been classified by humans into groups called "call types" from listening to the calls and observing their spectra. For killer whale sounds classification by eye and ear is consistent, and this type of classification has been useful to reveal group-specific acoustic repertoires and matching vocal exchanges (Yurk *et al.* 2002). It would, nonetheless, be useful to replace human classification with an automatic technique because of the large amounts of data to be classified, and the fact that automatic methods can be fully replicated in subsequent studies.

In a previous study we examined a group of captive killer whale sounds recorded in Marineland in the French Antilles and consisting of nine call types with at least five examples in each (Brown *et al.* 2006). We found that dynamic time warping (DTW) gives an accurate measure of the dissimilarity of calls and were able to classify this set automatically with near-perfect accuracy. Here we extend this work with a larger group of whale sounds recorded on the open sea and examine the effectiveness of four different DTW algorithms. This set of sounds consists of biphonic (two simultaneous, independently modulated) calls of northern resident whales and contains over 100 calls previously classified perceptually into seven call types. This is the first automatic classification study using frequency contours of biphonic calls as well as the first full-length article on classification of marine mammal calls using DTW. Preliminary results were reported by Brown and Miller (2006a, b).

II. BACKGROUND

A. Killer whale vocalizations

Killer whales produce three forms of vocalizations: clicks, whistles, and pulsed calls. Clicks consist of an impulse train (series of broadband pulses); whistles consist, for the most part, of a single sinusoid with varying frequency; and pulsed calls are more complex sounds with many harmonics. Among these pulsed calls are a number of highly stereotyped (repeated and recognizable) calls, which are thought to be learned within the pod or living group. Repertoires of these stereotyped calls are pod specific, and the time-frequency contours of shared stereotyped calls are also group specific from matrilineal lines (group with same mother) to larger pods (consisting of several matrilineal lines) to clans (larger groups sharing calls).

B. Fundamental frequency tracking and perceptual classification

One of the remarkable features of some northern resident killer whale pulsed calls is that they contain two overlapping but independently modulated contours or "voices" as shown in Fig. 1. Biphonation, as this is called, is common in birds but has been described for few marine mammal sounds (Tyson, 2006; Tyson *et al.*, 2006). One of the challenges of analyzing these complex sounds is to determine the fundamental frequency or to "pitch-track" these two components

[a)]Electronic mail: brown@media.mit.edu
[b)]Electronic mail: pm29@st-andrews.ac.uk

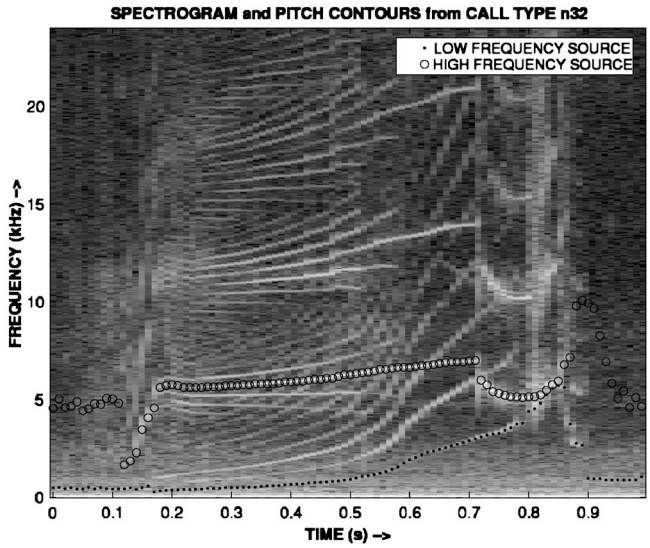


FIG. 1. Spectrogram showing pitch contours of the low frequency and high frequency sources in a killer whale pulsed call. Note there is noise before and after the onset of the calls.

from the same sound. See the example in Fig. 1 where the upper and lower frequency components are superposed on the spectrum.

Pitch tracking has had a long and abundant history in the speech literature (Hess, 1983). Some of these methods have proven successful for determining the repetition rate, or fundamental frequency, for pulsed killer whale sounds and have been described in Brown (1992), Wang and Seneff (2000), and Brown *et al.* (2004). The pitch contours of our northern resident group are arranged by perceptually determined call types in Fig. 2 for the low frequency contours and in Fig. 3 for the high frequency contours. As can be seen in these figures, the shapes of the contours within each group are similar though the lengths of the calls differ. The call types are graphed separately because of the considerable overlap in frequency range of several of the groups; this foreshadows difficulty for automatic classification.

C. Dynamic time warping

For automatic classification, a technique for quantitatively comparing curves of similar shape but different lengths is required. Dynamic time warping (DTW) is ideally suited to this task. It was used widely in the early days of speech recognition, and the different algorithms used by the speech community are described and evaluated in an excellent paper by Myers *et al.* (1980). See also Rabiner and Juang (1993). More recently DTW has been used for "query by humming" searches in musical information retrieval (Hu *et al.*, 2003).

For marine mammal sounds DTW was first used for the classification of 15 dolphin signature whistles into five groups by Buck and Tyack (1993). In the past year it has been applied to pulsed killer whale sounds by Deecke and Janik (2006) on a set of 20 calls in six categories, as well as our (Brown *et al.*, 2006) Marineland classification of 57 calls into nine call types. In the smaller sets of 15 and 20 calls, the contours within call types were virtually identical. While

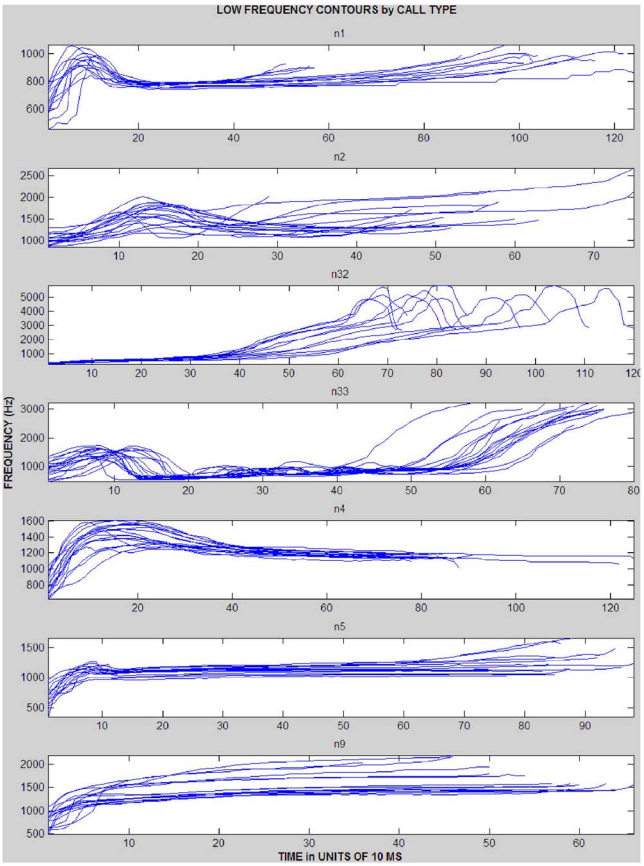


FIG. 2. (Color online) Pitch contours of the low frequency calls of the northern resident group of killer whales plotted in the perceptually designated call types. The calls are plotted separately since there is too much overlap to distinguish them if plotted on the same graph.

there was much more contour variation in the 57-call data set, these calls still separated sufficiently in absolute frequency to be identifiable on the same graph; this is not the case for our current, much larger set of calls.

We have chosen four very different DTW algorithms, including the three used previously in the marine mammal studies mentioned above, for our current classification to determine their relative performance on this extremely challenging set of calls.

III. CALCULATIONS

A. Dynamic time warping (DTW) and contour dissimilarity

As an example of a DTW calculation, we consider two calls of different lengths, both from call type n32. By convention the shorter call is referred to as the query $\mathbf{Q}[i]$ and is aligned along a vertical axis, and the longer call is the target $\mathbf{T}[j]$ aligned horizontally as shown in Fig. 4. For all algorithms the first step is to construct a difference matrix where each element $\mathbf{D}[i,j]$ is equal to the difference in corresponding elements,

$$\mathbf{D}[i,j] = |\mathbf{Q}[i] - \mathbf{T}[j]|. \quad (1)$$

From this difference matrix, a cost matrix $\mathbf{M}$ is constructed that keeps a running tab on the dissimilarities of the elements making up the curves while adding up these costs

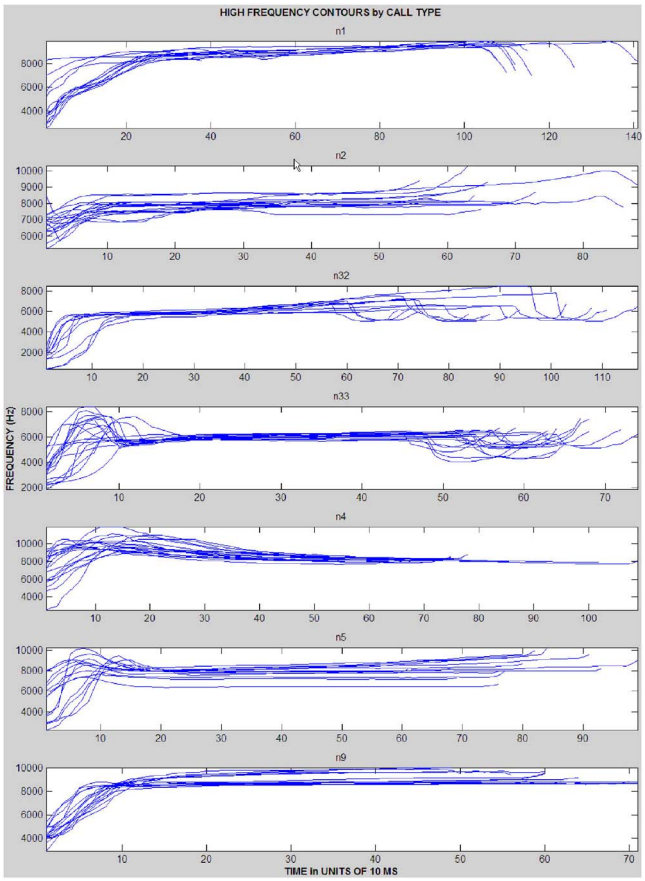


FIG. 3. (Color online) Pitch contours of the high frequency calls of the northern resident group of killer whales plotted in perceptually designated call types. They are plotted separately due to overlap as in Fig. 2.

to give a final number called the "dissimilarity" or distance between the query and target. We examine the cost matrices of our four algorithms below.

1. Ellis method

This is the simplest and most straightforward algorithm.[1] Each element of the cost matrix is obtained by adding the difference element for that position [obtained from Eq. (1)] to the minimum of the three previously determined elements of the cost matrix, which are (1) diagonal, (2) above, and (3) to the left:

$$\mathbf{M}[i,j] = \min\begin{pmatrix} \mathbf{M}[i-1,j-1] \\ \mathbf{M}[i-1,j] \\ \mathbf{M}[i,j-1] \end{pmatrix} + \mathbf{D}[i,j]. \quad (2)$$

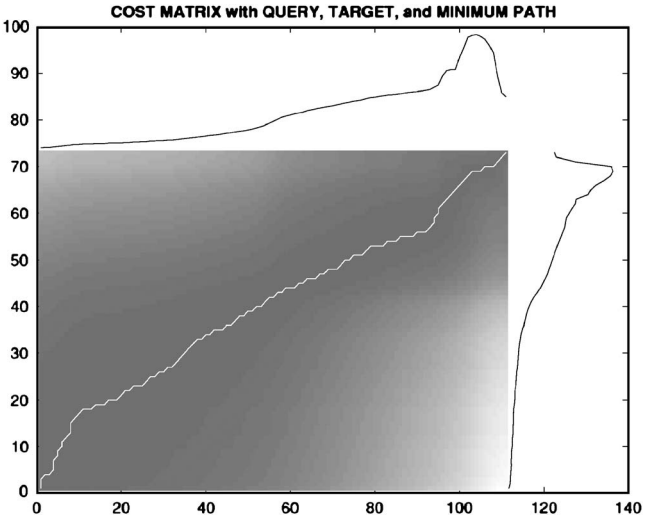


FIG. 4. Cost matrix with minimum cost path in white and input contour shapes above and to the right. The shorter sound is called the query and the longer sound the target.

2. Sakoe-Chiba method

The method of Sakoe and Chiba (1978) in an altered form was used by Deecke and Janik (2006). It is more complex and compares the weighted sum of difference elements from two columns and two rows distant with the weighted diagonal as shown in the equation below. We have chosen the form indicated by Sakoe and Chiba to give the best results:

$$\mathbf{M}[i,j] = \min\begin{pmatrix} \mathbf{M}[i-1,j-1] + 2\cdot\mathbf{D}[i,j] \\ \mathbf{M}[i-2,j-1] + 2\cdot\mathbf{D}[i-1,j] + \mathbf{D}[i,j] \\ \mathbf{M}[i-1,j-2] + 2\cdot\mathbf{D}[i,j-1] + \mathbf{D}[i,j] \end{pmatrix}. \quad (3)$$

3. Itakura method

This method (Itakura, 1975) was used by Buck and Tyack (1993):

$$\mathbf{M}[i,j] = \min\begin{pmatrix} \mathbf{M}[i-2,j-1] \\ \mathbf{M}[i-1,j-1] \\ \mathbf{M}[i,j-1] \end{pmatrix} + \mathbf{D}[i,j]. \quad (4)$$

It differs from other algorithms in that there is a constraint that two elements cannot be chosen sequentially from the same row, i.e., if $\mathbf{M}[i,j-1]$ is the minimum element, then it is not an option for the next element of the cost matrix in that row.

4. Chai-Vercoe method

This is the method often used in music information retrieval (Chai and Vercoe, 2003; Foote, 2000; Kruskal and Sankoff, 1983) and was extremely successful in classifying our killer whale calls from Marineland. The cost matrix is generated with

$$\mathbf{M}[i,j] = \min\begin{pmatrix} \mathbf{M}[i-1,j] + a, \\ \mathbf{M}[i,j-1] + b, \\ \mathbf{M}[i-1,j-1] + \mathbf{D}[i,j] \end{pmatrix}. \quad (5)$$

Here each element of the cost matrix can come from (1) the cost element directly above and adding a, the cost of an insertion; (2) the cost element to the left and adding b, the cost of an deletion; or (3) the previous element along the diagonal with the addition of the difference in corresponding elements. Since a deletion means a difference in lengths,

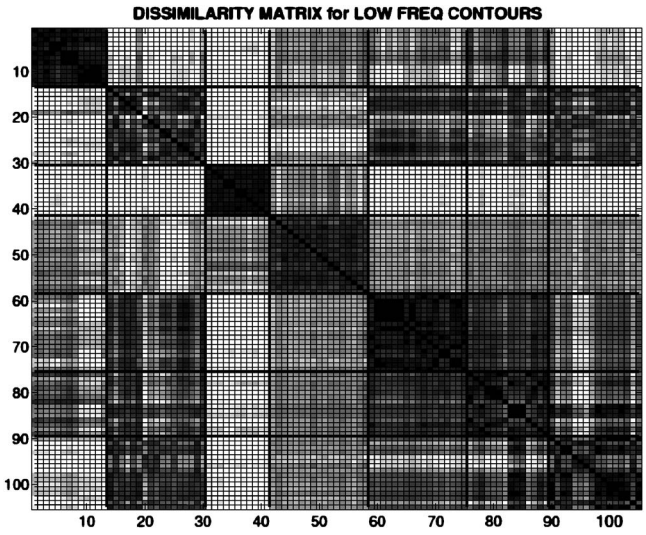


FIG. 5. Dissimilarity matrix of the LFC using the method of Sakoe and Chiba. Each point represents the dissimilarity of (or distance between) a pair of sounds.

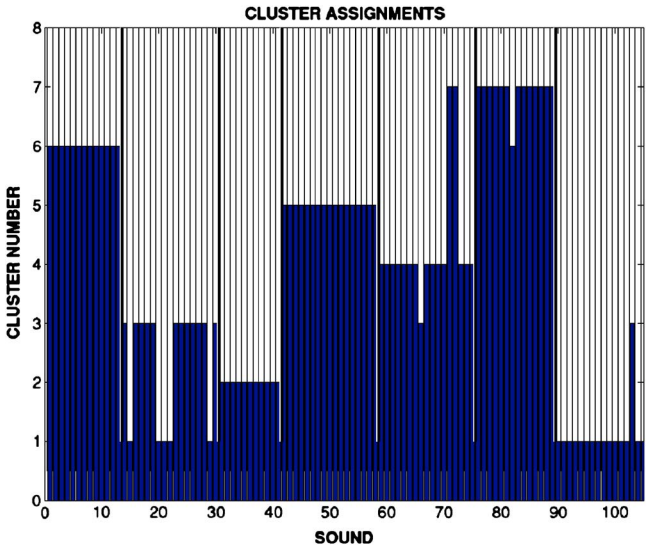


FIG. 6. (Color online) Clustering results for the difference matrix of Fig. 5.

which we do not want to penalize, b was chosen to be 0. The principal disadvantage of this method is that it contains the adjustable parameter "a."

For each of these four methods a running tab is kept of which choice is made for each element. Thus the minimum path can be retraced, and an example is shown in Fig. 4. The final "dissimilarity" is the number $\mathbf{M}[i_{\max}, j_{\max}]$ normalized by dividing by the length of the query; this is a measure of the difference in the two contours. Identical signals will have a diagonal best path and a cost of zero, while larger differences will increase the matching cost. For classification these costs are a means of grouping (clustering) the calls with the smallest dissimilarities.

5. Dissimilarity matrices

Dissimilarity matrices were obtained by calculating a cost matrix for each pair of the low frequency calls shown in Fig. 2 to give a matrix with elements equal to these dissimilarities. The frequencies in Hz, which are graphed, were transformed for the cost matrix calculation using

$$f_{\text{cents}} = 12 \log_2(f/f_{\text{ref}}), \tag{6}$$

where f_{ref}=440 Hz as described in Brown *et al.* (2006). This unit means that we are comparing ratios of frequencies rather than absolute values, and, for example, a difference of 100 Hz and 200 Hz will be weighted the same as a difference of 400 Hz and 800 Hz. An identical procedure was carried out for obtaining a dissimilarity matrix for the high frequency calls shown in Fig. 3 as well as the derivatives (point to point differences of each curve in Figs. 2 and 3) for both groups. The derivatives are a measure of the shape alone of the curves and are independent of absolute frequency.

An example of a dissimilarity matrix is given in Fig. 5. Each element of this matrix represents the result of calculating a cost matrix for a particular pair of calls. Since there were a total of 105 calls, each dissimilarity matrix represents the results of calculating a cost matrix for 105(104)/2 or 5460 pairs of calls. The matrix is not truly symmetric but comparison of the shorter (query) to the longer (target) sound has been found to be a more accurate measure of the difference; therefore the elements below the diagonal were obtained by transposition.

The example of Fig. 5 was calculated for the low frequency calls using the Sakoe-Chiba method. Here the dark elements indicate a small distance and the lighter ones a larger distance. The perceptual groupings of Fig. 2 are indicated with bold horizontal and vertical lines. Perfect agreement with the perceptual results would give black blocks along the diagonal corresponding to small distances for the perceptual groupings and white elsewhere indicating large distances. The third group (n32) is closest to this ideal with white for all other groups except the fourth (n33). The last three groups are mixed with each other as well as with the second group, and this was typical of all calculations.

B. Classification

For each method the distances given by the dissimilarity matrices were transformed into a Euclidean-like space using multi-dimensional scaling. They were then clustered using a kmeans algorithm (Brown *et al.*, 2006) from Matlab into seven call types to compare to the perceptual classification. An example classification corresponding to the dissimilarity matrix of Fig. 5 for the Sakoe-Chiba method is given in Fig. 6. There are ten errors in this example all involving clusters 2, 5, 6, and 7, as could be predicted from the dissimilarity matrix.

IV. RESULTS AND DISCUSSION

Classification results for each of the four algorithms used to classify the low frequency component (LFC), the high frequency component (HFC), and their derivatives are given numerically in Table I as well as in the corresponding bar graphs of Figs. 7–9, where they are more easily visualized. In Table I the column labeled "Double group" indicates

TABLE I. Summary of results. The upper third of the table gives the percent agreement with the perceptual classification of the LFC and HFC contours alone in columns 1 and 3 and for their sum in column 6. The middle third of the table does the same for the LFC and its derivative. The lower third does the same for the HFC and its derivative.

	Summary of results						
	Low frequency and high frequency components						
	Low freq	Double group	High freq	Double group	Ratio	Sum	Double group
Ellis	77		70		1.6	80	
Sakoe-Chiba	90		69		1.6	90	
Itakura	86		68	1	1.6	81	
Chai-Vercoe	83	1	79	1	1.8	90	1
	Low frequency component and its derivative						
	Low freq	Double group	Low freq derivative	Double group	Ratio	Sum	Double group
Ellis	77		81	1	12	77	
Sakoe-Chiba	90		82	1	12	88	
Itakura	86		70	1	11	73	
Chai-Vercoe	83	1	86	1	20	86	1
	High frequency component and its derivative						
	High freq		High freq derivative	Double group	Ratio	Sum	Double group
Ellis	70		76		14	77	
Sakoe-Chiba	69		76		16	86	
Itakura	68	1	57	1	14	73	
Chai-Vercoe	79	1	77	1	26	80	1

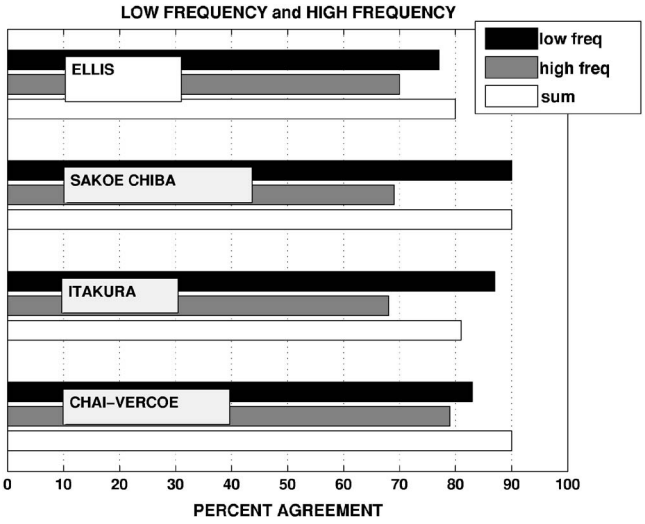


FIG. 7. Percent agreement with perceptual results for each method of calculating for the LFC, HFC, and their sum.

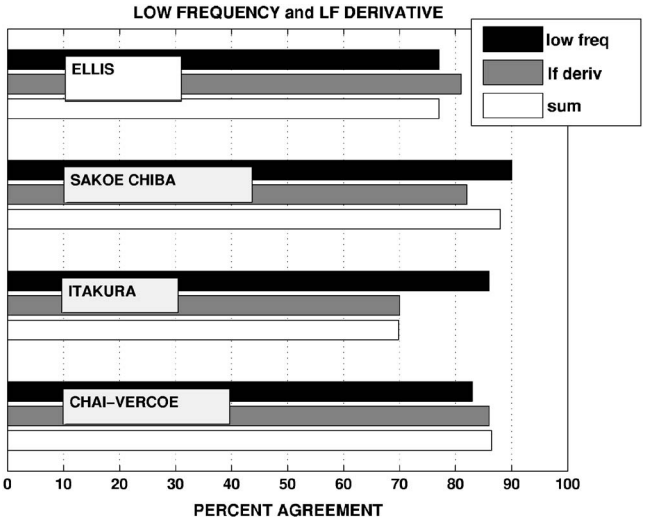


FIG. 8. Percent agreement with perceptual results for each method of calculating for the LFC, LFC derivative, and their sum.

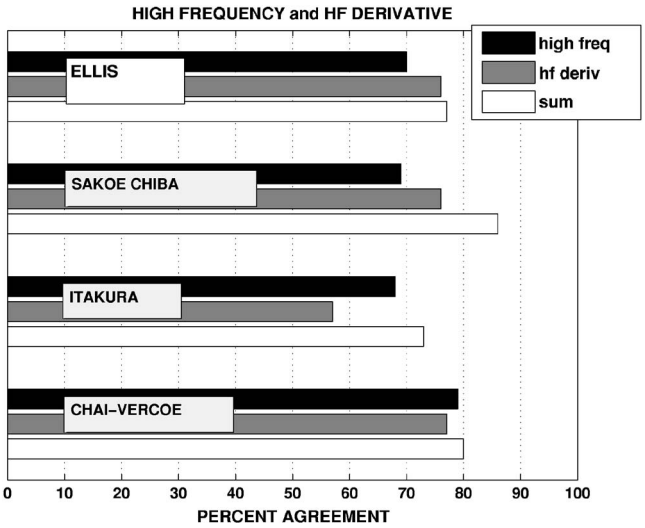


FIG. 9. Percent agreement with perceptual results for each method of calculating for the HFC, HFC derivative, and their sum.

that two of the perceptual call types were classified into the same cluster, so these results are not as good as the stated percentage would indicate.

A. Low frequency and high frequency contours

In the upper third of Table I (cf. Fig. 7), the agreement with perceptual results is greater for the low frequency component (LFC in column 1) than the high frequency component (HFC in column 3). Sakoe-Chiba does best on the LFC at 90% with Chai-Vercoe best for the HFC. To determine the effect of the LFC and HFC distances combined, their dissimilarity matrices were added with weighting corresponding to the ratio of the means (column 5), and a new classification calculation was carried out for each algorithm. The results are in column 6 with significant improvement for the Chai-Vercoe method.

B. Low and high frequency contour derivatives

Classification results for the LFC and HFC derivatives are found in column 3 of the remainder of Table I as well as in Figs. 8 and 9. With one exception these results are all over 70%, which is quite remarkable for these very irregular curves.

C. Sum of contours and derivatives

Results for the Marineland group (Brown *et al.*, 2006) were improved from 88% to 98% agreement with perceptual results by adding the dissimilarity matrix for the LFC derivative to the matrix for LFC alone weighted with the means of the two matrices. Results of the analogous calculation for this set of calls are given in column 6. Here they yield a marginal improvement for the Chai-Vercoe method. For the HFC summed with the HFC derivative calculation, Sakoe-Chiba was improved by 10% and 17%, respectively, over the HFC derivative and HFC alone results to 86%.

D. Summary

With the exception of the Itakura method on the HFC derivatives, the results of all algorithms were in agreement with the perceptual classification by near 70% or greater and could thus be considered successful given the difficulties of this data set. The shapes of the curves show variation within each perceptual call type, and there is considerable similarity among groups 2, 5, 6, and 7 (call types n2, n4, n5, and n9) in frequency range as well as shape. The best result was 90% using Sakoe-Chiba for the low frequency contours, which is truly outstanding.

It should be recalled that the perceptual classification was made by listening to the calls while observing their spectra, rather than by an examination of the contour alone. These perceptual decisions were probably influenced by spectral content (not present in the contours). Also DTW is most effective for curves differing in length by less than a factor of 2; in this set there was variation of lengths as great as a factor of 3. Thus, it is in fact remarkable that the computer classification reached 90% agreement.

V. CONCLUSIONS

These results with a maximum of 90% agreement with the perceptual data were not as impressive as the 98% reported previously on the Marineland set. However, this is easily understood in comparing Fig. 2 to the corresponding figure in Brown *et al.* (2006). The Marineland calls separated nicely in frequency and could be viewed on the same graph. In a similar graph (not included) for these northern resident calls, four of the call types were intermingled and unseparable visually. In other DTW studies on marine mammals (Buck and Tyack, 1993; Deeke and Janik, 2006) there were few contours, and they were virtually identical within groups. The current data set thus represented a severe test for DTW, and the 70%–90% agreement with perceptual classification is excellent.

Of the algorithms examined and combinations of dissimilarities, Sakoe-Chiba performed best on the LFC. While slightly more complicated than the other algorithms, it has the advantage of having no adjustable parameters. There is also a positive side to the fact that results were best for the low frequency component alone in that preprocessing reduces to pitch tracking a single component.

Dynamic time warping has proven to be an excellent technique for the automatic classification of killer whale call types. One of its most rewarding applications would be the ability to monitor the movements and habitat preferences of killer whales just by tracking sounds heard at remote monitoring stations. This will only be possible with systems developed to automatically process and identify calls heard at those locations so that the group producing them can be identified remotely.

ACKNOWLEDGMENTS

JCB is grateful to Dr. Eamonn Keogh for his MATLAB code, which was adapted to give Fig. 4. Funding was provided by WHOI's Ocean Life Institute and a Royal Society fellowship to PJOM.

[1]We are calling this the Ellis method as the code was obtained from Dan Ellis's website http://labrosa.ee.columbia.edu/matlab/dtw/.

Brown, J. C. (**1992**). "Musical fundamental frequency tracking using a pattern recognition method," J. Acoust. Soc. Am. **92**, 1394–1402.

Brown, J. C., Hodgins-Davis, A., and Miller, P. J. O. (**2004**). "Calculation of repetition rates of the vocalizations of killer whales," J. Acoust. Soc. Am. **116**, 2615.

Brown, J. C., Hodgins-Davis, A., and Miller, P. J. O. (**2006**). "Classification of vocalizations of killer whales using dynamic time warping," J. Acoust. Soc. Am. **119**, EL34–EL40.

Brown, J. C., and Miller, P. J. O. (**2006a**). "Dynamic time warping for automatic classification of killer whale vocalizations," J. Acoust. Soc. Am. **119**, 3434.

Brown, J. C., and Miller, P. J. O. (**2006b**). "Classifying killer whale vocalization using time warping," Echoes **16**, 45–47.

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.

Chai, W., and Vercoe, B. (**2003**). "Structural analysis of musical signals for indexing and thumbnailing," Proceedings of ACM/IEEE Joint Conference on Digital Libraries.

Deecke, V. B., and Janik, V. M. (**2006**). "Automated categorization of bioacoustic signals: Avoiding perceptual pitfalls," J. Acoust. Soc. Am. **119**, 645–653.

Foote, J. (**2000**). "ARTHUR: Retrieving orchestral music by long-term structure," Proc. of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000), pp. 1–6.

Hess, W. (**1983**). *Pitch Determination of Speech Signals: Algorithms and Devices* (Springer-Verlag, Berlin).

Hu, N., Dannenberg, R. B., and Tzanetakis, G. (**2003**). "Polyphonic audio matching and alignment for music retrieval," IEEE Workshop on Applications of Signal Processing to Audio, New Paltz, NY.

Itakura, F. (**1975**). "Minimum prediction residual principle applied to speech recognition," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-23**, 57–72.

Kruskal, J., and Sankoff, D. (**1983**). "An anthology of algorithms and concepts for sequence comparison," in *Time Warps, String Edits, and Macromolecules: The Theory and Practice of String Comparison*, edited by D. Sankoff and J. Kruskal (Addison-Wesley, Reading, MA).

Myers, C., Rabiner, L. R., and Rosenberg, A. E. (**1980**). "Performance tradeoffs in dynamic time warping algorithms for isolated word recognition," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-28**, 623–634.

Rabiner, L., and Juang, B. H. (**1993**). *Fundamentals of Speech Recognition* (Prentice Hall, Englewood Cliffs, NJ).

Sakoe, H., and Chiba, S. (**1978**). "Dynamic programming optimization for spoken word recognition," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-26**, 43–49.

Tyson, R. (2006). "The presence and potential functions of nonlinear phenomena in cetacean vocalizations," thesis, Florida State University, Tallahassee, FL.

Tyson, R., Nowacek, D. P., and Miller, P. J. O. (**2006**). "Nonlinear phenomena in the vocalizations of North Atlantic right whales (*Eubalaena glacialis*) and killer whales (*Orcinus orca*)," accepted by J. Acoust. Soc. Am..

Wang, C., and Seneff, S. (**2000**). "Robust pitch tracking for prosodic modeling in telephone speech," Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 3, pp. 1343–1346.

Yurk, H., Barrett-Lennard, L., Ford, J. K. B., and Matkin, C. O. (**2002**). "Cultural transmission within maternal lineages: vocal clans in resident killer whales in southern Alaska," Anim. Behav. **63**, 1103–1119.

Blue and fin whale call source levels and propagation range in the Southern Ocean

Ana Širović,[a] John A. Hildebrand, and Sean M. Wiggins
Scripps Institution of Oceanography, 9500 Gilman Drive, La Jolla, California 92093-0205

(Received 20 October 2006; revised 15 May 2007; accepted 21 May 2007)

Blue (*Balaenoptera musculus*) and fin whales (*B. physalus*) produce high-intensity, low-frequency calls, which probably function for communication during mating and feeding. The source levels of blue and fin whale calls off the Western Antarctic Peninsula were calculated using recordings made with calibrated, bottom-moored hydrophones. Blue whales were located up to a range of 200 km using hyperbolic localization and time difference of arrival. The distance to fin whales, estimated using multipath arrivals of their calls, was up to 56 km. The error in range measurements was 3.8 km using hyperbolic localization, and 3.4 km using multipath arrivals. Both species produced high-intensity calls; the average blue whale call source level was 189±3 dB *re*:1 μPa-1 m over 25–29 Hz, and the average fin whale call source level was 189±4 dB *re*:1 μPa-1 m over 15–28 Hz. Blue and fin whale populations in the Southern Ocean have remained at low numbers for decades since they became protected; using source level and detection range from passive acoustic recordings can help in calculating the relative density of calling whales. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2749452]

PACS number(s): 43.80.Ka, 43.30.Sf [WWA] Pages: 1208–1215

I. INTRODUCTION

Blue (*Balaenoptera musculus*) and fin whales (*B. physalus*) were the primary targets of the commercial whaling industry that developed in the Southern Ocean during the twentieth century. Populations of both species were brought to near extinction before their hunt was banned in the 1960's and 70's (Clapham and Baker, 2001), and their population recovery has been slow (Best, 1993; Branch and Butterworth, 2001; Branch *et al.*, 2004). Both species produce calls that are likely to be an important part of the mating and feeding behaviors (Watkins *et al.*, 1987; McDonald *et al.*, 2001; Croll *et al.*, 2002; Oleson *et al.*, 2007), and it has been established that certain baleen whale calls can be detected at ranges of hundreds of kilometers (Cummings and Thompson, 1971; Payne and Webb, 1971; Clark, 1995; Stafford *et al.*, 1998). Payne and Webb (1971) postulated that long-range propagation might be important for communication with conspecifics over large distances, and the low population densities resulting from commercial whaling (Branch and Butterworth, 2001) could make this type of communication even more important for species survival.

Several methods have been developed for acoustic localization and source level estimation in the marine environment (e.g., Frazer and Pecholcs, 1990; Cato, 1998; Jensen *et al.*, 2000; Spiesberger, 2001). The theory was developed predominately for naval and seismic purposes, but similar methods can be used to determine locations and source levels of calling cetaceans in the wild (Watkins and Schevill, 1972; McDonald *et al.*, 1995; Stafford *et al.*, 1998; McDonald and Fox, 1999; Clark and Ellison, 2000; Thode *et al.*, 2000; Charif *et al.*, 2002). Blue and fin whales make distinctive low-frequency, high-intensity calls that vary geographically (Cummings and Thompson, 1971; Watkins, 1981; Edds, 1982; 1988; Clark, 1995; McDonald *et al.*, 1995; Ljungblad *et al.*, 1998; Stafford *et al.*, 1999; McDonald *et al.*, 2006), and their source levels have been estimated at several worldwide locations. Cummings and Thompson (1971) estimated source level of blue whale moans off Chile in the 14 to 222-Hz band to be 188 dB *re*:1 μPa at 1 m. Calls of blue whales from the eastern North Pacific Ocean had maximum intensity 180–186 dB *re*:1 μPa at 1 m over the 10–110-Hz band (Thode *et al.*, 2000; McDonald *et al.*, 2001). Fin whale downswept call source levels have been reported at 160–186 dB *re*:1 μPa at 1 m in the western North Atlantic and between 159 and 184 dB *re*:1 μPa at 1 m in the eastern North Pacific Ocean (Watkins, 1981; Watkins *et al.*, 1987; Charif *et al.*, 2002). Northrop *et al.* (1968) reported fin whale downsweeps of even higher intensity in the Central Pacific Ocean, ranging between 164 and 199 dB *re*:1 μPa at 1 m, albeit assuming relatively high transmission loss.

Frequency and temporal characteristics of blue and fin whale calls in the Southern Ocean have been described previously (Ljungblad *et al.*, 1998; Širović *et al.*, 2004; Rankin *et al.*, 2005). Blue whale calls last up to 18 s and generally consist of three segments: a 9-s-long, 27-Hz tone, followed by a 1-s downsweep to 19 Hz and another, longer-lasting downsweep to 18 Hz (Širović *et al.*, 2004; Rankin *et al.*, 2005). Fin whales produce short (< 1 s) downsweeps from 28 to 15 Hz (Širović *et al.*, 2004, 2006). Calls of both species are usually repeated at regular intervals. No call source levels from either species have been reported for the Southern Ocean.

Call intensity may be important for successful intraspecific communication over long distances, and needs to be quantified before we can understand the potential impacts of

[a]Current address: Southwest Fisheries Science Center, NMFS/NOAA, 8604 La Jolla Shores Dr., La Jolla, CA 92037. Electronic mail: ana.sirovic@noaa.gov

 0001-4966/2007/122(2)/1208/8/$23.00

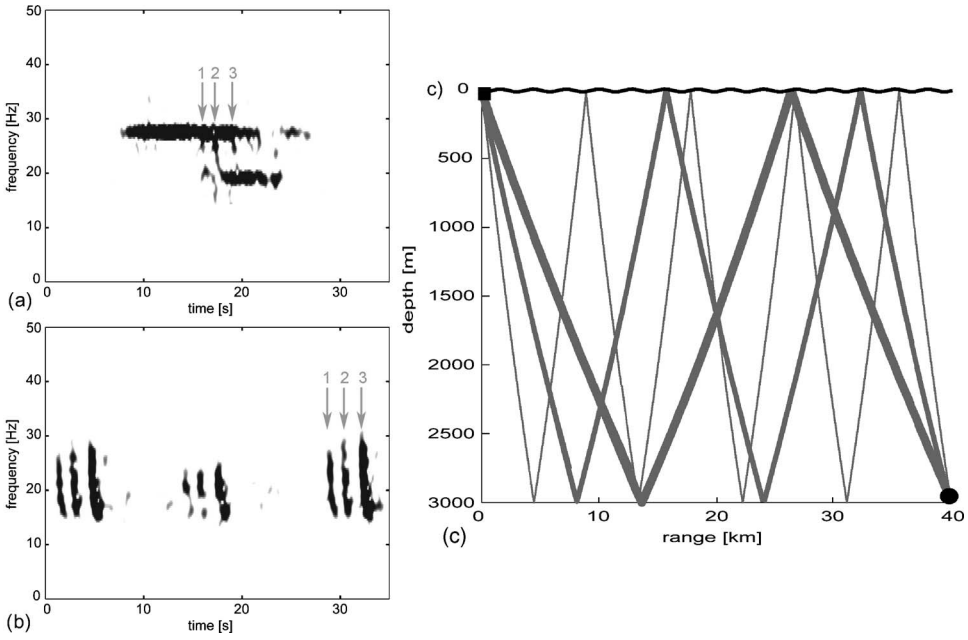


FIG. 1. Blue (a) and fin whale (b) calls recorded off the Western Antarctic Peninsula, showing multipath arrivals. In both examples, paths shown were first, second, and third bounces (marked 1, 2, and 3, respectively); direct path is not visible. Calculated ranges were 33 km for the blue whale and 40 km for the fin whale. Theoretical contributing bounces for the fin whale path arrivals are shown in part (c), with the thick line representing the first bounce, the medium thickness line for the second bounce, and the thin line for the third bounce. Calling whale location is denoted with a black square and the receiving ARP location is shown by a black circle.

anthropogenic noise on these animals. In this paper, we report the average estimated source levels for blue and fin whale calls recorded off the Antarctic Peninsula and investigate the variation in the source levels within the population. Also, we calculate the ranges over which these calls can be expected to propagate, using the average noise levels for this region.

II. METHODS

Acoustic data were recorded using Acoustic Recording Packages (ARPs) deployed off the Western Antarctic Peninsula between March 2001 and February 2003. Detailed information on ARPs, these deployments, and temporal characteristics of blue and fin whale calls used in the analyses is given in Wiggins (2003) and Širović *et al.* (2004). The ARPs were not navigated after deployment for precise locations and the maximum error in the deployment locations is less than 1 km, given the average ARP sinking speed (40 m/min to 3500-m depth) and assuming maximum speed of the Antarctic Circumpolar Current [15 cm/s, Pickard and Emery (1990)].

The goals of the study were to calculate blue and fin whale call source levels and to estimate the maximum range over which these calls can be heard. Data needed for call source level estimation are the instrument response, distance to the calling whales, and knowledge of the ocean propagation environment. Two methods were used to determine range to the calling whales: multipath arrivals and time difference of arrivals.

A. Multipath arrivals

As sound travels through the water column from the source to the receiver, it can follow a direct path, or it can be reflected off the surface and the bottom. The arrival time differences of those multipaths to a single receiver can be used to determine the distance between the source and the receiver. Both blue and fin whale calls were suitable for this analysis because the downswept parts of their calls made it possible to distinguish exact multipath arrival times (Fig. 1). Arrival time for the downsweep was measured in the time-frequency domain at the time of the highest frequency for all multipaths, and the differences between the multipath arrival times were calculated. Spectral parameters were set to 500-point FFT and 90% overlap. Calls with multiple arrivals were found only at one instrument at a time and only calls with three or more multipath arrival times and good signal-to-noise ratios were used in the analysis. The error in the calculation of the arrival time differences was determined by taking multiple measurements of the multipath arrival times of an individual whale call. The range to the calling whale was calculated separately for each measurement and the standard deviation of those ranges was reported.

The following assumptions were made in the multipath arrival model: whale calling occurred near the surface, instruments were located on the bottom, the sound-speed profile was homogeneous ($c=1480$ m/s), and the bottom was flat. Blue whales are known to make calls at depths of 20–30 m (Thode *et al.*, 2000; Oleson *et al.*, 2007), and the calling depth for fin whales is reported to be around 50 m (Watkins *et al.*, 1987). The hydrophone was suspended 10 m above the ocean floor. Given the water column depth of around 3000 m, differences in water column depth $<$ 100 m could reasonably be approximated as calling at the surface and receiver on the bottom. All the ARPs used in these analyses were deployed in locations close to the shelf break, but the regions away from the shelf break had a relatively flat or slightly sloping bottom. This region is an upward-refracting environment (Urick, 1983), so the calls produced in the relatively shallow water on the shelf and shelf break could not be recorded by the ARPs located in deep water (see Sec. II D below). Therefore, whales that were recorded on the ARPs were known be located in the region away from the shelf break, and flat bottom was a good assumption.

The range was determined by comparing the measured arrival time differences with the modeled data. The measured

arrivals were assigned to successive modeled bounce times to determine a possible range for each arrival separately, starting from the direct path and the first measured arrival and stopping at the sixth bounce and the last measured arrival. The average range and standard deviation were calculated for each sequence of measured arrival-bounce path pairs. The range with the lowest standard deviation was used for all further calculations.

Determining the range to calling animals using multipath arrivals was possible only at times when there were no overlapping calls. This method estimated only the distance to the calling whale from the ARP, not the location of the calling whale. The range information, however, was sufficient for source level calculations.

B. Time difference of arrival and hyperbolic localization

Blue whale calls were recorded on an array of ARPs, enabling comparisons of arrival times of the same call to multiple instruments. To use time difference of arrival (TDOA) for determining range and location, a minimum of three instruments needs to receive the same call (Spiesberger, 2001). Periods when the same calls were recorded on multiple instruments were identified by finding sections that had blue whale call sequences with matching intercall intervals. This was possible because the ambient noise at this frequency range is low in the Antarctic, there were not many other calling animals present, and blue whale calls are produced in long, repetitive sequences. Search times were limited by the maximum possible travel time difference between the instruments. Once a matching sequence was identified on three instruments, arrival times of blue whale calls to each instrument were measured by an analyst in the time-frequency domain (i.e., using spectrograms with 500-point FFT and 90% overlap). The point used as the arrival time was the beginning of the first downswept segment of the blue whale call (Fig. 1). Instrument clocks drifted between 2:54 and 5:57 over the course of the deployment period (321 days). We corrected the times assuming linear drift and calculated the TDOA for each instrument pair.

The TDOA between pairs of instruments confine possible locations of the calling animal in two dimensions to a hyperbola. When multiple pairs of instruments are used, the intersections of these hyperbolas give the location of the caller. Hyperbolic localization software developed and made available by Mellinger was used for localization. This localization method assumed homogeneous sound-speed profile ($c=1480$ m/s). The location of the caller was calculated using the Lavenberg-Marquardt nonlinear least-squares optimization of the resulting intersections of the three hyperbolas. Range from the whale to each instrument was calculated from the resulting location. The mean error was calculated as the difference between the actual and theoretically calculated optimized TDOA (Clark and Ellison, 2000). The geometry of the ARP array resulted in an east-west ambiguity for all the localizations. The ambiguity was resolved due to the bathymetric constraints of the environment (Spiesberger, 2001), using BELLHOP ray trace modeling (see Sec. II D below). However, the range value is the same for both solutions, so even if the ambiguities in the hyperbolic localization results were not resolved, the source level results would not be affected. This method was feasible only for blue whale localization.

We compared the two methods using blue whale calls which exhibited multipath arrivals and which could be located using TDOA. The range results were calculated from 14 blue whale calls, from assumed at least four different whales on three different days using the two methods. A paired t-test was performed to determine if the results obtained using these two methods were significantly different and the average difference between the results is reported.

C. Source level calculations

The call source level was calculated as the sum of the measured received level (RL) and the calculated transmission loss (TL). The received level was measured for all calls with calculated range from time-averaged power spectrum densities. Power spectra were calculated using 500-point FFT, 90% overlap, and Hanning window. Parseval's theorem was applied to calculate the total received level in the frequency band of interest. For blue whale calls, 6 s of the call over the 25–29-Hz frequency band prior to the first downsweep were used. Fin whale call received level was measured over a frequency band 15–28 Hz starting at the beginning of the call and lasting 1 s. The hydrophones used for received level measurements were calibrated at the U.S. Navy facility in Point Loma, CA. System frequency response from 10–250 Hz was measured and a calibration of -71.3 dB $re:\text{counts}^2/\mu\text{Pa}^2$ in the 20–30-Hz band was applied to the measured received levels (McDonald, 2005).

The transmission loss can be described as a function of range (r) as follows:

$$\text{TL} = X \log\left(\frac{r}{r_0}\right),$$

where X is the environment-dependent transmission loss coefficient, and r_0 is the reference range, taken to be 1 m. X has the value of 10 under cylindrical and 20 under spherical spreading conditions. While the ranges over which the calls propagated were much larger than the depth of the seafloor and thus spherical spreading did not apply, the polar environment is generally upward refracting (Urick, 1983) and is a propagation environment that is an intermediate between cylindrical and spherical spreading assumptions. To estimate the value of X applicable for this study, we used an empirical method where the transmission loss coefficient was calculated from the relationship between the received levels and the ranges of blue whale calls calculated using hyperbolic localization,

$$X = \frac{\text{RL}_2 - \text{RL}_1}{\log(r_1) - \log(r_2)}.$$

Data from all the blue whale calls had to be pooled to obtain a large enough range distribution to smooth out convergence effects and provide a robust X estimate. This empirical value of X was verified theoretically using BELLHOP incoherent

transmission loss models with the appropriate environmental parameters (see Sec. II D below). In this case, bathymetry was assumed to be upwards sloping, with a steep shelf break on one side.

The source level of each blue whale call was calculated separately for each instrument, giving three estimates. The average of these three values was used as the calculated source level of each call. Standard deviation of each estimate was calculated, and their average is reported and compared to the expected variation in the source level based on the error in range estimation. Only one source level estimate was available for each fin whale call because each range was calculated using only a single instrument recording.

D. Sound propagation modeling

BELLHOP ray-trace modeling was used to verify if calls produced on the shelf could be heard on the ARPs, to resolve the east-west ambiguity in the hyperbolic localization results, and to check the flat bottom assumption from the multipath model. For this problem, we assumed the calling whale was 5 km from the edge of the shelf (which was less than the minimum distance from the hyperbolic localization results) and that the depth increased from 500 m on the shelf to 3500 m off the shelf, over a 15-km distance, and then sloped gradually. The following assumptions were the same for both transmission loss modeling, and the resolution of the east-west ambiguity. The ocean and the bottom sound-speed properties were range independent. The sound-speed profile was obtained from the average of expendable bathythermograph (XBT) casts in the vicinity of the instruments during the seasons when calls were localized. Source depth was 30 m, and we used multiple receiver depths and ranges, at 100-m and 1-km intervals, respectively. The modeling was done for 27 Hz (the frequency of the blue whale tonal segment) and 22 Hz (the middle frequency of the fin whale call).

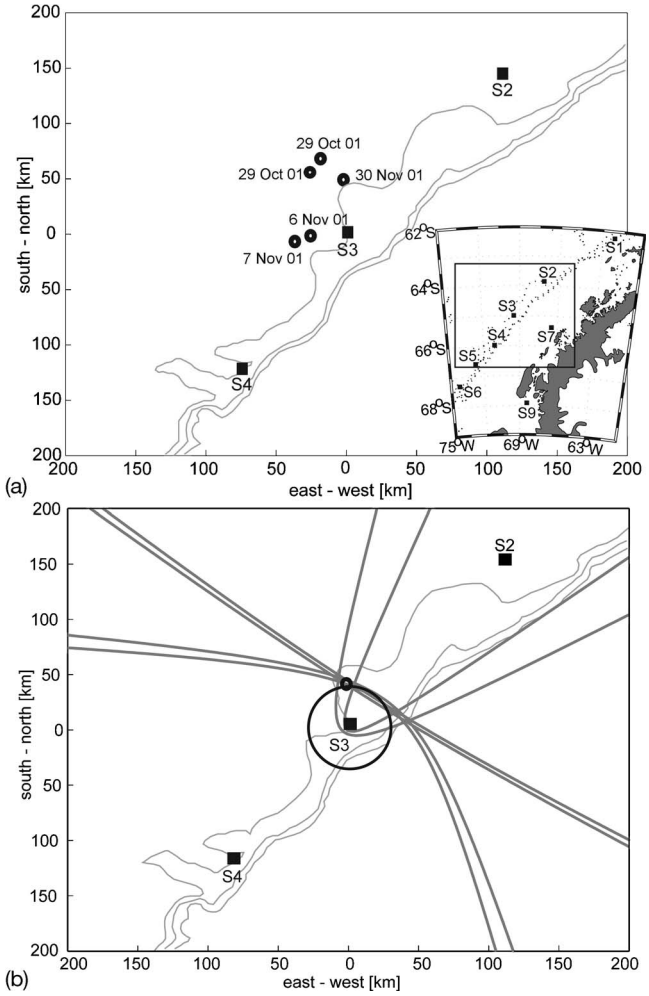


FIG. 2. (a) Locations of calling whales (circles) and dates when they were recorded. Squares show ARP locations and gray lines are 1000-, 2000-, and 3000-m bathymetry contours. Inset shows a larger area of the Western Antarctic Peninsula (dark gray) where the ARPs numbered S1 to S9 were deployed, with area of localizations indicated with a box. (b) Comparison of multipath and hyperbolic localization results for one of the calls recorded on 30 November 2001.

III. RESULTS

The range to calling blue and fin whales and the source levels of their calls were calculated using multiple calls. Detections useful for localization and range determination were limited to the austral spring for blue whales and the early fall for fin whales, because those were the times during which there was less calling (Širović *et al.*, 2004), making it possible to find periods without overlapping call sequences from multiple whales.

A. Blue whales

At least five blue whales were localized on four different days using 84 individual calls in October and November 2001 (Fig. 2). The longest track (a series of whale locations calculated from a number of sequential calls) lasted 1 h 17 min, while the shortest was 13 min. Owing to the changes in the ARP array geometry, calls from the same blue whale could be heard on instruments at sites 2, 3, and 4 only during one deployment year. It should be noted that the original experiment design intended each instrument to be independent and individual calls recorded on one instrument would not be recorded by other ARPs. The linear array geometry limited the detection area to a relatively tight region. However, due to high intensity of the sounds and good propagation characteristics, blue whale calls could be detected up to the 200-km range. The mean error in the TDOA method was 2.6 s, or the equivalent of 3.8 km. (We do not report percent error because it was different for each instrument used for localization.) Propagation modeling under typical spring conditions showed that sounds produced in shallow water do not propagate easily into deep water. Therefore, all localized animals were assumed to be calling off the shelf, in deep water, from where their calls could be recorded by the ARPs.

The transmission loss coefficient (X), corresponding to linear least-squares fit of call received levels and logarithmic of calculated ranges, was 17.8 dB/m (Fig. 3). This matched closely (within 2 dB $re:1$ m) the results of the modeled transmission loss at two depths (Fig. 4). The empirical value at short ranges (< 80 km) fit the propagation model at 2000-m depth better, while for ranges over 80 km the fit was

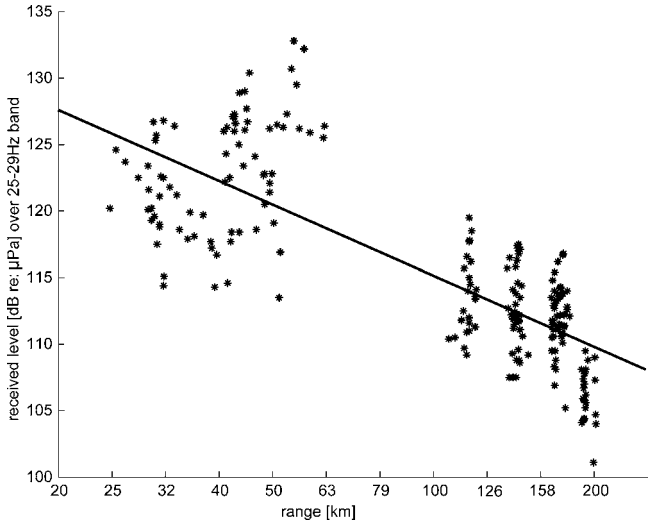


FIG. 3. Plot of blue whale received levels versus log of calculated range (N=252). Black line is the best-fit line through the data; the slope of this line corresponds to the value of the transmission loss coefficient, X, and is 17.8 dB/m. An increase in X leads to an increase in the difference from the theoretical model at higher ranges, while a decrease in X leads to an increase in the difference at lower ranges.

better for the 200-m depth. The difference between propagation models of 200 and 2000 m, however, was generally not larger than 5 dB *re*:1 m.

The average source level of blue whale calls off the Western Antarctic Peninsula was estimated to be 189±3 dB *re*:1 μPa at 1 m over the 25–29-Hz band [Fig. 5(a)]. The average standard deviation of each source level calculation was 2.8 dB *re*:1 μPa at 1 m, which estimated the measurement error of our system. If the difference in the range to a calling whale between two consecutive calls was greater than 10 km, we assumed there were at least two different blue whales calling. We also assumed two calling whales if the intercall interval between the calls was less than 60 s (Širović *et al.*, 2004; Rankin *et al.*, 2005). With those assumptions, we found that the received levels of an individual blue whale during a calling bout on one instrument had a maximum variation up to 6 dB *re*:1 μPa at 1 m.

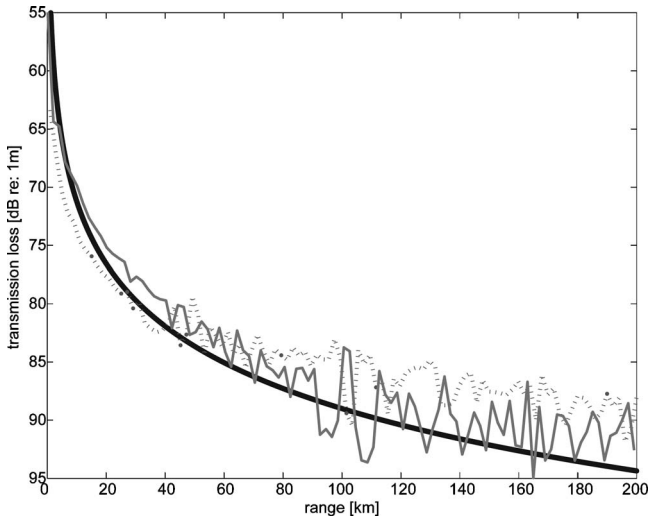


FIG. 4. Results of BELLHOP incoherent transmission loss calculations for Antarctic Peninsula spring conditions at 27 Hz. Solid gray line is the transmission loss at 200-m depth, and the dashed line is the loss at 2000-m depth. Black line is the empirically determined transmission loss, TL =17.8*log(r).

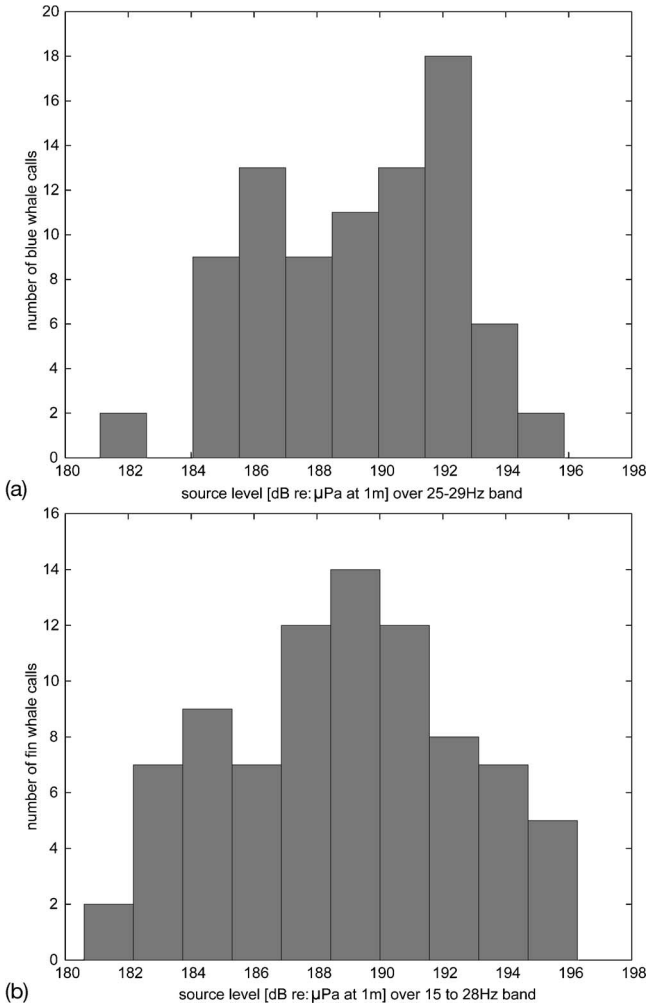


FIG. 5. Distribution of (a) blue whale call source levels, with the mean of 189±3 dB *re*:1 μPa at 1 m over the 25–29-Hz band (N=84) and (b) fin whale call source levels, with the mean of 189±4 dB *re*:1 μPa at 1 m over the 15–28-Hz band (N=83).

There was a significant difference between the results of the hyperbolic localization and multipath arrival methods (df=13, t=−1.17, p=0.262), and the average difference between the calculated ranges to calling blue whales was 1.8±5.6 km (between 3% and 7%), which is of the same order as the error in each method. Since the downswept part of the blue whale call used in these measurements is very similar to the fin whale call, it is reasonable to assume that the method works equally well for both species, and that the range results obtained for the two species using these different methods are comparable.

B. Fin whales

A total of 83 fin whale calls from 12 different days between March and June 2001 were analyzed for range and source levels. The longest period during which ranges to fin whale calls were determined was 21 min. Calls with clear, three or more multipath arrivals, however, generally occurred only every 3–4 min and it was not possible to determine

whether the calls originated from the same animal, so the variation in received levels is not reported. The maximum range to calling fin whales determined using multipath arrivals was 56 km. The average error in the measurement of multipath arrival times was 0.1 s, and the error in range determination resulting from this measurement error was 3.4 km (6%). There were no differences between the transmission loss at 27 and 22 Hz at different depths and different seasons, so we used the transmission loss coefficient calculated from the blue whale data ($X=17.8$ dB/m) for the estimation of transmission loss for fin whale calls. The average source level of fin whale calls was estimated to be 189 ± 4 dB *re*: 1 μPa at 1 m, over the 15–28-Hz band [Fig. 5(b)].

IV. DISCUSSION

Blue and fin whale call source levels reported here are among the highest intensity calls reported for these two species. The maximum source levels reported for previous studies at other locations (e.g., Cummings and Thompson, 1971; Thode *et al.*, 2000; Watkins, 1981) are close to the mean levels reported here. Given the low population densities of these two species in the Southern Ocean (Branch and Butterworth, 2001), these high source level calls would be beneficial for long-range propagation and successful communication with conspecifics. Our empirical estimate of transmission loss was comparable to the theoretical transmission loss calculations across the range, with better correspondence at ranges below 60 km, where all the calls with the range determined from multipath arrivals occurred. Even though there was some discrepancy between the empirical and theoretical transmission losses at longer ranges, the average difference between blue whale call source levels obtained from calculations at three different ranges was low (2.8 dB *re*: 1 μPa at 1 m) and without a consistent pattern between near and far calls, indicating that our transmission loss method did not create a bias. So, the observed difference between this study and previous ones is not likely caused by biased transmission loss estimation.

From the source levels reported here and the calculated transmission loss coefficient, it is possible to estimate theoretical maximum range over which these calls could be detected by conspecifics. The average noise levels in this region are 75 dB *re*: 1 μPa^2/Hz at 220 Hz (McDonald *et al.*, 2005), and at lower frequencies where blue and fin whale calls occur (15–30 Hz), they were up to 5 dB *re*: 1 μPa^2/Hz higher during periods when call ranges and source levels were calculated for this study. Even though there are no reports on threshold signal-to-noise (S/N) ratios for blue and fin whales, critical ratio functions are similar among vertebrates (Richardson *et al.*, 1995), so if we assume zero threshold S/N ratio for the calls to be intelligible by conspecifics (Miller *et al.*, 1951; Scharf, 1970), these whales could be heard out to a distance of about 1300 km. This theoretical range, however, is shortened by the real-life constraints imposed on call propagation by the changes in the physical properties, such as the sound-speed profile, at the fronts of the Antarctic Circumpolar Current.

The detection of a call by a conspecific also depends on the product of call duration and bandwidth. Long calls with narrow bandwidth and short, broadband calls can have similar detectability. Blue whale calls have the highest intensity in a very narrow, 1-Hz band, but they last several seconds (8–18 s). Fin whale calls, on the other hand, are short (<1 s) and cover 5–10 Hz of effective bandwidth. These different temporal and frequency characteristics make blue whale calls about 2 times easier to detect than fin whale calls. Production of repetitive calls further increases the probability they will be detected by a conspecific (Payne and Webb, 1971) and both species regularly repeat calls.

The range over which calls were detected in this study are comparable to earlier results. Stafford *et al.* (1998) reported detecting blue whales in the North Pacific over ranges of 400 to 600 km and Clark (1995) detected them in the Atlantic Ocean at ranges of up to 1600 km. Cummings and Thompson (1971) detected fin whales to a distance of 100 mi. The sensors Clark (1995) and Stafford *et al.* (1998) used, however, were placed in the sound channel, and they summed multiple beams to enhance the S/N ratios. Our instruments were in approximately 3000 m of water, in the polar region where the sound channel comes close to the surface (Jensen *et al.*, 2000), so the propagation was less than optimal and the signal was not enhanced by processing.

The accuracy in the measured arrival times of both methods was limited by the ability of the human analyst to pick the arrival times, and the difference between the methods was comparable to levels of measurement error. Multipathing, which was the result of the complex propagation environment, made it impossible to automatize call cross correlation, as in Tiemann *et al.* (2004), for example. This produced errors of several kilometers in the range estimation, so it was impossible to determine blue and fin whale swim speeds. But, as the calls were detected over long ranges, the relative percentage errors are comparable to other localization studies (e.g., Clark and Ellison, 2000).

Variation in source levels of 5 dB has been reported previously for fin whales (Watkins, 1981), and we found a variation in individual blue whale received levels of 6 dB *re*: 1 μPa. By using received levels we eliminated the 2.8-dB error introduced by range determination. We assumed that this variation is a result of a single calling animal, but it is possible there were multiple animals calling close to each other, each at a different source level. Usually, however, the calls were repeated at very regular intervals, which indicate that a single whale was likely calling. Even though many calls showed multipath arrivals, the full range could not be accounted for by the changes in the multipath, because the movement of the whale between successive calls (always less than 2 min) would not be large enough to cause large changes in the propagation characteristics over these distances. Likewise, the variation is not likely caused by variations in the calling depth since blue whales appear to produce calls at consistent depth (Oleson *et al.*, 2007). Therefore, it appears that the total variation in the source levels of the analyzed population sample is comparable to the variation in the calls of individual whales.

Although we found there was likely some variation in the call source levels within an individual blue whale, we could not establish if there was a seasonal difference in call levels. Our ability to localize and range on animals during very short seasonal periods was not caused by the seasonal changes in the propagation characteristics, but by the number of calling animals. While hundreds of thousands of calls were present in the data set (Širović *et al.*, 2004), calls could be used for the analyses only when calls were not too abundant, as it was necessary to distinguish between individual calls. Therefore, the methods used here would not be useful in areas with a large number of calling animals, or times with overlapping calls.

Another correlation worth investigating is possible change in the source levels during periods of high acoustic noise. Fin whales present in the northern region of the array create a "noise band" in the 15–28-Hz band during peak presence (Širović *et al.*, 2004). If blue whales, for example, use the calls for communication with conspecifics, they would have to overcome that noise by increasing their source levels, or changing their call frequency. The blue and fin whale calls measured in this study, however, occurred at times when there was no fin whale "noise band." As blue whale calls in the Southern Ocean have a consistent frequency (Širović *et al.*, 2004; Rankin *et al.*, 2005), it would be interesting to determine if blue and fin whale call source levels exhibit a Lombard effect (higher source levels) during periods of higher noise, which was not possible in this study.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation Office of Polar Programs Grants OPP 99-10007 and OPP 05-23349 as part of the Southern Ocean GLOBEC program, with program guidance by Polly Penhale, Roberta Martinelli, and Marie Bundy. BELLHOP acoustic modeling software was developed by M. Porter and is available from the Ocean Acoustic Library. The authors would like to thank the Masters and crew of the ARSV LAURENCE M GOULD during LMG01-03, LMG02-01A, and LMG03-02, as well as the staff at Raytheon Polar Services who provided logistical assistance. The manuscript was improved by comments from M. A. McDonald, J. Barlow, C. Berchok, and two anonymous reviewers. This work represents a portion of A.Š.'s dissertation.

Best, P. B. (**1993**). "Increase rates in severely depleted stocks of baleen whales," ICES J. Mar. Sci. **50**, 169–186.

Branch, T. A., and Butterworth, D. S. (**2001**). "Estimates of abundance south of 60°S for cetacean species sighted frequently on the 1978/79 to 1997/98 IWC/IDCR-SOWER sighting surveys," J. Cetacean Res. Manage. **3**, 251–270.

Branch, T. A., Matsuoka, K., and Miyashita, T. (**2004**). "Evidence for increases in Antarctic blue whales based on Bayesian modeling," Marine Mammal Sci. **20**, 726–754.

Cato, D. H. (**1998**). "Simple methods of estimating source levels and location of marine animal sounds," J. Acoust. Soc. Am. **104**, 1667–1678.

Charif, R. A., Mellinger, D. K., Dunsmore, K. J., Fristrup, K. M., and Clark, C. W. (**2002**). "Estimated source levels of fin whale (*Balaenoptera physalus*) vocalizations: Adjustments for surface interference," Marine Mammal Sci. **18**, 81–98.

Clapham, P. J., and Baker, C. S. (**2001**). "How many whales were killed in the Southern Hemisphere in the 20th century?," Paper SC/53/O14 presented to IWC Scientific Committee, July 2001 (unpublished). 3 pp. Available from secretariat@iwoffice.org.

Clark, C. W. (**1995**). "Matters arising out of the discussion of blue whales," Rep. Int. Whal. Comm. **45**, 210–212.

Clark, C. W., and Ellison, W. T. (**2000**). "Calibration and comparison of the acoustic location methods used during the spring migration of the bowhead whale, *Balaena mysticetus*, off Pt. Barrow, Alaska, 1984–1993," J. Acoust. Soc. Am. **107**, 3509–3517.

Croll, D. A., Clark, C. W., Acevedo, A., Tershy, B. R., Flores, S., Gedamke, J., and Urban, J. (**2002**). "Only male fin whales sing loud songs," Nature (London) **417**, 809.

Cummings, W. C., and Thompson, P. O. (**1971**). "Underwater sounds from the blue whale, *Balaenoptera musculus*," J. Acoust. Soc. Am. **50**, 1193–1198.

Edds, P. L. (**1982**). "Vocalizations of the blue whale, *Balaenoptera musculus*, in the St. Lawrence River," J. Mammal. **63**, 345–347.

Edds, P. L. (**1988**). "Characteristics of finback *Balaenoptera physalus* vocalizations in the St. Lawrence estuary," Bioacoustics **1**, 131–149.

Frazer, L. N., and Pecholcs, P. I. (**1990**). "Single-hydrophone localization," J. Acoust. Soc. Am. **88**, 995–1002.

Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H. (**2000**). *Computational Ocean Acoustics* (Springer, New York).

Ljungblad, D., Clark, C. W., and Shimada, H. (**1998**). "A comparison of sounds attributed to pygmy blue whales (*Balaenoptera musculus brevicauda*) recorded south of the Madagascar Plateau and those attributed to 'true' blue whales (*Balaenoptera musculus*) recorded off Antarctica," Rep. Int. Whal. Comm. **49**, 439–442.

McDonald, M. A. (**2005**). "Calibration of Acoustic Recording Packages (ARPs) at Pt. Loma Transducer Evaluation Center (TRANSDEC)." SIO Tech. Report (Marine Physical Laboratory, La Jolla, CA), 16 pp.

McDonald, M. A., and Fox, C. G. (**1999**). "Passive acoustic methods applied to fin whale population density estimation," J. Acoust. Soc. Am. **105**, 2643–2651.

McDonald, M. A., Calambokidis, J., Teranishi, A. M., and Hildebrand, J. A. (**2001**). "The acoustic calls of blue whales off California with gender data," J. Acoust. Soc. Am. **109**, 1728–1735.

McDonald, M. A., Hildebrand, J. A., and Webb, S. C. (**1995**). "Blue and fin whales observed on a seafloor array in the Northeast Pacific," J. Acoust. Soc. Am. **98**, 1–10.

McDonald, M. A., Hildebrand, J. A., Wiggins, S. M., Thiele, D., Glasgow, D., and Moore, S. E. (**2005**). "Sei whale sounds recorded in the Antarctic," J. Acoust. Soc. Am. **118**, 3941–3945.

McDonald, M. A., Mesnick, S. L., and Hildebrand, J. A. (**2006**). "Biogeographic characterization of blue whale song worldwide: Using song to identify populations," J. Cetacean Res. Manage. **8**, 55–65.

Miller, G. A., Heise, G. A., and Lichten, W. (**1951**). "The intelligibility of speech as a function of the context of the test materials," J. Exp. Psychol. **41**, 329–335.

Northrop, J., Cummings, W. C., and Thompson, P. O. (**1968**). "20-Hz signals observed in the Central Pacific," J. Acoust. Soc. Am. **43**, 383–384.

Oleson, E. M., Calambokidis, J., Burgess, W. C., McDonald, M. A., LeDuc, C. A., and Hildebrand, J. A. (**2007**). "Behavioral context of Northeast Pacific blue whale call production," Mar. Ecol.: Prog. Ser. **330**, 269–284.

Payne, R., and Webb, D. (**1971**). "Orientation by means of long range acoustic signaling in baleen whales," Ann. N.Y. Acad. Sci. **188**, 110–141.

Pickard, G. L., and Emery, W. J. (**1990**). *Descriptive Physical Oceanography* (Butterworth Heinmann, Oxford).

Rankin, S., Ljungblad, D., Clark, C. W., and Kato, H. (**2005**). "Vocalizations of Antarctic blue whales, *Balaenoptera musculus intermedia*, recorded during the 2001–2002 and 2002–2003 IWC-SOWER circumpolar cruises, Area V, Antarctica," J. Cetacean Res. Manage. **7**, 13–20.

Richardson, W. J., Greene Jr., C. R., Malme, C. I., and Thomson, D. H., editors (**1995**). *Marine Mammals and Noise* (Academic, San Diego).

Scharf, B. (**1970**). "Critical bands," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. 1, pp. 157–202.

Širović, A., Hildebrand, J. A., Wiggins, S. M., McDonald, M. A., Moore, S. E., and Thiele, D. (**2004**). "Seasonality of blue and fin whale calls and the influence of sea ice in the Western Antarctic Peninsula," Deep-Sea Res., Part II **51**, 2327–2344.

Širović, A., Hildebrand, J. A., and Thiele, D. (**2006**). "Baleen whales in the Scotia Sea in January and February 2003," J. Cetacean Res. Manage. **8**, 161–171.

Spiesberger, J. L. (**2001**). "Hyperbolic location errors due to insufficient

numbers of receivers," J. Acoust. Soc. Am. **109**, 3076–3079.
Stafford, K. M., Fox, C. G., and Clark, D. S. (**1998**). "Long-range acoustic detection and localization of blue whale calls in the northeast Pacific Ocean," J. Acoust. Soc. Am. **104**, 3616–3625.
Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**1999**). "Low-frequency whale sounds recorded on hydrophones moored in the eastern tropical Pacific," J. Acoust. Soc. Am. **106**, 3687–3698.
Thode, A. M., D'Spain, G. L., and Kuperman, W. A. (**2000**). "Matched-field processing, geoacoustic inversion, and source signature recovery of blue whale vocalizations," J. Acoust. Soc. Am. **107**, 1286–1300.
Tiemann, C. O., Porter, M. B., and Frazer, L. N. (**2004**). "Localization of marine mammals near Hawaii using an acoustic propagation model," J. Acoust. Soc. Am. **115**, 2834–2843.
Urick, R. J. (**1983**). *Principles of Underwater Sound* (Peninsula, Los Altos, CA).
Watkins, W. A., and Schevill, W. E. (**1972**). "Sound source location by arrival-times on a non-rigid three-dimensional hydrophone array," Deep-Sea Res. **19**, 691–706.
Watkins, W. A., (**1981**). "Activities and underwater sounds of fin whales," Sci. Rep. Whales Res. Inst. **33**, 83–117.
Watkins, W. A., Tyack, P., Moore, K. E., and Bird, J. E. (**1987**). "The 20-Hz signal of finback whales (*Balaenoptera physalus*)," J. Acoust. Soc. Am. **82**, 1901–1912.
Wiggins, S. (**2003**). "Autonomous Acoustic Recording Packages (ARPs) for long-term monitoring of whale sounds," Mar. Technol. Soc. J. **37**(2), 13–22.

Variation in chick-a-dee calls of tufted titmice, *Baeolophus bicolor*: Note type and individual distinctiveness

Jessica L. Owens and Todd M. Freeberg[a)]
Department of Psychology, Austin Peay Building, University of Tennessee, Knoxville, Tennessee 37996

(Received 10 November 2006; revised 16 May 2007; accepted 22 May 2007)

The chick-a-dee call of chickadee species (genus *Poecile*) has been the focus of much research. A great deal is known about the structural complexity and the meaning of variation in notes making up calls in these species. However, little is known about the likely homologous "chick-a-dee" call of the closely related tufted titmouse, *Baeolophus bicolor*. Tufted titmice are a prime candidate for comparative analyses of the call, because their vocal and social systems share many characteristics with those of chickadees. To address the paucity of data on the structure of chick-a-dee calls of tufted titmice, we recorded birds in field and aviary settings. Four main note types were identified in the call: ***Z***, ***A***, $\boldsymbol{D_h}$, and ***D*** notes. Several acoustic parameters of each note type were measured, and statistical analyses revealed that the note types are acoustically distinct from one another. Furthermore, note types vary in the extent of individual distinctiveness reflected in their acoustic parameters. This first step towards understanding the chick-a-dee call of tufted titmice indicates that the call is comparable in structure and complexity to the calls of chickadees. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2749459]

PACS number(s): 43.80.Ka, 43.80.Lb, 43.80.Ev [MCH] Pages: 1216–1226

I. INTRODUCTION

The "chick-a-dee" call of the genus *Poecile*, the North American chickadees, is a structurally complex vocal signal. Calls are composed of distinct note types, and note composition is guided by rules of note ordering (Hailman *et al.*, 1985). By using these rules of note ordering an extremely large number of distinct calls can be produced by varying the number of note types within a call. A great deal has been discovered about variation in note composition and acoustic structure and how these relate to possible messages and meanings of chick-a-dee calls (Hailman, 1989; Hailman and Ficken, 1996; Lucas and Freeberg, 2007; Sturdy *et al.*, 2007; Templeton *et al.*, 2005). For example, variation in acoustic structure of notes or in note composition in chick-a-dee calls, or both, may potentially encode information about individuals (Bloomfield *et al.*, 2004, 2005; Charrier *et al.*, 2004), flock membership (Mammen and Nowicki, 1981; Nowicki, 1983), local population membership (Freeberg *et al.*, 2003), energetic status of the signaler (Lucas *et al.*, 1999), food detection and flight behavior of the signaler (Freeberg and Lucas, 2002; Smith, 1972), detection of predator by, and distance of predator to, the signaler (Baker and Becker, 2002; Ficken *et al.*, 1994), and level of threat related to predator size (Templeton *et al.*, 2005).

These efforts at understanding the chick-a-dee call have been directed towards chickadee species (and some *Parus* tit species, e.g., willow tit, *P. montanus*, Haftorn, 1993, 2000; black-lored tit, *P. xanthogenys*, Hailman, 1994). As described in Hailman and Ficken (1996) and Freeberg *et al.* (2007), however, very little is known about the likely homologous chick-a-dee call of the closely related titmice (genus *Baeolophus*). The aim of this study is to begin to understand the structure of the chick-a-dee call of tufted titmice, *B. bicolor*, and to assess some of the potential information that may reside in the notes of the signal.

Recent work by Bloomfield and Sturdy and colleagues (mountain chickadees, *Poecile gambeli*, Bloomfield *et al.*, 2004; Carolina chickadees, *P. carolinensis*, Bloomfield *et al.*, 2005; black-capped chickadees, *P. atricapillus*, Charrier *et al.*, 2004), has used a consistent approach to analyzing call structure across these different species. This approach involves assessing multiple acoustic parameters of the different note types that construct calls, to determine the potential for those acoustic characteristics to be used by the birds in note type and individual perceptual discriminations. Another study of Carolina chickadees found relationships between acoustic structure of notes in a call and the local population of the signaler, as well as the note composition of the rest of the call (Freeberg *et al.*, 2003). In the present study, we have taken approaches used by these earlier studies to assess call and note variation in tufted titmice. We aimed to address the paucity of data on the chick-a-dee call of tufted titmice by classifying their note types into natural categories using acoustic features, by determining which spectral and temporal features of each note type may be used in note type and individual discrimination, and by beginning to describe the effects of note composition of calls and sex of signaler on note structure. The current study is our first step to a more comprehensive analysis of structural and functional variation of this call system in titmouse species.

II. METHODS

A. Subjects

Chick-a-dee calls used in the current study were recorded from 16 individually color-banded tufted titmice that

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

 0001-4966/2007/122(2)/1216/11/$23.00

served as subjects in two separate aviary studies or were part of naturalistic field studies from December of 2005 through April of 2006. For aviary studies, 14 titmice were housed in social groups with other titmice and some also with Carolina chickadees. All birds in the aviaries could see and hear wild birds in the wooded areas around the aviaries. High quality vocalizations were also recorded from two color banded free-range birds. One aviary individual, after being released from the aviary study, was recorded at the field site mentioned above. One call from this bird's field recordings is included in this analysis.

We captured birds recorded in the aviaries at several different sites at the University of Tennessee Forest Resources, Research, and Education Center (UTFRREC), Ijams Nature Center, and a residential site, all within Knox and Anderson counties, TN. After capture with potter (treadle) traps set on platform feeding stations or with mist nets set near the feeding stations, all titmice were individually color marked with unique color combination leg rings and then placed in one of three-outdoor aviaries ($6\times9\times3.5$ m) located at UTFRREC. Although ages of birds were not known, no birds used in this study produced the highly variable and plastic vocal signals characteristic of young birds first learning their vocalizations. We measured wing chord length of birds as a criterion for estimating sex. We were able to classify sex of most of the tufted titmice in our sample based upon wing chord measurements—we categorized birds with wing chords of $\leq$77 mm as females and birds with wing chords of $\geq$80 mm as males, with birds falling between 77 and 80 mm being classified as unknown (based upon Thirakhupt, 1985). In our set of 16 birds, six were classified as males, five as females, and six as unknown (five of the unknown birds had wing chord measures that fell in the range of overlap between females and males, and one was never measured). This wing chord criterion for sexing birds was supported with capture data of 11 titmice (none were subjects in this study) in May of 2007. At this time of year, titmice are nesting and incubating young (Grubb and Pravosudov, 1994), with females exhibiting complete brood patches and males exhibiting fairly swollen cloacal protuberances with minimal brood patch presence (Pyle, 1997). Using brood patches and cloacal protuberances as our means of sexing these 11 birds, we found that two of three females had wing chords $\leq$77 mm (with the third female having a wing chord of 78.5 mm, which would have resulted in her being classified as "unknown" in our study), and eight of eight males had wing chords $\geq$80 mm.

In the aviaries, diet was composed of *ad libitum* black-oil sunflower seed and safflower seed, mixed songbird seed, suet, and crushed oyster shell. In addition, birds were provided with Bronx Zoo diet for omnivorous birds, chopped fresh vegetables or fruit, and fresh vitamin treated water each day. We also added to the aviary feeding bowls roughly 2–3 mealworms per bird each day. We acclimated the titmice to human presence for at least the first two weeks they were in captivity. During the acclimation period JO or TF or both would enter the aviary and sit where they would during a recording session and/or walk around the aviary. Recording began after this acclimation period. After recording was completed birds were recaptured in the aviaries using potter traps and transported back to their sites of capture and released.

B. Recording of chick-a-dee calls

We recorded chick-a-dee calls using Sennheiser ME-64 and ME-66 microphones and either Fostex FR-2 digital field memory recorders (with a sampling rate of 22 050 and a 16 bit resolution) or a Marantz-PMD 22 portable cassette recorder on Maxell-XLII tape. In the aviaries the microphone was mounted on a stand approximately 3 m off the ground and angled toward nearby perches. The observer sat as motionless as possible at a far end of the aviary and documented the identity of the bird after it called directly on the sound file. For field recordings the microphone was set up roughly 1 m from the platform feeding station and was aimed up at the feeding station. The feeding stations were stocked with a 1:1 ratio of black-oil sunflower seed and safflower seed to attract birds. The observer sat as motionless as possible roughly 10 m away and partially concealed by vegetation.

C. Call selection and note categorization

Chick-a-dee calls included in this analysis were chosen based first on recording quality: we chose calls that were produced within 45° of the axis of the microphone and within 3 m of the microphone (as stated by the observer on the recording file), and that had minimal wind and background noise. Once we had compiled a group of high quality calls from a bird, we semirandomly picked ten calls from the bird to go into subsequent analyses. Before random picking of calls, we first attempted to increase the diversity of note types included in the call sample for each individual. If a note type was rare for a particular individual, we specifically chose 1 or 2 calls that included that note type from its set of calls. After we had selected 1 or 2 calls for the rare note type of an individual, we used a random number generator to pick the remaining calls needed to reach ten calls per bird. Calls with only one note were omitted from the current study; for our purposes here, a call is defined as having two or more notes.

To classify call notes JO scored all 508 notes from 160 calls and assigned each to either a ***Z***, ***A***, $\boldsymbol{D_h}$ or ***D*** note-type category (described below; Fig. 1). These categories are similar to those used in the chick-a-dee call descriptions of chickadees. TF independently scored notes for all 160 calls, and agreement was high (Cohen's Kappas for inter-rater reliability on note classifications: $\boldsymbol{Z}=0.94$, $\boldsymbol{A}=0.95$, $\boldsymbol{D_h}=1.00$, $\boldsymbol{D}=1.00$; JO's note classifications were used for analyses). ***Z*** and ***A*** are whistled note types that begin at a high frequency and differ in duration and extent of frequency change. ***A*** notes are shorter in duration than ***Z*** notes and frequently have a pronounced drop in ending frequency, whereas ***Z*** notes are longer and do not typically drop below 4 kHz. ***D*** notes are harsh sounding note types that appear as stacked harmonics in spectrogram form (similar to the ***D*** note types in several chickadee species). $\boldsymbol{D_h}$ notes are a hybrid category. These notes begin at an upper frequency similar to an ***A***.

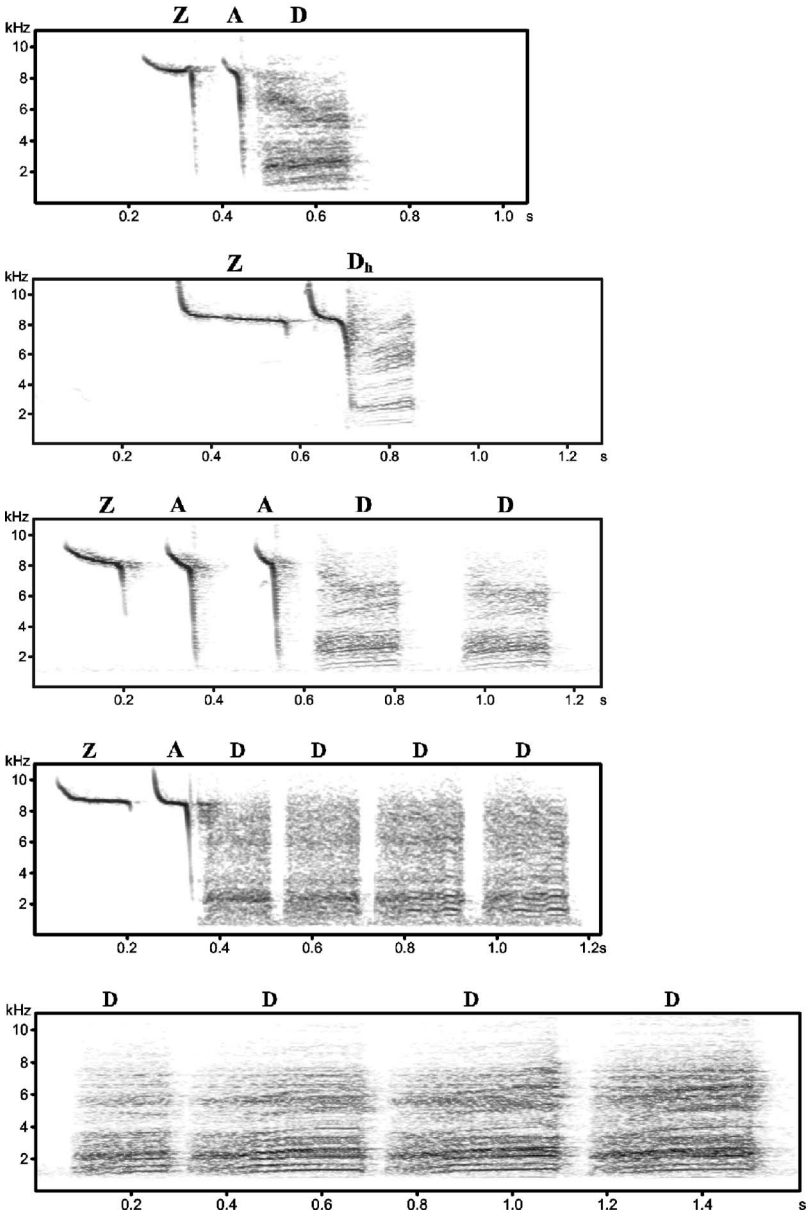


FIG. 1. Sound spectrograms of examples of chick-a-dee calls from tufted titmice. *Y* axis=frequency (kHz); *X* axis=time (s). Spectrograms were generated in Avisoft SASLab Pro, with a FFT length of 512, frame 75%, and Blackman window. Note type classifications (***Z***, ***A***, $\boldsymbol{D_h}$, and ***D***) for each note are indicated above the note in each spectrogram.

However, the ending frequency drop does not terminate, but levels off and is joined by the harmonic-like structure of a ***D*** note.

D. Analyses of the chick-a-dee calls

Calls recorded with the Marantz PMD-22 cassette recorder were digitized using Cool Edit Pro (version 2) run on a Windows XP platform, with a sampling rate of 22 050 at 16 bit resolution. Once all recording files were digitized or uploaded to the computer, Cool Edit Pro (version 2) was used to assess the large (5–15 min) sound files for high quality calls. All calls were viewed in the Blackman-Harris window with a resolution of 256 bands. Individual calls were separated in Cool Edit and acoustical analyses were completed using SASLab Pro sound analysis and synthesis laboratory software (version 4.2; Avisoft, hereafter, SASLab). All of our parameter measurements of notes of calls were made in a 512 pt fast Fourier transform (FFT) Blackman spectrogram window using the automatic parameter settings feature within SASLab. We used a volume range of 25–45% in the Normalize window of the "Change Volume" menu in SASLab. We used this volume range to adjust the volume of individual files to make note types of calls recorded at different distances and orientations from the microphone comparable to one another for automatic parameter measurements by SASLab. To decrease low frequency background noise in the sound files we applied a high pass Butterworth filter set at 0.8 kHz, and occasionally at 1.0 kHz, depending on the nature of background noise in the individual call file. Within the automatic parameter settings we selected for automatic two threshold measurements, and therefore could manually change the "threshold" and the "start/end threshold" options. To ensure that all note types within a call were individually measured by SASLab, we used a threshold range of −34 to −40 dB, and a start/end threshold range of −10 to −20 dB, depending on the specific quality of each call. We used the same 17 automatic parameter measures for each note type. The automatic parameter measures included

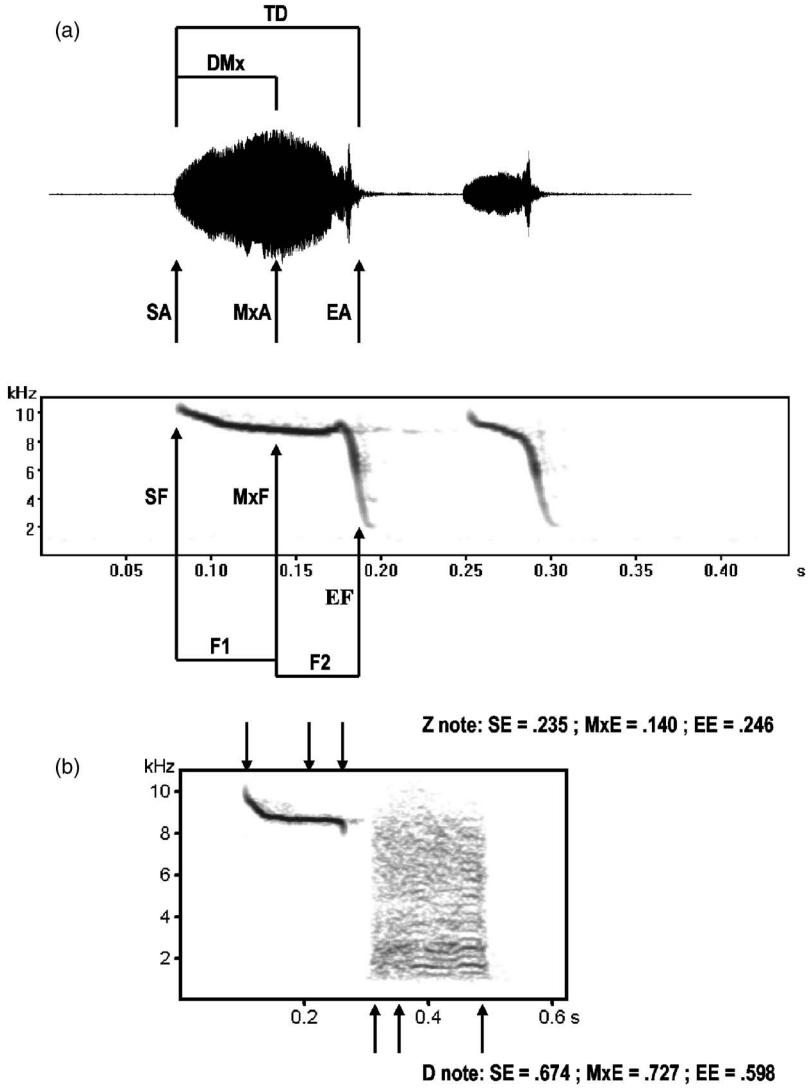


FIG. 2. Illustration of some of the parameters measured in notes of chick-a-dee calls in this study, (a) Wave form and sound spectrogram of a ***ZA*** call, illustrating the temporal and some of the amplitude and frequency measures. (b) Sound spectrogram of a ***ZD*** call, illustrating some of the entropy measures. TD=total note duration; DMx=distance to maximum amplitude; SA =amplitude at start of note; MxA=amplitude at point of maximum amplitude; EA=amplitude at end of note; SF=frequency at start of note; MxF=frequency at point of maximum amplitude; EF=frequency at end of note; F1=frequency difference between SF and MxF; F2=frequency difference between MxF and EF; SE =entropy at start of note; MxE=entropy at point of maximum amplitude; EE=entropy at end of note. Not illustrated are the mean measures of frequency, amplitude, and entropy for the entire note, and RMS of the entire note.

two temporal, six frequency, five amplitude, and four entropy parameters (Fig. 2; several of our measures are adapted from those described in Bloomfield *et al.*, 2004, 2005; Charrier *et al.*, 2004; Freeberg *et al.*, 2003).

1. Temporal parameters

We assessed two different temporal parameters: *total note duration* (TD) which measures the duration (s) of each note from start to end, and *distance to max* (DMx) which measures the duration (s) from the start of a note to the point of maximum amplitude in the note.

2. Frequency parameters

We assessed six frequency-based parameters: *start frequency* (SF), *end frequency* (EF), *max frequency* (MxF), *mean frequency* (MnF), *frequency change 1* (F1), and *frequency change 2* (F2), each measured in Hz. SF is a measure of the frequency at the start of the note, and EF is the same measure at the end of the note, based upon the threshold settings described above. MxF is the frequency in the note at the point with maximum amplitude. MnF is a parameter derived from the averaged spectrum of the entire note. F1 measures change in frequency from SF to MxF (calculated as SF–MxF), and F2 measures change in frequency from MxF to EF (calculated as MxF–EF).

3. Amplitude parameters

The five amplitude parameters are *start amplitude* (SA), *end amplitude* (EA), *max amplitude* (MxA), *mean amplitude* (MnA), each measured in dB, and *root mean square* (RMS), measured in 1 V units. SA is a measure of the amplitude at the start of the note, and EA is the same measure at the end of the note, based upon the threshold settings described above. MxA measures the maximum amplitude of the note. MnA is a derived parameter that averages the amplitude across the spectrum of the entire note; rms is a wave form parameter derived from the entire note.

4. Entropy parameters

Entropy measures provide an assessment of structural disorder or "noisiness" within a note. For example, a relatively pure tone note would have entropy fairly close to 0, whereas a harsh, noisy note or syllable spanning a wide range of frequencies would have an entropy closer to 1 [see Fig. 2(b) for typical entropy differences seen in comparisons of ***Z*** notes and ***D*** notes]. We measured *start entropy* (SE), the

entropy at the beginning of the note, *end entropy* (EE), the entropy at the end of the note, based upon the threshold settings described above. *Max entropy* (MxE) is the entropy at the point in the note of maximum amplitude, and *mean entropy* (MnE) is the entropy averaged across the entire note.

E. Statistical analyses

The basic goal of our study was to describe the structure of note types of the chick-a-dee call of the tufted titmouse. Our high inter-rater agreement in categorizing note types into ***Z***, ***A***, $\boldsymbol{D_h}$ and ***D*** notes (discussed above), suggested these note categories may be natural categories. We next aimed to determine which acoustic features of each note type may be useful in note type and individual discrimination, and to begin to describe the effects of a call's note composition on these acoustic features. We first sought to establish which acoustic features of these note types would have the potential to allow receivers to discriminate note types, and possibly individuals, from one another. As a converging measure of this approach, and since many of our measured parameters were likely highly correlated with one another, we reduced the 17 parameters using factor analysis, and then used the factors that emerged to test for effects of note type, individual, and sex and note composition of call on factor scores.

1. Potential for note-type coding (PNTC): Can acoustic features help discriminate between note types?

We determined which spectral and temporal features of notes of the chick-a-dee call may be used to discriminate between note types by examining the potential for note type coding (PNTC, Bloomfield *et al.*, 2004, 2005; Charrier *et al.*, 2004). PNTC is a variance measure. Specifically, it is the ratio of the coefficient of variation (Sokal and Rohlf, 1995) between note types (CV_b) and the mean of the coefficients of variation within note types (CV_w). This measure indicates which spectral and temporal features are less variable within note types than between note types, and therefore which of these parameters can potentially be used in note type discrimination. A parameter with a PNTC value greater than 1 (i.e., $CV_b > CV_w$) means it may be useful in note-type discrimination (Bloomfield *et al.*, 2004, 2005; Charrier *et al.*, 2004). If a PNTC value for an acoustic parameter is greater than two, it is considered to be a parameter that may be highly useful in note-type discrimination (Robisson *et al.*, 1993). We used the small samples formula to determine PNTC:

$$\mathrm{PNTC} = (\mathrm{CV_b} * 100)/[\mathrm{CV_W} * 100 * (1 + 1/4n)].$$

CV_b is the SD for the acoustic parameter across all the note types, divided by the mean for that acoustic parameter across all note types. CV_w, calculated for each note type, is the SD for the acoustic parameter of the note type, divided by the mean for that acoustic parameter for that note type. The $(1+1/4n)$ at the end of the equation is the correction for small samples (e.g., Sokal and Rohlf, 1995). The "*n*" denotes the number of cases analyzed for each note type. The PNTC analysis was applied to all 508 notes.

2. Potential for individual coding (PIC): Can acoustic features help discriminate between individuals?

We tested whether note-type acoustic characteristics may be useful in individual discrimination (PIC, Bloomfield *et al.*, 2004, 2005; Charrier *et al.*, 2001, 2002, 2004). For individual discrimination to be possible, acoustic parameters must have lower intra-individual variation than inter-individual variation. This means that for an individual titmouse to be distinguishable via one of its note types, it would have to produce reliably distinctive features in that note type that are less variable within its repertoire than in the total set of that note type produced in calls of members of the entire group of birds. To test for this possibility we examined, for each note type, which of the 17 parameters were less variable within an individual than across individuals, to determine which parameters could potentially be useful for each note type in discriminating across individuals. For each note type separately, we examined the PIC using the small samples formula as described above:

$$\mathrm{PIC} = (\mathrm{CV_b} * 100)/[CV_w * 100 * (1 + 1/4n)].$$

CV_b is the SD for the acoustic parameter across all the individuals, divided by the mean for that acoustic parameter across all individuals. CV_w, calculated for each individual, is the SD for the acoustic parameter, divided by the mean for that acoustic parameter. The $(1+1/4n)$ at the end of the equation is the correction for small samples. The "*n*" denotes the number of cases analyzed for each individual. For a parameter to be useful in individual discrimination the PIC value must be greater than 1, which would indicate that the parameter's variation among individuals is greater than variation within an individual bird (Bloomfield *et al.*, 2004, 2005; Charrier *et al.*, 2004). A PIC value greater than 2 is considered to be highly useful at permitting individual discrimination (Robisson *et al.*, 1993). The PIC analysis was run on all individuals by note type. ***Z*** and ***A*** note types were produced by all 16 individuals within the sample. The $\boldsymbol{D_h}$ note type was only produced by three individuals. The ***D*** note type was produced by all birds, but one was excluded from the PIC analysis because it produced only one ***D*** note.

3. Factor analysis: Are note types and individuals distinct?

As a converging measure of note type and individual distinctiveness in the chick-a-dee call of titmice, we reduced the 17 parameters using factor analysis (SPSS, Statistical Package for the Social Sciences, version 13.0). We used the principal components method with a varimax rotation. Factor scores for each note of each call of each bird were then used in analyses of varience (ANOVAs) to evaluate the effects of note type and individual. When significant effects of note type or individual were found for a particular factor, post hoc tests with corrections for multiple comparisons were used to test for significant contrasts between note types or between individuals. If error variance was equal across groups we used Tukey's Honestly Significant Difference (HSD) post hoc test. If error variance was not equal across groups we used Dunnett's *C* post hoc test.

TABLE I. Prevalence of note types in our sample of chick-a-dee calls of tufted titmice.

Unit	Mean No. ± SD	Maximum No.	Percent of calls with ≥1 of note type
Total notes	3.17±1.63	18	
Z	0.92±0.55	2	80.6
A	0.81±0.60	2	71.3
D_h	0.10±0.30	1	10.0
D	1.34±2.02	18	64.4

4. Effect of note composition and signaler sex on a note's factor scores

To examine the possible effects of note composition or sex of signaler we conducted two tests. Only Factor 1 and Factor 2 scores (see below) were analyzed in this analysis since these two factors account for most of the variation. First, for ***Z***, ***A***, and ***D*** note types separately, we tested whether the factor scores for the first note to occur in a call was different if it were the only note, compared to if it were followed by at least one more of the note type in a string in the call, as in Freeberg *et al.* (2003) for Carolina chickadees. This addressed, for example, whether a ***Z*** note followed by non-***Z*** notes in the call was acoustically different from a ***Z*** note followed by one or more ***Z*** notes (and possibly other note types) in the call. Our second test of an effect of note composition on the first ***Z***, ***A***, and ***D*** note types to occur in calls addressed whether the overall number of notes in a call affected factor scores for those notes. Note numbers were grouped into three classes: two note calls, three note calls, four+ note calls. The $\boldsymbol{D_h}$ note type was excluded from note composition analysis because of small sample size and the fact that it was never repeated within a call in this sample. We also tested for an effect of sex on ***Z***, ***A***, and ***D*** note types, and excluded the $\boldsymbol{D_h}$ note because of the small sample size for that note type.

III. RESULTS

We analyzed a total of 508 notes from 160 calls (ten calls per bird). Our final call sample had 145 ***Z***, 132 ***A***, 16 $\boldsymbol{D_h}$, and 215 ***D*** note types. The average tufted titmouse call in our sample had 3.2 notes, and averaged 0.9 ***Z***, notes, 0.8 ***A*** notes, 0.1 $\boldsymbol{D_h}$ notes, and 1.3 ***D*** notes (Table I). Calls containing at least 1 ***Z*** note occurred with the highest frequency (roughly 80% of the calls in our sample; Table I) and calls containing at least 1 $\boldsymbol{D_h}$ note were least common (10% of the calls in our sample). Over half the calls in our sample had ***ZAD*** or ***ZA*** note compositions, and nearly 80% of the calls were composed of two or three total notes (Table II). We observed only one case (0.006 of the sample) of a call structure violating the ***Z-A-D_h-D*** rule of note ordering, a call with ***AZAD*** note composition.

A. Potential for note-type coding

The PNTC analysis assessed which acoustic parameters of the tufted titmouse chick-a-dee call notes may potentially be useful in distinguishing between note types (Table III). All parameters had PNTC scores greater than 1, indicating the potential for note types to be discriminated from one another on the basis of each parameter. Parameters with the greatest potential to be useful in note-type discrimination—that is, those with PNTC values >2 (Robisson *et al.*, 1993)—are TD, SF, MxF, MnF, and SE. Thus, three of the four frequency measures, note duration, and "noisiness" at the start of the note, appear to be best at potentially allowing discrimination of these four note types from one another.

TABLE II. The different call structures (note compositions) and their frequencies observed in titmouse chick-a-dee calls.

Note composition	Percent in sample	Cumulative percent
ZAD	33.8	33.8
ZA	17.5	51.3
ZZD_h	5.0	56.3
DDD	3.8	60.0
ZD_h	3.8	63.8
ZZ	3.8	67.5
DDDD	3.1	70.6
ZAA	2.5	73.1
AAD	2.5	75.6
ZADD	1.9	77.5
ZAAD	1.9	79.4
ZADDD	1.9	81.3
AA	1.9	83.1
ADDD	1.9	85.0
ZD	1.3	86.3
DDDDD	1.3	87.5
DDDDDD	1.3	88.8
ZZA	0.6	89.4
ZZAD	0.6	90.0
ZZADDDD	0.6	90.6
ZZD	0.6	91.3
ZADDDD	0.6	91.9
ZAADD	0.6	92.5
ZD_hDDD	0.6	93.1
ZD_hDDDDD	0.6	93.8
ZDD	0.6	94.4
ZDDDDD	0.6	95.0
AD	0.6	95.6
ADD	0.6	96.3
AADD	0.6	96.9
AZAD	0.6	97.5
DD	0.6	98.1
DDDDDDD	0.6	98.8
DDDDDDDD	0.6	99.4
DDDDDDDDDDDDDDDDDD	0.6	100.0

TABLE III. Potential for note-type coding (PNTC) values of our 17 acoustic parameters.

Parameter	PNTC
Duration	
TD	2.04
DMx	1.45
Frequency	
SF	3.79
EF	1.36
MxF	2.98
MnF	2.21
F1	1.34
F2	1.47
Amplitude	
SA	1.26
EA	1.22
MxA	1.38
MnA	1.08
RMS	1.23
Entropy	
SE	2.01
EE	1.28
MxE	1.70
MnE	1.93

B. Potential for individual coding

The PIC analysis assessed for each note type which acoustic parameters could be useful in distinguishing between individuals. We found that all parameters are >1 for the ***Z*** note type, which suggests it may be the note type most likely to carry individual signature features, followed by $\boldsymbol{D_h}$ (14 parameters), ***A*** (13 parameters), and ***D*** (5 parameters) note types (Table IV). Even though all 17 parameters for ***Z*** notes are >1, none of these parameters are considered to be highly useful in individual distinction, though TD, with a PIC value of 1.91, may be fairly useful. $\boldsymbol{D_h}$ notes do present PIC values >2, but this may be an effect of small sample size and should be interpreted with caution.

C. Factor analysis

A factor analysis on the 17 parameters generated four factors with eigenvalues >1 (Table V). We determined that a parameter loaded onto the factors if it had a value of 0.6 or greater. Of the 17 total parameters, 14 loaded onto the four factors. Factors 1 and 2 together account for 60.9% of the total variation, and are our main factors for analyses of possible note composition and sex effects described below. Frequency and entropy parameters load heavily onto Factor 1, amplitude measures load heavily onto Factor 2, duration parameters load heavily onto Factor 3, and an entropy and frequency change measure associated with the ending of the notes load heavily onto Factor 4.

D. Factor analysis: Are note types distinct?

We found a significant effect of "note type" on all four factors (Fig. 3; Factor 1: $F_{3,33}=242.9$, $P<0.001$; Factor 2: $F_{3,34}=8.5$, $P<0.001$; Factor 3: $F_{3,31}=31.0$, $P<0.001$; Factor 4: $F_{3,34}=79.3$, $P<0.001$). Post-hoc analyses (corrected for multiple comparisons) indicated that all four note types were significantly different from one another for Factor 1 (factor scores: $\boldsymbol{D_h}>\boldsymbol{Z}>\boldsymbol{A}>\boldsymbol{D}$) and for Factor 4 (factor scores: $\boldsymbol{D_h}>\boldsymbol{A}>\boldsymbol{D}>\boldsymbol{Z}$). For Factor 2, ***Z*** notes had significantly higher factor scores than did ***D*** and $\boldsymbol{D_h}$ notes. For Factor 3, $\boldsymbol{D_h}$ notes had significantly higher factor scores than did ***D*** and ***Z*** notes, and those three notes had significantly higher factor scores than did ***A*** notes. Taken together, all four factors do a reasonably good job of potentially discriminating the note types, with Factors 1 and 4 (frequency and entropy parameters) providing the strongest potential discrimination.

E. Factor analysis: Are individuals distinct?

"Individual" had a significant effect on at least one factor score for each note type (Table VI). For both ***Z*** and ***D*** note types, individual had a significant effect on all four factor scores. Factor 1 produced the largest number of significant individual-individual distinctions for ***Z*** notes and Factors 1 and 3 produced similar numbers of such individual-individual distinctions for ***D*** notes. In these cases, there were typically two or three birds with factor scores very different from other birds, accounting for most of the significant bird *x* bird comparisons.

F. Note composition and sex effects

Overall there was little effect of note composition detectable within this call set. Factor 2 scores for first ***A*** notes in calls were significantly affected by the number of ***A*** notes in calls ($F_{1,7}=48.55$, $p<0.001$; Fig. 4). The overall amplitude of the first ***A*** note in a call is higher when it is followed by 1 or more ***A*** notes than if it is the only ***A*** note in a call. Factor 1 scores for ***A*** notes were not significantly affected by the number of following ***A*** notes in the call ($F_{1,9}=2.98$, $p=0.118$). Factor scores for ***Z*** and ***D*** note types were also not significantly impacted by the number of ***Z*** notes or ***D*** notes, respectively, to follow (***Z*** notes: Factor 1 $F_{1,12}=2.64$, $p=0.130$; Factor 2 $F_{1,14}=1.75$, $p=0.208$; ***D*** notes: Factor 1 $F_{1,12}=0.02$, $p=0.889$; Factor 2 $F_{1,9}=2.33$, $p=0.161$). Similarly, we found no effect of the total number of notes in a call on factor scores for either ***Z*** (Factor 1 $F_{2,22}=1.07$, $p=0.361$; Factor 2 $F_{2,22}=0.02$, $p=0.978$), ***A*** (Factor 1 $F_{2,22}=0.85$, $p=0.44$; Factor 2 $F_{2,22}=0.10$, $p=0.905$) or ***D*** notes (Factor 1 $F_{2,17}=0.16$, $p=0.854$; Factor 2 $F_{2,18}=0.04$, $p=0.961$; Fig. 5). Finally, sex had no detectable effect on Factor 1 or Factor 2 scores for either ***Z*** (Factor 1 $F_{1,9}=0.09$, $p=0.769$; Factor 2 $F_{1,9}=0.60$, $p=0.458$, ***A*** (Factor 1 $F_{1,8}=0.08$, $p=0.780$; Factor 2 $F_{1,8}=0.01$, $p=0.944$), or ***D*** note types (Factor 1 $F_{1,9}=0.04$, $p=0.850$; Factor 2 $F_{1.9}=0.04$, $p=0.853$).

IV. CONCLUSIONS

In this study we used the acoustic parameters measured from chick-a-dee calls of tufted titmice to classify their note types into categories, to determine which of the note types may potentially be useful in note type and individual discrimination, and to describe possible effects of note composition or sex of signaler on note type acoustic structure. Simi-

TABLE IV. Potential for individual coding (PIC) values of our 17 acoustic parameters for each note type (first row for each parameter), and means ± SD (second row for each parameter). PIC values greater than 1 are potentially useful for discriminating individuals.

	Note type			
Parameter	***Z***	***A***	$\boldsymbol{D_h}$	***D***
TD	1.91	0.99	1.17	1.01
(s)	0.12±0.04	0.05±0.01	0.24±0.02	0.17±0.02
DMx	1.29	1.08	1.04	0.87
(s)	0.07±0.03	0.03±0.01	0.04±0.02	0.09±0.05
SF	1.69	1.56	3.73	1.02
(kHz)	9.21±0.51	8.94±0.46	9.56±0.76	2.52±0.68
EF	1.14	1.01	1.43	1.03
(kHz)	7.34±0.95	5.54±1.74	2.92±1.44	2.75±0.98
MxF	1.47	1.04	1.64	1.06
(kHz)	8.36±0.38	8.08±0.76	8.28±0.68	2.79±0.80
MnF	1.64	1.07	1.25	1.03
(kHz)	8.36±0.40	8.16±0.73	5.75±2.80	2.56±0.41
F1	1.54	1.11	2.14	0.92
(kHz)	0.85±0.51	0.86±0.74	1.28±0.89	−0.35±1.24
F2	1.24	1.03	1.26	0.95
(kHz)	1.02±1.02	2.53±1.75	5.36±1.69	−0.04±1.37
SA	1.31	1.03	0.99	0.87
(dB)	−31.93±4.65	−34.47±4.81	−36.56±3.67	−39.51±5.49
EA	1.35	1.03	1.20	0.86
(dB)	−31.92±4.77	−34.82±4.77	−38.34±4.11	−39.59±6.17
MxA	1.18	0.98	0.95	0.90
(dB)	−18.34±4.27	−21.71±4.30	−22.04±3.74	−27.78±5.21
MnA	1.20	0.98	1.44	0.88
(dB)	−30.47±4.83	−32.15±4.98	−40.49±3.71	−36.38±5.04
RMS	1.24	1.09	0.87	0.95
(V)	0.08±0.04	0.07±0.03	0.05±0.02	0.04±0.02
SE	1.69	1.20	1.46	0.97
	0.27±0.09	0.30±0.07	0.45±0.08	0.64±0.05
EE	1.13	1.05	2.19	0.98
	0.49±0.15	0.68±0.09	0.61±0.07	0.60±0.05
MxE	1.32	1.11	1.52	0.97
	0.17±0.04	0.28±0.14	0.32±0.10	0.54±0.07
MnE	1.13	0.95	2.99	0.96
	0.38±0.08	0.58±0.12	0.81±0.06	0.70±0.05

lar analyses have been completed on the chick-a-dee call of chickadee species for many years (see reviews in Sturdy *et al.* (2007) and Lucas and Freeberg (2007). Our study was prompted by the fact that *Baeolophus* and *Poecile* are closely related, and share many features of their natural histories and social organizations (Grubb and Pravosudov, 1994; Mostrom *et al.*, 2002), but very little is known about the structure and nature of the chick-a-dee calls of titmice.

We classified the notes of the chick-a-dee call of the tufted titmouse into four distinct note types: ***Z***, ***A***, $\boldsymbol{D_h}$, and ***D***, with high inter-observer agreement. Like chickadee (*Poecile*) species that have been documented, tufted titmouse chick-a-dee calls obey basic rules of note ordering, with over 99% of our call sample obeying a $\boldsymbol{Z \rightarrow A \rightarrow D_h \rightarrow D}$ rule (see Table II). The potential for note-type coding (PNTC) analysis indicated that all of our 17 measured parameters could potentially be useful in note-type discrimination. ANOVAs used here to detect an effect of note type on factor score can be considered a converging method of potential note-type distinctiveness. We found that note type had a significant effect

TABLE V. Factor loadings and variance explained for the factor analysis of the 17 acoustic parameters. See text for definitions of factor properties.

Parameter	Factor 1	Factor 2	Factor 3	Factor 4
SF	0.904			
SE	−0.776			
MxF	0.917			
MxE	−0.816			
MnF	0.869			
RMS		0.882		
SA		0.898		
EA		0.906		
MxA		0.868		
MnA		0.922		
TD			0.812	
DMx			0.829	
EE				0.744
F2				0.676
Eigenvalue	5.34	5.01	1.97	1.79
Variance explained	31.42	29.45	11.59	10.53
Cumulative variance	31.42	60.87	72.46	82.99

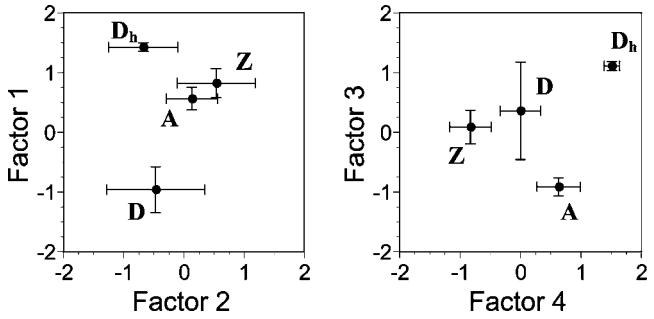


FIG. 3. Means (±1 SD) for each of the four factors emerging from factor analysis for ***Z***, ***A***, $\boldsymbol{D_h}$, and ***D*** note types.

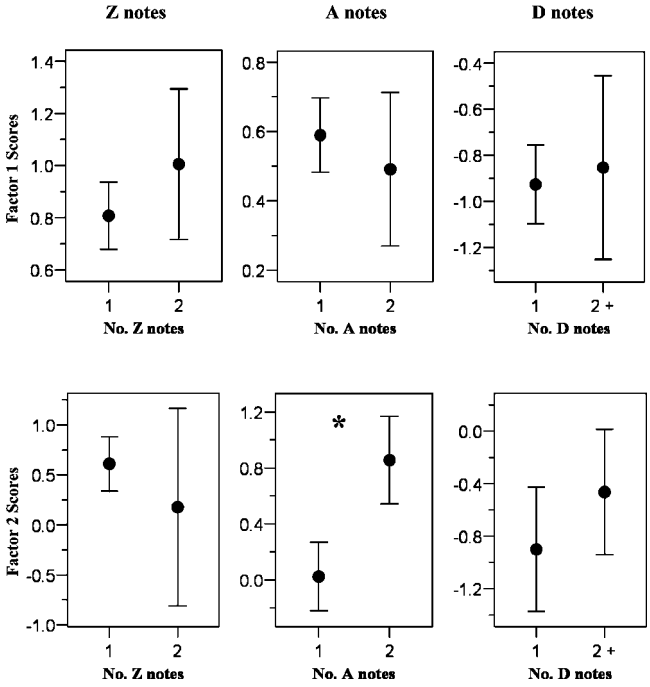


FIG. 4. Effect of note composition on factor scores for ***Z***, ***A***, and ***D*** notes: for the first ***Z***, ***A***, and ***D*** note type to occur in a call, the effect of one or more ***Z***, ***A*** or ***D*** note (respectively) on its factor 1 and factor 2 scores. Data are plotted as means and 95% confidence intervals.

on factor scores for all four factors that emerged in our analysis. Taken together, these two tests show, unsurprisingly, that acoustic characteristics vary reliably with note type (see also Dawson *et al.*, 2006 for a recent example of using artificial neural networks to distinguish note types of chick-a-dee calls of black-capped chickadees, *P. atricapillus*). Experiments like those by Sturdy *et al.* (2000), now need to be conducted to determine if titmice actually perceive the difference between note types, and which perceptual mechanisms guide this discrimination.

The chick-a-dee call of tufted titmice is also capable of encoding individual identity information, as seen in chickadee species (Bloomfield *et al.*, 2004, 2005; Charrier *et al.*, 2004). The potential for individual coding (PIC) analysis, as well as ANOVAs run on factor scores, indicate that individuals in our sample can potentially be discriminated on the basis of acoustical features of the note types. However, the results of the PIC analysis offer some different interpretations than the results of the ANOVAs run on factor scores. For example, for our 17 parameters, all 17 provided PIC scores greater than 1 (suggesting the potential for individual distinctiveness) for ***Z*** notes, but only five provided PIC scores greater than 1 for ***D*** notes (Table IV). This would suggest that ***Z*** note types in tufted titmice may allow for strong individual-level discrimination, but ***D*** notes would be more limited in their ability to do so (i.e., limited to note duration and some frequency measures). However, analysis of variance of factor scores for ***D*** and ***Z*** notes indicate that both provide a strong potential for individual distinctiveness (Table VI). Additionally, ANOVAs indicated a fairly weak effect of individual on the factor scores for both ***A*** and $\boldsymbol{D_h}$ notes note types (Table VI), but PIC analysis indicated that these note types should allow for fairly strong individual-level discrimination (Table IV). The explanation for these differences in PIC analyses compared to the factor score analyses will, as described above, need to be addressed with perceptual discrimination studies as conducted by Sturdy *et al.* (2000) or with call playback studies as conducted by Nowicki (1983). However, both analyses indicate that indi-

TABLE VI. Effect of individual on factor score for each note type. The *F* statistic for the influence of individual on factor score for each note type is displayed on line one with its *p*-value below it, and line three is the number of unique individual (bird *x* bird) comparisons that are significantly different, given that a significant effect of individual was detected. Individual comparisons were significant after corrections for multiple comparisons.

	Z	***A***	$\boldsymbol{D_h}$	***D***
(*ndf*, *ddf*)	(15, 129)	(14, 117)	(2, 13)	(14, 199)
Factor 1	$F=6.70$	$F=2.04$	$F=0.16$	$F=10.99$
	$p<0.001$	$p<0.05$	$p<0.85$	$p<0.001$
	19	0		17
Factor 2	$F=7.96$	$F=1.94$	$F=4.95$	$F=5.83$
	$p<0.001$	$p<0.05$	$p<0.05$	$p<0.001$
	8	0	0	7
Factor 3	$F=4.70$	$F=0.83$	$F=0.55$	$F=13.32$
	$p<0.001$	$p=0.64$	$p=0.59$	$p<0.001$
	5			20
Factor 4	$F=1.82$	$F=1.55$	$F=0.13$	$F=2.29$
	$p<0.05$	$p=0.11$	$p=0.88$	$p<0.05$
	2			1

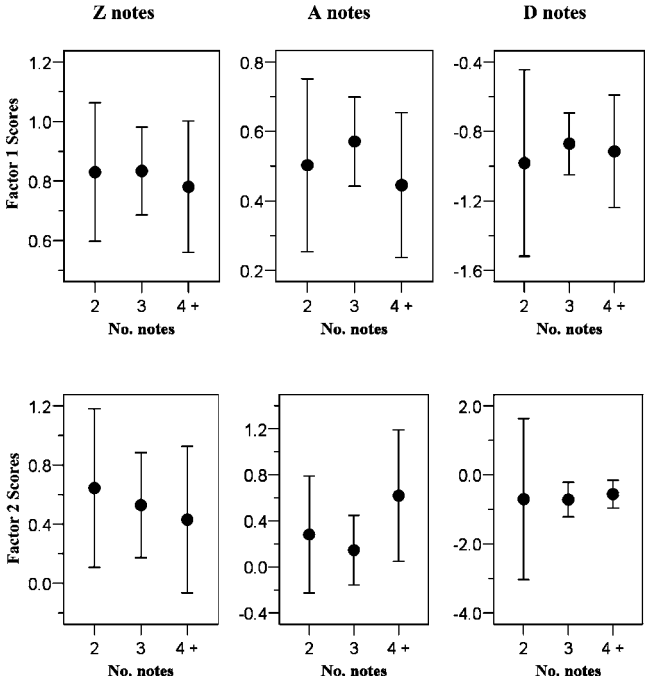


FIG. 5. Effect of note composition on factor scores for ***Z***, ***A***, and ***D*** notes: for the first ***Z***, ***A***, and ***D*** note type to occur in a call, the effect of the total number of notes in the call on its factor 1 and factor 2 scores. Data are plotted as means and 95% confidence intervals.

viduals are potentially distinguishable from one another by the acoustic characteristics of the note types of their chick-a-dee calls.

We suggest that the ***Z*** note type may be the primary note type to encode individual identity in tufted titmouse chick-a-dee calls, taking a number of lines of evidence into consideration. For example, the ***Z*** note occurs in the highest percentage of calls (Table I), and if it is present in a call, it virtually always occurs first (Table II). Also, PIC values for ***Z*** notes were >1.0 for all 17 acoustic parameters we measured (Table IV) and strong effects of individual were detected for all four factors that emerged from our factor analysis (Table VI). Playback studies manipulating ***Z*** note types (as well as other note types) are now needed to test this assertion.

We detected very little effect of note composition (call structure) on the acoustic structure of notes of calls. This finding contrasts with the pervasive effects of note composition on the acoustic structure of notes of the chick-a-dee call observed in Carolina chickadees (Freeberg *et al.*, 2003). Differences in call length between tufted titmice and Carolina chickadees are a likely contributor to the differences in the effect of note composition seen in these two species. The average number of note types in a Carolina chickadee chick-a-dee call is roughly twice that seen in tufted titmouse chick-a-dee calls. As described by Hailman *et al.* (1985), calls with more note types or more notes may be physiologically more difficult for an individual to produce than calls with fewer note types or fewer notes. Physiological constraints therefore may be an important causal factor in the note composition effect seen in Freeberg *et al.* (2003). Unlike Carolina chickadees, however, most tufted titmouse calls have few notes. In our sample, roughly 80% of the calls were composed of only two or three total notes. If the production of a four-note call is physiologically no more taxing to a titmouse than production of a two-note call, then this could well explain our finding of a general lack of a note composition effect. Larger samples of chick-a-dee calls of tufted titmice, that would uncover larger numbers of calls with greater numbers of notes, may yet reveal an effect of note composition on acoustic features of note types.

We also detected no effects of sex on Factor 1 or Factor 2 scores for the ***Z***, ***A***, and ***D*** note types in our sample. The earlier study with Carolina chickadees likewise found minimal sex effects on acoustic parameters of note types (Freeberg *et al.*, 2003). The lack of a strong sex effect on acoustic features of notes of chick-a-dee calls may not be too surprising, though, given the fact that sexes in these two species are highly monomorphic, and the call is used by both sexes throughout the year in facilitating social cohesion (Grubb and Pravosudov, 1994; Hailman, 1989; Hailman and Ficken, 1996; Mostrom *et al.*, 2002).

Individual distinctiveness in vocal signals is common in many avian (and nonavian) species (Stoddard, 1996). Individual distinctiveness has been hypothesized to be useful to birds with fairly stable social organization, like the Parids, in which individuals of a group may frequently be out of sight from one another and need to communicate (e.g., Marler, 1960). Note type and individual discrimination have been demonstrated in a few chickadee species, and we have now extended these findings to tufted titmice, a species related to chickadees, but one whose vocal behavior is much less understood. Given the complexity of the chick-a-dee call, including its structural similarities with human language (e.g., Hailman and Ficken, 1986; Hailman *et al.*, 1985), increased effort is needed to understand variation in structure and function of chick-a-dee calls of titmouse species, as well as other under-studied Parid species.

ACKNOWLEDGMENTS

We thank Richard Evans and the staff of the University of Tennessee Forest Resources, Research, and Education Center for assistance with the establishment of field sites and the construction of aviaries. This research was conducted under approved University of Tennessee Animal Care and Use Committee protocols (1248 and 1326). We thank two anonymous reviewers, Brad Bishop, Ellen Harvey, and Ami Padget for helpful comments on earlier drafts of this manuscript.

Baker, M. C., and Becker, A. M. (**2002**). "Mobbing calls of black-capped chickadees: Effects of urgency on call production," Wilson Bull. **114**, 510–516.

Bloomfield, L. L., Charrier, I., and Sturdy, C. B. (**2004**). "Note types and coding in parid vocalizations. II: The chick-a-dee call of the mountain chickadee (*Poecile sambeli*)," Can. J. Zool. **82**, 780–793.

Bloomfield, L. L., Phillmore, L. S., Weisman, R. G., and Sturdy, C. B. (**2005**). "Note types and coding in parid vocalizations. III: The chick-a-dee call of the Carolina chickadee (*Poecile carolinensis*)" Can. J. Zool. **83**, 820–833.

Charrier, I., Bloomfield, L. L., and Sturdy, C. B. (**2004**). "Note types and coding in parid vocalizations. I: The chick-a-dee call of the black-capped chickadee (*Poecile atricapillus*)," Can. J. Zool. **82**, 769–779.

Charrier, I., Jouventin, P., Mathevon, N., and Aubin, T. (**2001**). "Individual identity coding depends on call type in the South Polar Skua *Catharacta mccormicki*," Polar Biol. **24**, 378–382.
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.
Dawson, M. R. W., Charrier, I., and Sturdy, C. B. (**2006**). "Using an artificial neural network to classify note types in the "chick-a-dee" call of the black-capped chickadee (*Poecile atricapillus*)," J. Acoust. Soc. Am. **119**, 3161–3172.
Ficken, M. S., Hailman, E. D., and Hailman, J. P. (**1994**). "The chick-a-dee call system of the Mexican chickadee," Condor **96**, 70–82.
Freeberg, T. M., and Lucas, J. R. (**2002**). "Receivers respond differently to chick-a-dee calls varying in note composition in Carolina chickadees, *Poecile carolinensis*," Anim. Behav. **63**, 837–845.
Freeberg, T. M., Baker, M. C., Bloomfield, L. L., Charrier, I., Gammon, D. E., Hailman, J. P., Lee, T. T-Y., Lucas, J. R., Mennill, D. J., and Sturdy, C. B. (**2007**). "Synopsis: Complexities in vocal communication," in *Ecology and Behaviour of Chickadees and Titmice: An Integrated Approach*, edited by K. A. Otter (Oxford University Press, Oxford), pp. 235–240.
Freeberg, T. M., Lucas, J. R., and Clucas, B. (**2003**). "Variation in chick-a-dee calls of a Carolina chickadee population, *Poecile carolinensis*: Identity and redundancy within note types," J. Acoust. Soc. Am. **113**, 2127–2136.
Grubb, T. C., Jr., and Pravosudov, V. V. (**1994**). "Tufted titmouse," in *The Birds of North America*, (No. 86), edited by A. Poole and F. Gill (Academy of Natural Sciences and American Ornithologists' Union, Philadelphia and Washington, DC).
Haftorn, S. (**1993**). "Ontogeny of the vocal repertoire of the willow tit *Parus montanus*," Ornis Scand. **24**, 267–289.
Haftorn, S. (**2000**). "Contexts and possible functions of alarm calling in the willow tit, *Parus montanus*: The principle of 'better safe than sorry," Behaviour **137**, 437–449.
Hailman, J. P. (**1989**). "The organization of major vocalizations in the Paridae," Wilson Bull. **101**, 305–343.
Hailman, J. P. (**1994**). "Constrained permutation in "chick-a-dee"-like calls of a black-lored tit (*Parus xanthogenys*)," Bioacoustics **6**, 33–50.
Hailman, J. P., and Ficken, M. S. (**1986**). "Combinatorial animal communication with computable syntax: Chick-a-dee calling qualifies as 'language' by structural linguistics," Anim. Behav. **34**, 1899–1901.
Hailman, J. P., and Ficken, M. S. (**1996**). "Comparative analysis of vocal repertoires, with reference to chickadees," in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell University Press, Ithaca, NY), pp. 136–159.
Hailman, J. P., Ficken, M. S., and Ficken, R. W. (**1985**). "The 'chick-a-dee' call of *Parus atricapillus*: A recombinant system of animal communication compared with written English," Semiotica **56**, 191–224.
Lucas, J. R., Schraeder, A., and Jackson, C. (**1999**). "Carolina chickadees (Aves, Paridae, *Poecile carolinensis*) vocalization rates: Effects of body mass and food availability under aviary conditions," Ethology **105**, 503–520.
Lucas, J. R., and Freeberg, T. M. (**2007**). " 'Information' and the chick-a-dee call: Communicating with a complex vocal system," in *Ecology and Behaviour of chickadees and titmice: An integrated approach*, edited by K. A. Otter (Oxford University Press, Oxford), pp. 199–213.
Mammen, D. L., and Nowicki, S. (**1981**). "Individual differences and within-flock convergence in chickadee calls," Behav. Ecol. Sociobiol. **2**, 271–290.
Marler, P. (**1960**). "Bird songs and mate selection," in *Animal Sounds and Communication*, edited by W. E. Lanyon and W. Tavolga (AlBS, Washington, DC) pp. 348–367.
Mostrom, A. M., Curry, R. L., and Lohr, B. (**2002**). "Carolina chickadee," in *The birds of North America*, (No. 636), edited by A. Poole and F. Gill (Academy of Natural Sciences and American Ornithologists' Union, Philadelphia and Washington, DC).
Nowicki, S. (**1983**). "Flock-specific recognition of chickadee calls," Behav. Ecol. Sociobiol. **12**, 317–320.
Pyle, P. (**1997**). *Identification Guide to North American Birds* (Slate Creek Press, Bolinas, CA).
Robisson, P., Aubin, T., and Brémond, J. (**1993**). "Individuality in the voice of emperor penguin *Aptenodytes forsteri*: Adaptation to a noisy environment," Ethology **94**, 279–290.
Smith, S. T. (**1972**). "Communication and other social behavior in *Parus carolinensis*," Publ. Nuttal Ornithol. Club **11**, 1–125.
Sokal, R. R., and Rohlf, F. J. (**1995**). *Biometry* (Freeman, New York).
Stoddard, P. K., (**1996**). "Vocal recognition of neighbors by territorial passerines," in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Comstock, Ithaca, NY), pp. 356–374.
Sturdy, C. B., Phillmore, L. S., and Weisman, R. G. (**2000**). "Call-note discrimination in black-capped chickadees (*Poecile atricapillus*)," J. Comp. Psychol. **114**, 357–364.
Sturdy, C. B., Bloomfield, L. L., Charrier, I., and Lee, T. T.-Y. (**2007**). "Chickadee vocal production and perception: An integrative approach to understanding acoustic communication," in *Ecology and Behaviour of Chickadees and Titmice: An integrated Approach*, edited by K. A. Otter (Oxford University Press, Oxford), pp. 153–166.
Templeton, C. N., Greene, E., and Davis, K. (**2005**). "Allometry of alarm calls: Black-capped chickadees encode information about predator size," Science **308**, 1934–1937.
Thirakhupt, K. (1985). "Foraging ecology of sympatric parids: Individual and population responses to winter food scarcity," Unpublished doctoral dissertation, Purdue University, West Lafayette, IN.

The hydrodynamic footprint of a benthic, sedentary fish in unidirectional flow

Sheryl Coombs[a)]
Department of Biological Sciences and J.P. Scott Center for Neuroscience, Mind and Behavior, Bowling Green State University, Bowling Green, Ohio 43402

Erik Anderson
Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138 and Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Christopher B. Braun
Department of Psychology, Hunter College, New York 10021

Mark Grosenbaugh
Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 12 July 2006; revised 21 May 2007; accepted 22 May 2007)

Mottled sculpin (*Cottus bairdi*) are small, benthic fish that avoid being swept downstream by orienting their bodies upstream and extending their large pectoral fins laterally to generate negative lift. Digital particle image velocimetry was used to determine the effects of these behaviors on the spatial and temporal characteristics of the near-body flow field as a function of current velocity. Flow around the fish's head was typical for that around the leading end of a rigid body. Flow separated around the edges of pectoral fin, forming a wake similar to that observed for a flat plate perpendicular to the flow. A recirculation region formed behind the pectoral fin and extended caudally along the trunk to the approximate position of the caudal peduncle. In this region, the time-averaged velocity was approximately one order of magnitude lower than that in the freestream region and flow direction varied over time, resembling the periodic shedding of vortices from the edge of a flat plate. These results show that the mottled sculpin pectoral fin significantly alters the ambient flow noise in the vicinity of trunk lateral line sensors, while simultaneously creating a hydrodynamic footprint of the fish's presence that may be detected by the lateral line of nearby fish. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2749455]

PACS number(s): 43.80.Ka, 43.80.Lb, 43.80.Nd [MCH] Pages: 1227–1237

I. INTRODUCTION

The mottled sculpin (*Cottus bairdi*) is a benthic freshwater scorpaeniform fish that has a dorso-ventrally compressed head with a fusiform body shape very common in current-swept, freshwater and marine habitats (Webb *et al.*, 1996). These negatively buoyant fish can hold position in currents up to at least 12 cm/s (~1.2 body lengths/s) without being swept downstream (Webb *et al.*, 1996; Kanter and Coombs, 2003; Coombs and Grossmann, 2006). To prevent downstream displacement, mottled sculpin may theoretically use at least two different behavioral strategies. One is to orient upstream, thus presenting as low of a drag profile as possible to the downstream forces. Mottled sculpin also make substantial use of their broad pectoral fins, typically displayed outward at an angle of ~45–60° from the body surface, to resist downstream displacement (Webb *et al.*, 1996). That is, when fish are headed upstream in a current, the leading edges of their pectoral fins are angled downwards towards the substrate, thereby generating a downward lift force, which holds the fish to the substrate (Wilga and Lauder, 2001).

Recent studies have shown that visually deprived mottled sculpin from both Lake Michigan and Appalachian stream populations exhibit positive rheotaxis that increases in vector strength with increasing current velocity (Kanter and Coombs, 2003; Coombs and Grossmann, 2006). In the absence of vision, rheotaxis is likely mediated by the superficial neuromasts of the flow-sensing lateral line, which can theoretically determine both the magnitude and direction of the surrounding current (Montgomery *et al.*, 1997, Baker and Montgomery, 1999a,b). In contrast, Lake Michigan mottled sculpin require lateral line canal, but not superficial neuromasts to detect nearby (~1/2 body length away) artificial prey (50 Hz vibrating sphere) (Coombs *et al.*, 2001) and they can readily detect this artificial prey in both still and running water up to current velocities of 8 cm/s, with only a modest decrease in sensitivity resulting from ambient flow (Kanter and Coombs, 2003).

Not only do the splayed pectoral fins assist sculpin in holding station, they also alter the spatial and temporal char-

[a)]Author to whom correspondence should be addressed. Electronic mail: scoombs@bgnet.bgsu.edu

acteristics of the surrounding current. These alterations could provide a hydrodynamic footprint of the fish's presence that is detectable by nearby predators. They could also influence the information that the animal's own lateral line system receives about the direction or magnitude of ambient currents and the presence and location of other hydrodynamic sources (e.g., prey). Although a number of studies have described near-body flow fields and hydrodynamic trails generated behind active swimming organisms (e.g., Stamhuis and Videler, 1995; Müller *et al.*, 1997; Wolfgang *et al.*, 1999; Hanke *et al.*, 2000; Anderson *et al.*, 2001; Hanke and Bleckmann, 2004), very few have described the passive, near-body hydrodynamic effects caused by the presence of stationary benthic organisms in flowing water (e.g., Wilga and Lauder, 2001).

In this study, we use DPIV (digital particle imaging velocimetry) to determine the spatial and temporal characteristics of flow alterations caused by the body and pectoral fins of Lake Michigan mottled sculpin when orienting upstream at different flow velocities. Flow characterizations were done under experimental conditions identical to those used in rheotactic and prey-orienting studies (Kanter and Coombs, 2003), so that results could be directly related to the measured behavioral abilities of mottled sculpin as a function of flow velocity.

II. METHODS

A. Animal care and collection

Mottled sculpin (*Cottus bairdi*) (6–8 cm in standard length) were collected from Lake Michigan using baited minnow traps placed at depths of 1–4 m in near-shore waters and transported to the Coastal Research Center at Woods Hole Oceanographic Institution, where they were housed in 10 gal aquaria. Water in both home and experimental flow tanks was de-chlorinated tap water maintained at $15\pm2\ ^{\circ}\mathrm{C}$. Fish were hand fed small pieces of squid several times a week. Protocols used in the handling of animals were approved by Loyola University Chicago's Institutional Animal Care and Use Committee, the home institution of S. Coombs during the course of these studies.

B. Experimental setup

Fish were tested in a long Plexiglas rectangular channel of the same dimensions ($44\ \mathrm{cm}\times18\ \mathrm{cm}\times17\ \mathrm{cm}$) and of similar design to that used by Kanter and Coombs (2003). The channel was immersed in a downstream section of a large, recirculating, oval-shaped flume (7.6 m long, 76 cm wide, 30 cm deep), so that water could run through the channel and parallel to its long axis. Flume currents were driven by a conveyor belt of rotating paddles. Two collimators placed at the upstream end of the test channel helped to reduce turbulence in the flow. Fish were placed in a 27 cm long $\times$ 18 cm wide test section bounded by the second collimator on the upstream side and a mesh screen on the downstream side to prevent the fish from escaping into the larger flume. Freestream average velocities in the tank were measured and calibrated against each other with two techniques: DPIV and a Marsh-McBirney flow meter (Model 2000).

C. Digital particle image velocimetry (DPIV)

The flow field around the body of a sedentary, benthic fish heading directly upstream was imaged with DPIV (e.g., Adrian, 1991; Willert and Gharib, 1991). Fluid flow around the fish was illuminated by a 0.5-mm-thick, horizontal laser sheet, which was imaged from below with a high-resolution (1008×1012) digital camera. The flow field was seeded with nearly neutrally buoyant, light-reflective (silver-coated) glass spheres 10 μm in diameter. The elevation of the laser sheet was adjusted to between 4 and 12 mm above the channel substrate. For the normal resting position of sculpin, an 8–9 mm elevation is at the approximate level of the trunk lateral line canal. The laser sheet was pulsed as a strobe, synchronized to flash once in each frame acquired by the digital camera precisely recording the frame by frame positions of the illuminated fish and seed particles. Particle velocities were then determined by dividing the distance traveled in successive frames by the time interval (dt) between the laser pulses. Laser sheets were pulsed on and off in pairs and the displacement of the particles over time, as well as any movement of the fish, was imaged with a high-resolution, digital video camera. The laser pulses were synchronized with the digital camera in pairs such that the first laser pulse of each pair occurred at the end of one video frame and the second at the beginning of the next; the rate of image pair acquisition (15 Hz) was thus approximately half that of video frame acquisition (30 Hz). The time interval (dt) between laser pulse pairs varied according to flow speed, ranging from 8 (highest flow speed) to 24 ms (lowest flow speed).

D. Data analysis

To systematically examine the effects of four different flow speeds on both spatial and temporal variations in the flow field, repeated measurements were analyzed from a single individual, 7.5 cm in standard length. Although this does not allow us to say anything about individual differences, it does rule out inter-individual differences as a source of spatial and temporal variation. Video frame sequences of flow around the fish were examined frame by frame to ensure that only those free of fish-generated movements (other than opercular motions) were selected for further analysis. Selected sequences ranged in total duration from 1.3 to 6.7 s (20–100 image pairs), with longer sequences at slower flow rates. Each sequence was long enough for a single freestream particle to travel a downstream distance of at least one fish body length. Cross correlation of successive DPIV images was used to produce a two-dimensional (2D) picture of the velocity field. Velocity fields for sequential imaged pairs were then time averaged to determine the mean flow magnitude and direction over a 30×30 matrix of points (900 total) in a 10×10 cm field of view (Figs. 1(a)–1(c)). Vorticity plots, derived from the instantaneous two-dimensional flow field, were also time averaged to yield a measure of the mean vorticity magnitude and rotational direction (counterclockwise positive) at each point in the matrix (Figs. 1(d)–1(f)).

In addition to time-averaged plots, instantaneous regional (i.e., spatial) averages of velocity and vorticity were

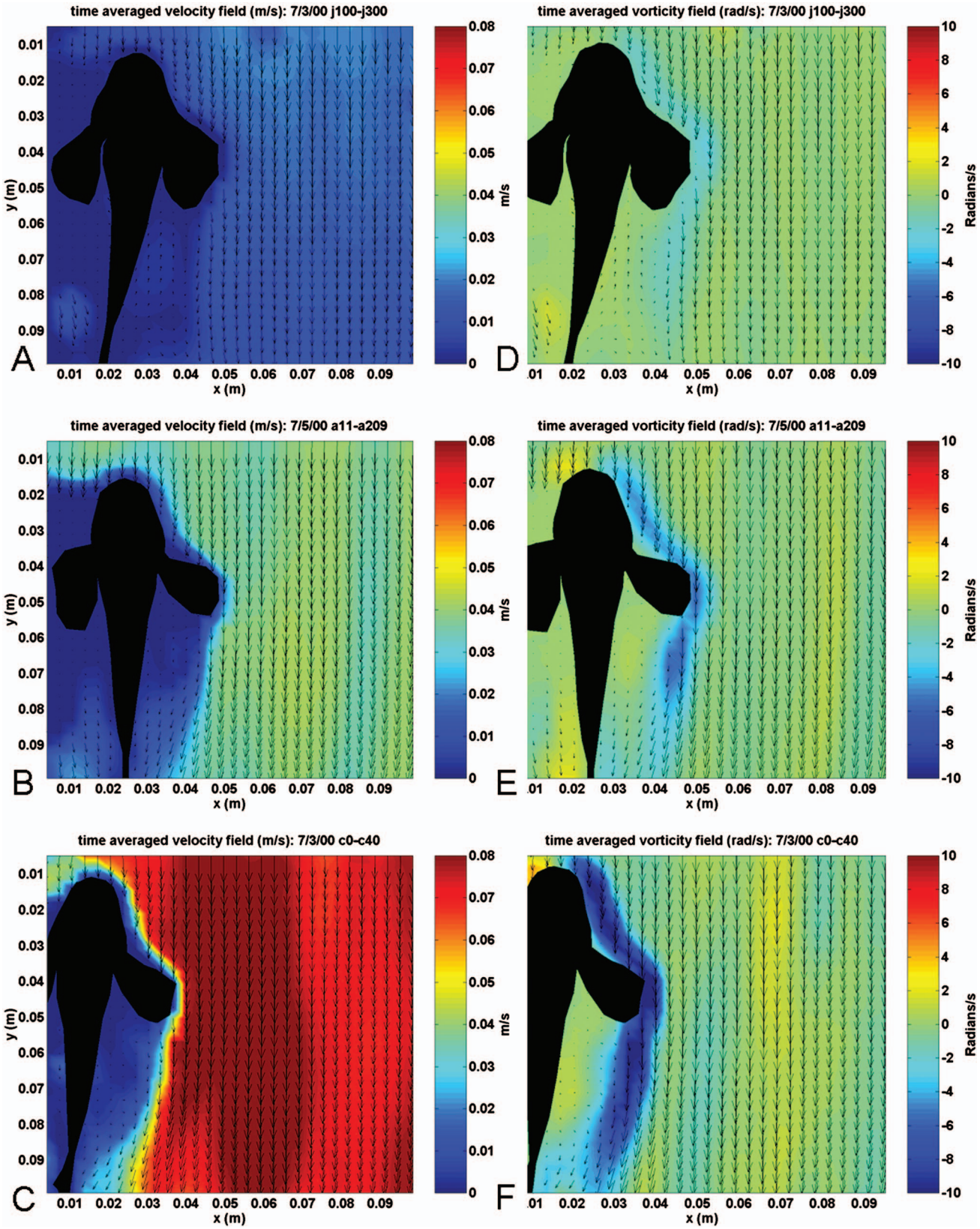


FIG. 1. Time-averaged velocity (a–c) and vorticity (d-e) plots for a single individual at three different background flow levels: 2 (a,d), 4 (b,e) and 8 (c,f) cm/s. Laser beam elevation=8 mm. Arrows depict mean flow directions at a total of 900 different locations in the 10×10 cm interrogation window and the color spectrum represents mean flow magnitude. Numbers at the upper right-hand corner of each plot indicate the range of frames that were averaged in each case. The average velocity and vorticity plots in a and d, for example, were based on a total of 200 frames (100–300) over a time period of 14 s (0.07 s/frame$\times$200 frames). The color scale for velocity (a–c) represents magnitude only, whereas that for vorticity (d–f) represents both magnitude and direction. Clockwise: negative, blue end of spectrum; counterclockwise: positive, red end of spectrum. Note that the light source for illuminating the particles is at the right of the fish and thus, flow information to the left of the fish should be ignored.

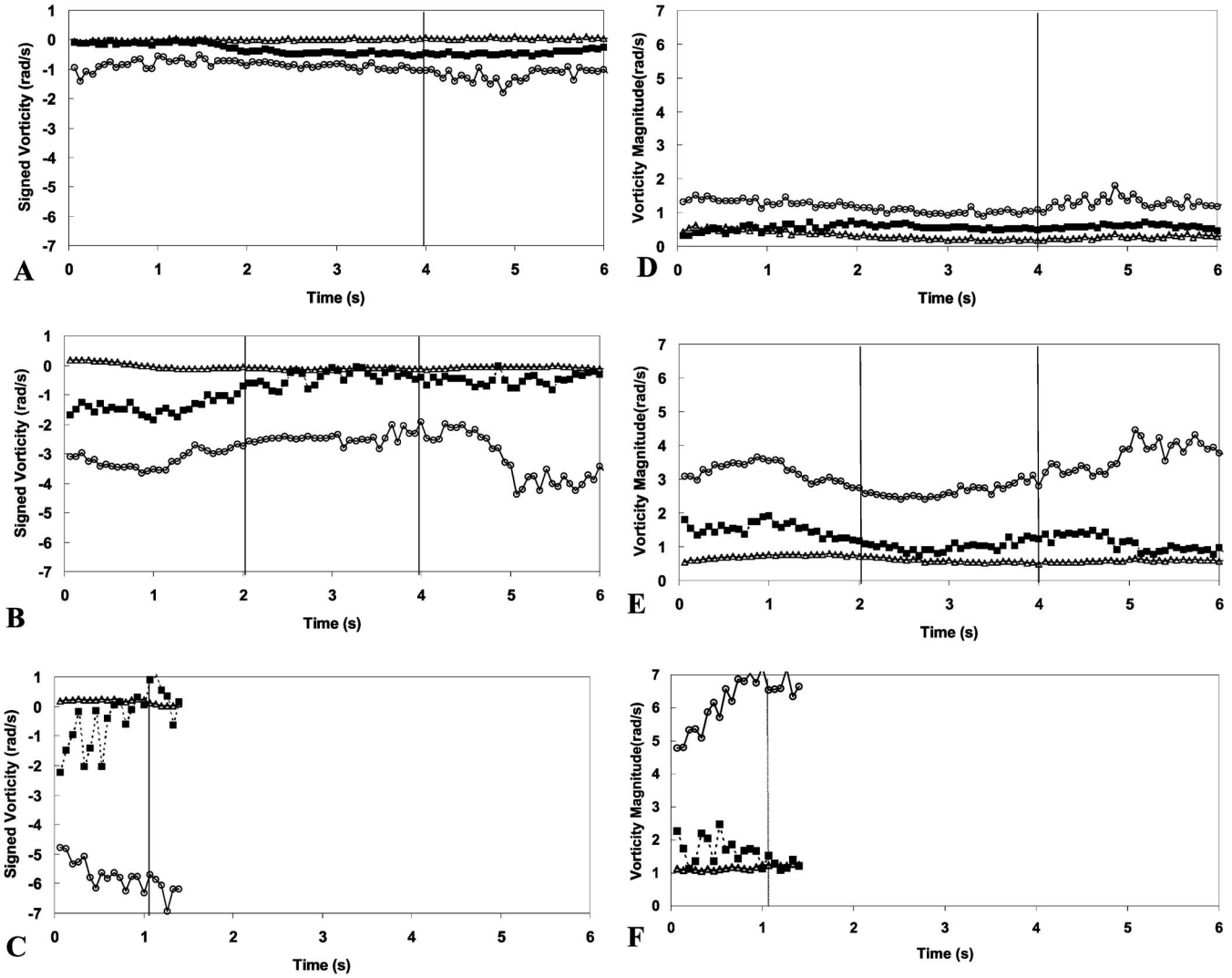


FIG. 2. Time-dependent changes in the signed vorticity (a, b, and c) and vorticity magnitude (d, e, and f) for the following regions of interest: freestream (solid line, open triangles), recirculating (filled squares) and shear layer (open circles) regions. Temporal variations in vorticity are plotted for three different freestream flow rates: 2 (a,d), 4 (b,e) and 8 (c,f) cm/s. Vertical lines in each panel indicate the approximate time that it takes for a freestream particle to traverse the length of the fish at different flow velocities.

determined for three specific regions: (1) The freestream region where the flow field was unaffected by the fish's body, (2) a reduced velocity region or separation bubble along the trunk and behind the pectoral fin and (3) the shear layer between the freestream region and the reduced velocity region. Because fish were free to move and their position within the video frame varied, the freestream region was user defined as a $12-16\ \mathrm{cm}^2$ rectangular region sufficiently far from the fish (>3 cm away from the fish's midline) and aligned with the velocity vectors and long axis of the fish. The reduced velocity region was defined empirically with a software algorithm that searched a user-defined area ($\sim 12\ \mathrm{cm}^2$) behind the pectoral fin and along the body surface to determine where the time-averaged velocity fell below a threshold criteria of $0.25\times$ mean freestream velocity. Similarly, the shear layer was defined by locations within the search region for which the time-averaged vorticity exceeded a threshold criteria of twice the maximum freestream vorticity. Empirically defined recirculating and shear layer regions were then graphed onto 2D, time-averaged velocity and vorticity plots and visually inspected to confirm that the algorithms had adequately captured these regions.

Two methods are used to report instantaneous regional vorticity. In the first, we computed the average vorticity, taking into account the sign of the vorticity. The major drawback to this method is that equal amounts of clockwise $(-)$ and counterclockwise $(+)$ vorticity lead to the false conclusion of no vorticity. The advantage of this method is that dominant rotational directions in the vorticity can be revealed when there are unequal amounts of clockwise and counterclockwise rotation. The second method, which avoids the mathematic canceling of counterrotating vorticity, is to simply average the magnitude (i.e., absolute value) of the vorticity in the region. These two methods represent the vector and scalar averages of vorticity, respectively. The vector and scalar averages of vorticity in each of the three regions of interest mentioned above were determined for each image pair of each video sequence and then plotted as a function of time (Fig. 2). Average values over the time series in Fig. 2 were then plotted as a function of the freestream flow veloc-

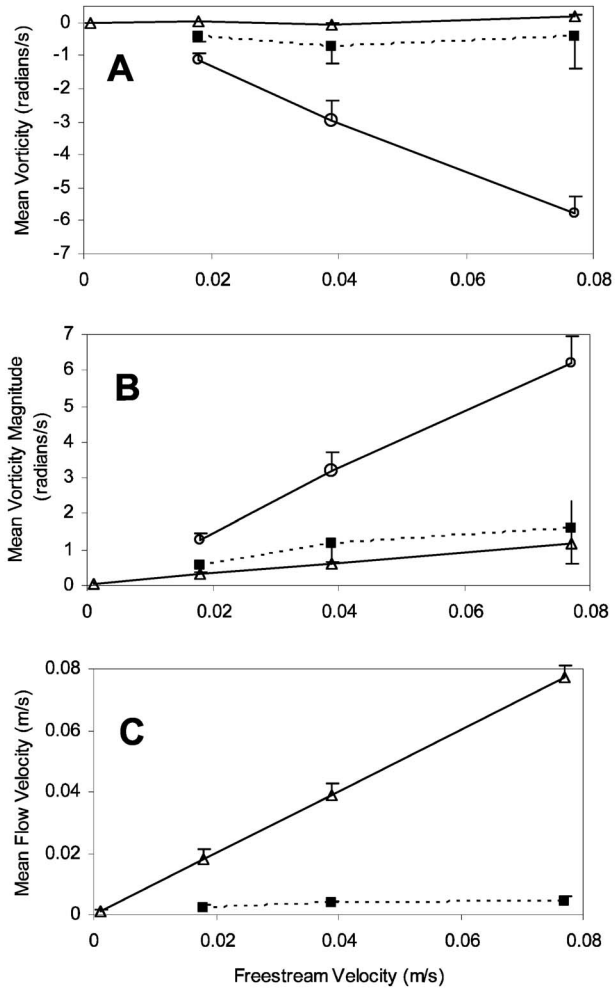


FIG. 3. Time averages of signed vorticities (a), vorticity magnitudes (b) and velocity in the streamwise direction (c) in the separation bubble (filled squares), shear layer (open circles) and freestream (open triangles) regions of the flow field. Instantaneous, spatially averaged values in each region were averaged over time to produce a final mean and standard deviation from the mean. Time averages come from the same data sets used in Fig. 2.

ity (Fig. 3(a)). Likewise, average magnitudes of velocity over time in the three regions of interest were calculated and plotted as a function of freestream flow velocity (Fig. 3(b), magnitude only). Additional time-varying features of the flow field were examined by viewing computer-generated sequences of the instantaneous (1) streamlines and (2) 2D velocity or vorticity plots over time. All measures were calculated using custom routines written in MATLAB (Mathworks, USA).

III. RESULTS

A. Regional differences in the flow field

At sufficient distances from the midline of the fish ($>\sim 3$ cm away), the time-averaged velocity magnitudes and directions were spatially uniform at all flow speeds relative to those closer to the body surface (Figs. 1(a)–1(c)). In this freestream region, the time-averaged vorticity field also revealed some spatial structure, i.e., a tendency for alternating "columns" of clockwise and counterclockwise vortices, presumably due to the presence of the upstream collimator (Figs. 1(d)–1(f)). Nevertheless, clockwise and counterclockwise vortices were nearly equal in abundance, if not randomly distributed relative to near-body regions (Figs. 1(d)–1(f)). This can best be seen from the near-zero values of the signed vorticity, which incorporates both magnitude and rotational direction (sign), both as a function of time for a given velocity (Figs. 2(a)–2(c), open triangles) and as time averages for different velocities (Fig. 3(a), open triangles). In contrast, the time averaged magnitude of vorticity increased with increasing flow velocity (Figs. 1(d)–1(f) and Fig. 3(b), open triangles).

Within 1–2 cm of the body surface, freestream flow rates as low as 2 cm/s ($\sim$0.25 body lengths/s) produced significant spatial and temporal nonuniformities in the flow field, including a reduced velocity region or separation bubble behind the pectoral fin (Figs. 1(a)–1(c)). The time-averaged, velocity magnitude in this region (2–8 mm/s) was approximately tenfold less than that of the freestream region (2–8 cm/s) (Fig. 3(c)) and the time-averaged vorticity (taking both magnitude and direction into account) revealed a slight bias in the clockwise (negative) direction (Fig. 3(a), solid squares) as would be expected for separated flow behind a flat plate at a high angle of attack to the flow. Separation of the flow at the trailing edge of the large, extended

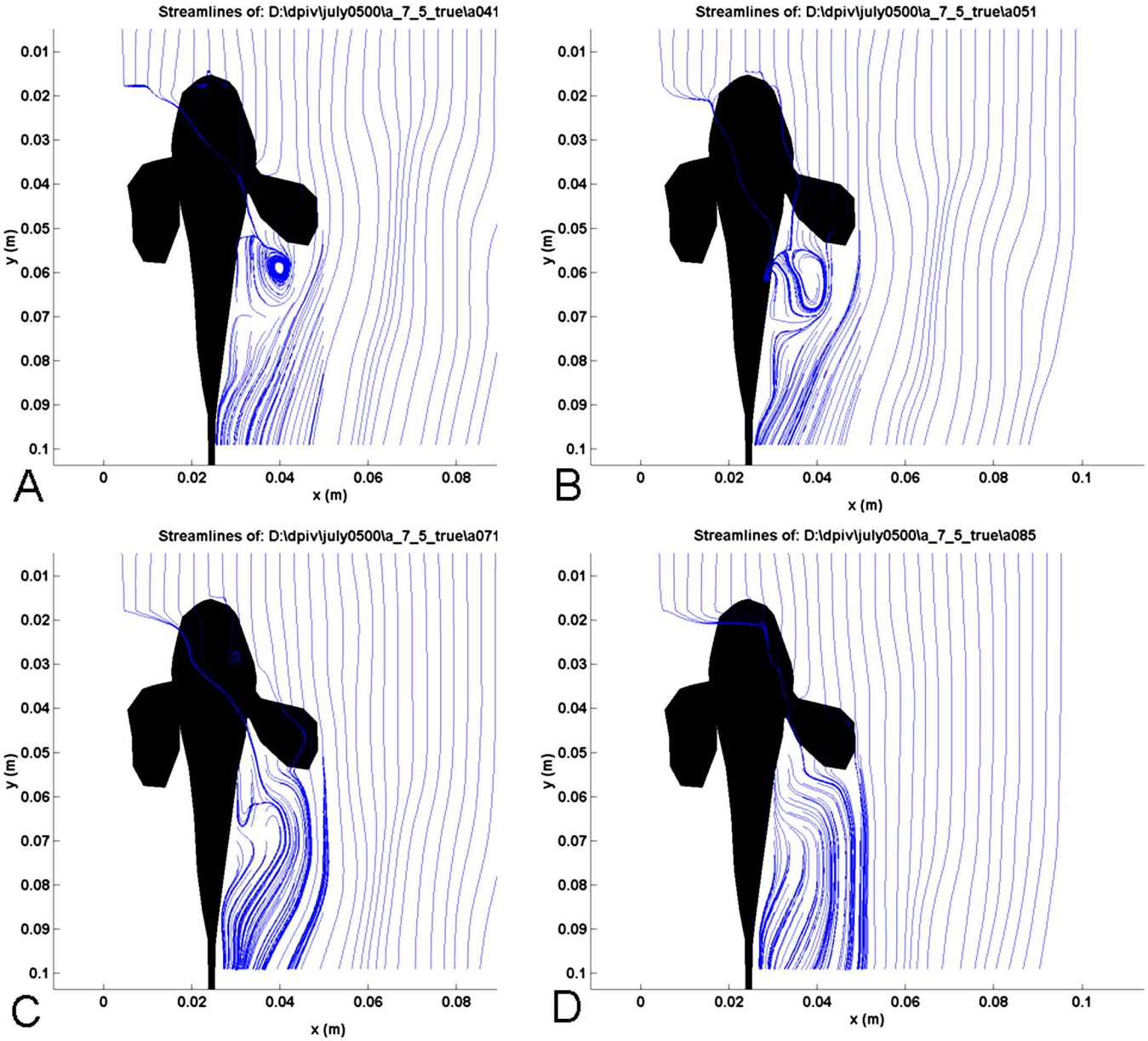


FIG. 4. (Color online) Streamline pictures for the 4 cm/s flow condition showing the formation (a) and subsequent shedding (b–d) of a vortex behind the pectoral fin over a time span of approximately 1.5 s. Note that the light source for illuminating the particles is at the right of the fish and, thus, flow information to the left of the fish should be ignored.

pectoral fin and the formation of a separation bubble is clearly revealed in both velocity field (Fig. 1) and streamline plots (Fig. 4). The magnitude of vorticity in the shear layer between the freestream and recirculating regions increased with increasing flow velocities (Figs. 1(d)–1(f), Fig. 3(b), open circles); vorticity in this region also had the expected clockwise (−) rotation (Fig. 1(d)–1(f), Fig. 3(a)).

Closer to the bottom of the test tank (elevation=4 mm) and well within the boundary layer of the substrate, mean flow velocities and vorticity were reduced (Figs. 5(a) and 5(b)) relative to those at the 8 mm elevation (Figs. 1(c) and 1(f)) for the same freestream velocity (8 cm/s).

B. Time-varying changes in the flow field

Because the DPIV sampling (image-pair acquisition) rate was 15 Hz, we were able to examine temporal changes in the flow field that were slower than ~7.5 Hz, or half the sampling rate. Plotted in Fig. 2 are time series of the spatial averages of the signed vorticity (a, b, and c) and vorticity magnitude (d, e, and f) for different flow regions and freestream flow rates. As this figure illustrates, the recirculating (filled squares) and shear layer (open circles) zones showed greater temporal oscillations than the freestream region (open triangles) at all flow speeds.

Time-varying changes in these spatially averaged values do not adequately capture all of the temporal and spatial changes that occur within a given region, however. As the flow separates off the pectoral fin, for example, the resulting shear layer boundary between the freestream and the recirculating region exhibits a slow (<1 Hz), wave-like undulation towards and away from the body surface, as observed from animated film sequences of 2D vorticity plots. As a

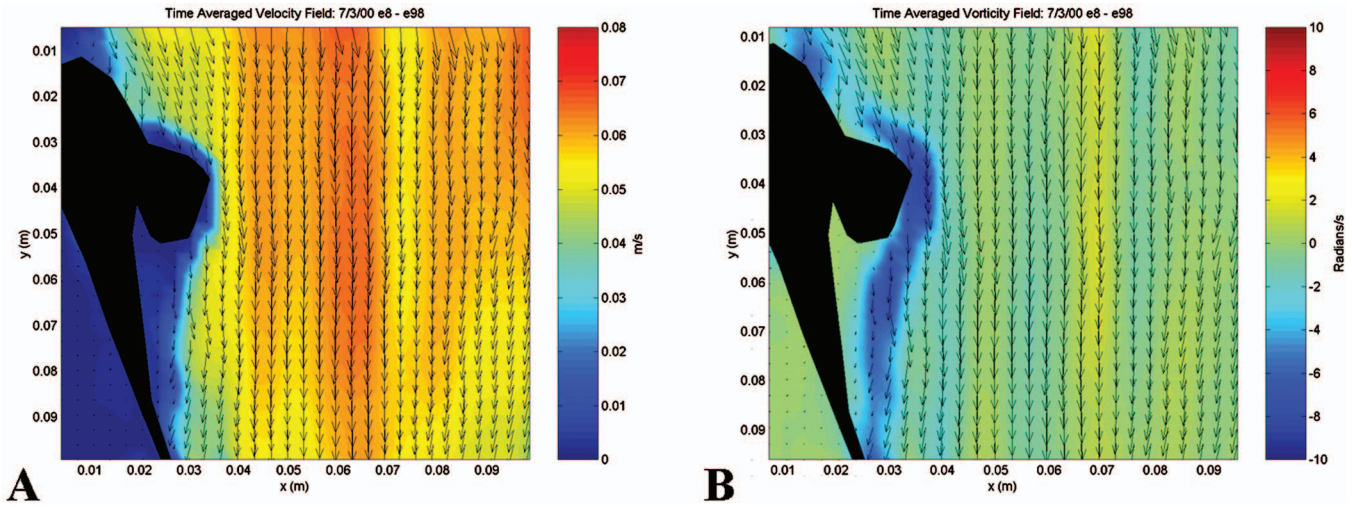


FIG. 5. Time-averaged velocity (a) and corresponding vorticity (b) plots for an 8 cm/s flow at a laser beam elevation of 4 mm. Compare with results from 8 mm beam elevation at same flow speed (Figs. 1(d) and 2(d)). Note that flow information to the left of the fish should be ignored.

result, the size and shape of the recirculating region behind the pectoral fin also changes. These slow fluctuations are most likely due to the periodic shedding of vortices, which form at the trailing edge of the pectoral fin (Figs. 4(a) and 4(b)) and are correlated with flow changes at the fish body surface. After a shedding event, a new vortex forms and the recirculating flow becomes more prominent. Fluid along the body behind the pectoral fin was observed to flow slowly upstream. Further downstream, near the peduncle, the flow was generally streamwise. Periodic vortex shedding adds a highly unsteady series of flow fields which "dilute" steady aspects of the time-averaged flow as the vortex moves downstream. Thus, in plots of time-averaged velocity, the average circulatory flow directly behind the pectoral fin is weaker than one might expect (Figs. 1(a)–1(c)). If the vortex behind the pectoral fin were stable, i.e., not-shedding periodically, the time-averaged velocity plots would reveal a more obvious circulatory flow. As one would expect, the vortex was more stable at slower flows; that is, shedding was less frequent, and indeed the circulatory flow directly behind the pectoral fin in the time-averaged velocity plots is most easily observed in the slower, 2 cm/s case (e.g., note upstream-directed arrows behind the pectoral fin in Fig. 1(d)).

Although not a planned part of this study, we were also able to discern respiratory flows due to the slow motions of the fish's operculum in the absence of imposed flow (Figs. 6(a) and 6(b)). The respiratory flow appeared to consist of a inhalant flow near the mouth and an exhalent flow caudal to the operculum and pectoral fin. The exhalant flow consisted of a slow dc (~2 mm/s) component that was ac modulated at the rate of ~0.7 Hz (Fig. 6(d)). AC fluctuations in surrounding flow velocity were also seen for several centimeters rostral to the pectoral fin, but without much evidence for a significant dc component (Fig. 6(c)). The *pk-pk* amplitude of ac flow modulations was greater near the trunk (~2 mm/s; Fig. 6(d)) than head (~1 mm/s; Fig. 6(c)).

C. Regional differences in signal-to-noise ratios in a prey-detection context

In order to gain an appreciation for how signal-to-noise (S/N) ratios might vary with ambient flow conditions in different regions of the flow field, we replotted signal detection data from previous experiments (Kanter and Coombs, 2003) conducted under ambient flow conditions nearly identical to those used in this study. In these experiments, the prey-orienting responses of mottled sculpin were used to determine threshold signal levels required for a 50% probability of orienting towards a small (6 mm diam), artificial prey (a 50 Hz vibrating sphere). Fish were oriented upstream at the time of signal onset and the signal source was approximately 5 cm (measured from the centerline) to the side of the fish at about the same rostro-caudal level as the point of pectoral fin insertion.

For the purpose of computing S/N ratios, threshold signal levels at the source (in *pk-pk* velocity units), rather than at the body surface of the fish, were used because these values do not involve untested assumptions about signal attenuation and presumed levels at the sensory surface of the fish. Noise levels were characterized in terms of mean velocity (as displayed in Fig. 3(c)), using the average values from the time series of spatially averaged magnitudes over the freestream region and reduced velocity (separation bubble) region behind the pectoral fin. S/N ratios in dB ($20^*\log$ [S/N]) were then plotted as a function of mean noise level in dB re: the noise level in no flow conditions (i.e., $20^*\log[N_x/N_0]$, where N_x=noise level for either the 0, 2, 4 or 8 cm/s flow conditions and N_0=noise level for the 0 cm/s flow condition) (Fig. 7). Since the no-flow condition did not produce distinct freestream and reduced velocity regions, a single noise level, based on a spatial average in the freestream region (as defined previously), was used to compute S/N ratios in the 0 cm/s flow condition. The nonzero noise values for this condition are presumably due to substrate vibrations, convective currents and the animal's own respiratory movements (Fig. 6).

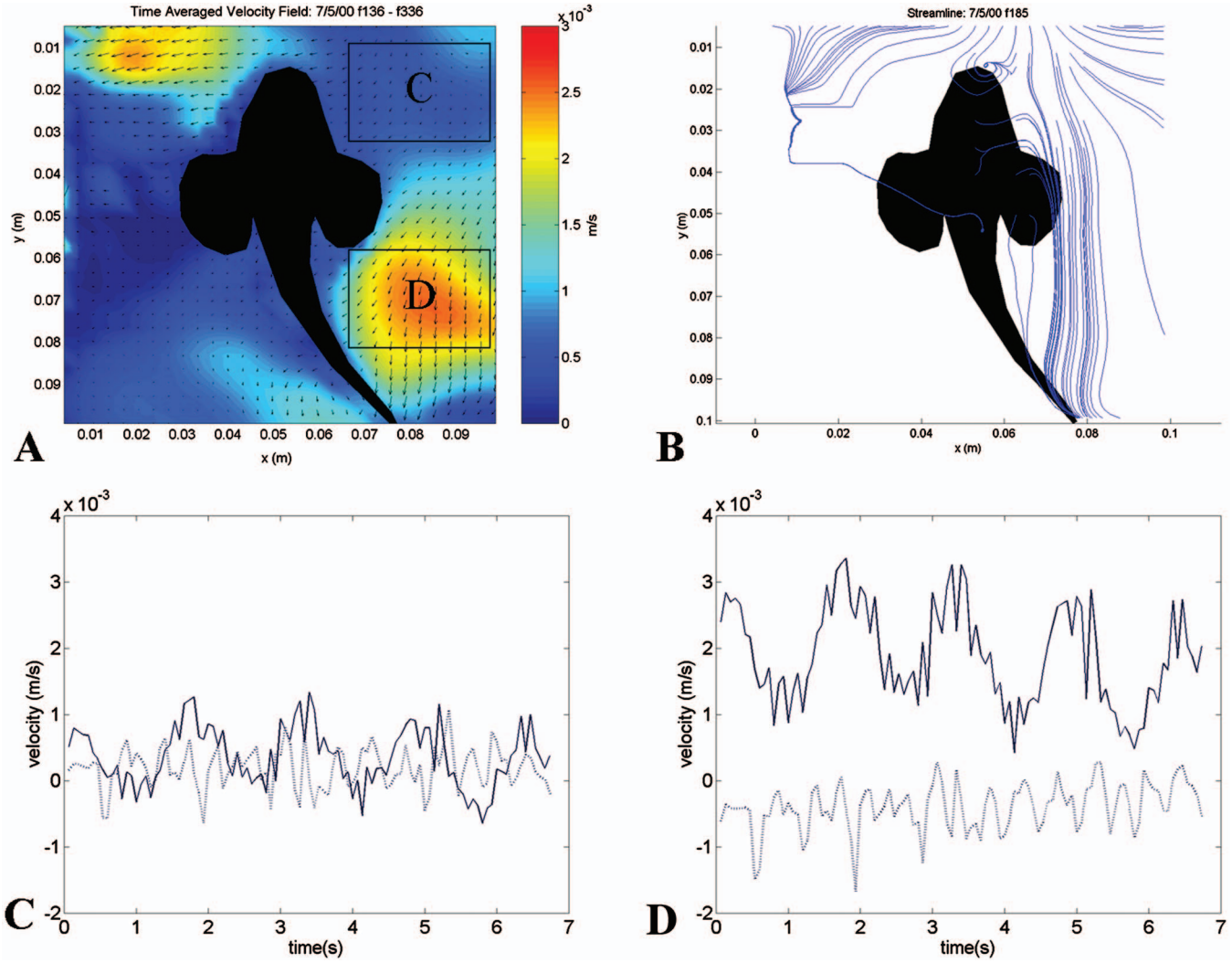


FIG. 6. Time-averaged velocity data are scaled to minimum and maximum velocity levels to show spatial patterns of fish-generated respiratory flow in the absence of imposed flow (a). Corresponding streamlines for a given image pair at a single moment in time are depicted in (b). Low- and high-velocity regions in front of (low) and behind (high) the pectoral fin were spatially averaged for each image pair of the video sequence to show time-varying changes in the mean respiratory flow for each region (low, c; high, d). Solid lines in c and d track spatially averaged velocity magnitudes in the streamwise (parallel to the long axis of the flow tank) direction, whereas dashed-lines track velocity magnitudes in the crosswise direction.

We should also point out that threshold signal levels from behavioral studies were measured within a 3 Hz bandwidth with an analog wave analyzer centered at 50 Hz. In contrast, DPIV sampling rates limited noise level measurements to frequencies below ~7.5 Hz. Thus, time wave forms of slow ac flows like those produced by the animal's own respiratory movements could be recovered (Figs. 6(c) and 6(d)), but those for higher frequency noises could not. In essence, the S/N ratios reported here represent ac signal levels at 50 Hz relative to low-frequency (dc-7 Hz) noise levels. Thus, it is important to point out that they can be nothing more than crude estimates of how S/N ratios vary with location and flow speed.

S/N ratios based on behavioral thresholds for the detection of a 50 Hz signal varied according to the spatial location of the noise (Fig. 7). Theoretical predictions for the complete presence or absence of noise interference (masking) are plotted for comparison (heavy solid lines in Fig. 7). Perfect masking predicts that, e.g., a twofold increase in noise will produce a twofold increase in the minimum level of the sig-

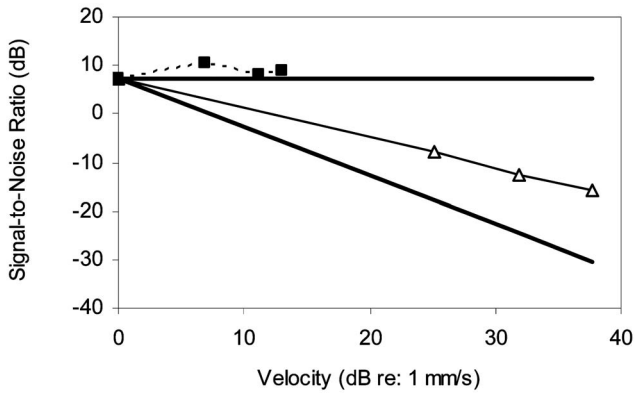


FIG. 7. Signal-to-noise ratio functions for two different regions of the flow field: freestream (solid line with open triangles) and recirculating region behind the pectoral fin (dashed line with solid squares). Thick solid lines show theoretical predictions for the complete presence (slope=0) and absence (slope=−1) of noise interference. See text for further details. Slope, *R* and *P*-value (slope significantly different from zero) regression statistics for S/N functions are −0.65, 0.99, 0.00023 (freestream) and −0.13, 0.77, 0.226 (separation bubble), respectively.

nal required for detection, resulting in S/N ratios that remain constant (solid line with slope of 0). A complete absence of masking, on the other hand, predicts that threshold signal levels will stay the same, regardless of ambient noise levels; S/N ratios will thus decline by the same amount as the noise level increases (solid line with slope of −1). As Fig. 7 shows, the S/N function for the freestream region has a negative slope near −1 that is significantly different from 0, whereas the S/N function for the recirculating region has a slope that is not significantly different from zero (see figure legend for regression statistics).

IV. DISCUSSION

A. Species comparisons

Our results show that the large pectoral fin of the mottled sculpin significantly alters ambient currents near the body surface in the vicinity of the lateral line and also behind the fin to leave a hydrodynamic wake. For fish heading upstream, the ambient flow is deflected by the pectoral fin and separates along the fin's edges, resulting in a low-velocity, recirculating region behind the pectoral fin and a trailing wake that presumably travels beyond the length of the fish.

In principle, our results are similar to those obtained by Wilga and Lauder (2001) in their study of pectoral fin function during station holding by the benthic bamboo shark (*Chiloscyllium plagiosum*). These investigators used DPIV to construct a 2D velocity matrix in the vertical plane for the purpose of computing lift forces on the pectoral fin of bamboo shark during station holding behavior in an upstream direction. Velocity profiles in this plane revealed a wake of clockwise vortices and a region of upstream flow behind and below the dorsal and trailing edge of the pectoral fin similar to those observed in the horizontal plane for the mottled sculpin behind the lateral edge of the pectoral fin and between the fin and the body surface.

In currents varying from 0 to 1 body length/s (up to ~50 cm/s), bamboo sharks were also observed to adopt station-holding behaviors similar to those of the mottled sculpin, including positive rheotaxis and an adjustment of the pectoral fin angle so that the leading edge was increasingly more ventral to the trailing edge as flow velocity increased. Although we did not directly measure fin angle in mottled sculpin as a function of flow velocity, we did observe a change in the pectoral fin position from one that was more nearly horizontal at 0 and 2 cm/s flow velocities to one that was more nearly vertical at higher flow velocities, with the leading (upstream) edge being more ventral than the trailing (downstream) edge. This effect can be seen as a corresponding change in the horizontal aspect of the fin, which has been traced from videotape images onto the velocity and vorticity plots (e.g., compare outline of right fin in Figs. 1(a) (2 cm/s) and 1(c) (8 cm/s)). In bamboo sharks, the vertical adjustment of the pectoral fin leads to negative lift forces as high as −0.084 N for a flow velocity of 0.75 body lengths/s (Wilga and Lauder, 2001). Thus, it is quite likely that pectoral fin positioning by mottled sculpin results in negative lift as well, as previously hypothesized by Webb *et al.* (1996). The friction of the animal against the substrate due to these negative lift forces helps fish to maintain a stationary position. Downstream drag forces, which result from high pressure areas in front of the fish and low pressure areas in the low-velocity, recirculating region behind the pectoral fin, and fluid shear in the boundary layer along the fish surface (Anderson *et al.*, 2001), work against this friction. Should these drag forces exceed the friction with the substrate, the fish will lose position.

Although it is tempting to speculate that species-specific differences in body form (e.g., the shape of the body or the shape, size, and insertion of the pectoral fin) might lead to quantitative or qualitative differences in the hydrodynamic signatures of these two species, this proposition is difficult to evaluate based on these two studies alone. For an equivalent flow speed of $\frac{1}{2}$ body lengths/s, the maximum vorticity magnitude in the wake of the bamboo shark's pectoral fin (~45 radians/s @ 17 cm/s; Wilga and Lauder, personal communication) was considerably higher than that measured for the mottled sculpin (~6 radians/s @ 4 cm/s) (Fig. 1(e)). How much of this difference is due to (1) morphological or kinematic differences between the two species, (2) differences in absolute flow speeds or (3) methodological differences in the plane of measurement or the exact location of the plane relative to the fish remains to be seen. Our measurements in the horizontal plane at two elevations (4 and 8 mm) (compare Figs. 1 and 5) clearly show that elevation can make a dramatic difference, not only in terms of vorticity magnitudes in the wake, but also in terms of vorticity and velocity magnitudes in the freestream region. Future DPIV studies should address all of these factors, and should include measurements taken at different elevations.

The wake behind the pectoral fin of the benthic sculpin is potentially a rich source of information to other nearby fishes. Although wakes actively generated by moving appendages and swimming fish have been widely studied (e.g., Hanke and Bleckmann, 2004), those generated by the passive interaction of an appendage in flowing water is less widely studied and appreciated (Fish and Lauder, 2006). Nevertheless, both behavioral (Pohlmann *et al.*, 2001, 2004) and physiological (Chagnaud *et al.*, 2006) experiments have shown that actively generated wakes and their vortex structures are potent lateral line stimuli. Catfishes preying on guppies are able to follow the wake left behind a swimming guppy for several seconds (up to 10 s) and at substantial distances from the prey (Pohlmann *et al.*, 2001; 2004). Moreover, primary lateral line afferents can code information about the size, shape and rotational direction of passing vortices in the wake of a stationary cylinder in a stream (Chagnaud *et al.* (2006)). It is likely that fish can analyze the hydrodynamic structure of a wake to determine the producer's size, swimming speed, mode of locomotion, and perhaps more. Future research should provide greater characterization of the hydrodynamic structures present in both passively and actively generated fish wakes and use similar structures in behavioral and physiological experiments to determine if and how fish make use of this hydrodynamic information.

B. Respiratory signals

It is well known that the ac modulation of dc electric currents across the gill epithelia of fish produce weak bio electric signals that can be detected by the electrosensory systems of nearby hetero- (e.g., predators) or conspecifics in the context of prey acquisition and mating behaviors (Sisneros and Tricas, 2002). In contrast, respiratory flow as a biologically relevant signal detected by the mechanosensory lateral line has received far less attention. For mottled sculpin, the nearby (<3 cm away from the fish) respiratory flow, which is amplitude modulated at the frequency (f) of 0.7 Hz, is relatively weak ~2 mm/s (*pk-pk* velocity, u) or ~9 mm/s^2 (*pk-pk* acceleration, a, where $a=2\ \mathrm{pi}\ f\ u$). Nevertheless, these levels are above threshold velocity and acceleration levels of response (~0.01 m/s and 1 mm/s^2, respectively) for mottled sculpin lateral line nerve fibers (Coombs and Janssen, 1990).

According to Hughes and Morgan (1973), a continuous, anterior-to-posterior flow of water through the mouth and across the gills in nonram ventilating fishes involves two pumps that are phase locked to alternately push (pressure pump) and pull (suction pump) water across the gills from the oropharyngeal to the parabranchial cavity. Our DPIV results are consistent with this description in that we see a continuous, downstream flow (Fig. 6(b)) that is ac modulated by gill movements (Figs. 6(c) and 6(d)). The modulation is strongest behind the operculum and pectoral fin (Figs. 6(a) and 6(d)), but weaker modulations are also observed frontally and laterally (Figs. 6(a) and 6(c)). It is worth pointing out that the pumping action of the opercula also changes the size and shape of the head. Thus, if sculpin maintain active branchial ventilation when exposed to externally imposed flows, changes in opercular positions are likely to cause further alterations in the surrounding (imposed+self-generated) flow field. Actively respiring animals may thus be surrounded by a bubble of temporal modulations that could give away their presence, whether ambient currents are present or not.

C. Effects of regional noise differences on signal detection and processing

The extent to which local flow alterations impact received information by the lateral line is still largely an open question, but it is likely that the effects will vary, depending on the behavioral task at hand, the location and type of both signal sources and lateral line sensors (i.e., superficial vs. canal neuromasts), the type and character of the ambient noise, and the overall size, shape and position of the fish's body and fins. In still-water conditions, the sensitivity of the mottled sculpin lateral line system to both live (e.g., *Daphnia*) and artificial (vibrating sphere) prey varies according to prey location. Sensitivity to prey along the trunk is poorer than that to prey along the head, though not directly in front of the head (Hoekstra and Janssen, 1986; Coombs and Janssen, 1990). Regional differences in sensitivity have been correlated with a variety of anatomical differences including the fact that neuromasts on the head are more densely packed and have greater numbers of hair cells per neuromast than those on the trunk (Janssen *et al.*, 1987).

There is presently very little information on how these sensitivity differences might affect prey detection when flow noise is present. Given that the sculpin's body alters the ambient noise field in the vicinity of the lateral line, the signal-to-noise ratio is likely to vary at different sensor locations along the body surface. It is conceivable, for example, that the ability of sculpin to detect small epibenthic prey within the reduced velocity region behind the pectoral fin would be nearly equivalent or even enhanced relative to their ability to detect the same prey at an equivalent distance from the head, where noise levels approach those in the freestream region. In any event, it is clear that regional differences in sensitivity need to be reexamined in terms of S/N ratios at different body locations.

In the context of prey detection by mottled sculpin in stream conditions, the choice of where flow noise is measured can lead to dramatically different conclusions about signal-to-noise processing capabilities (Fig. 7). For noise levels measured in the recirculating region behind the pectoral fin, S/N ratios were largely independent of the noise level, following the theoretical prediction for the presence of noise interference. Taken at face value, this "view" of the results is consistent with the idea that both signal and noise were "passed" through the same, low-pass channel (i.e., the superficial neuromast submodality of the lateral line), thus allowing the noise to interfere with signal detection. In contrast, when S/N ratios were based on noise levels measured in the freestream region, S/N ratios were seen to decrease with increasing noise levels, following the theoretical prediction for the absence of noise interference. In other words, this "view" of the results leads to a very different explanation—that the noise, but not the signal, was largely rejected by the known, high-pass filtering actions of lateral line canals.

In this particular case, there are several independent lines of evidence to suggest that the latter conclusion is most likely correct (as reviewed in Kanter and Coombs, 2003), including neurophysiological evidence (Engelman *et al.*, 2002) from goldfish showing that superficial neuromast, but not canal neuromast responses to dipole signals are reduced in the presence of flowing water relative to those in still water. Clearly, measurements of both signal and noise levels in different regions adjacent to the lateral line are needed before firm conclusions can be reached on this question. Nevertheless, this exercise illustrates the complexity of the problem and the care that must be taken to reach valid conclusions about S/N processing capabilities of the spatially distributed lateral line system.

Ironically, it is usually assumed that lateral line function, in particular that of superficial neuromasts, will be compromised in the presence of ambient flows, but in fact, local alterations in the flow field around the mottled sculpin's body predict that neuromasts behind the pectoral fin and along nearly the entire length of the trunk may not be compromised at all—at least with respect to the detection of nearby prey in this region. This puts a slightly new spin on the old hypothesis that lateral line sensors in some species may have been evolutionarily "rerouted" around the pectoral fin to circum-

vent self-stimulation by the animal's own fin movements (Dijkgraaf 1963). Indeed, the arching of the trunk canal above the pectoral fin is correlated with the general shape and size of the pectoral fin in a wide range of actively swimming fishes, but this does not appear to be the case for many sedentary, benthic fishes like the mottled sculpin (Webb, 1989). Rather than acting as a constant source of self-stimulation, the huge pectoral fin may in this case actually provide shelter from high levels of ambient flow noise to sensors on the trunk.

ACKNOWLEDGMENTS

This work was funded in part by an NIDCD program project grant to the Parmly Hearing Institute, Loyola University Chicago (W. Yost, PI, S. Coombs, Co-PI). S.C. and C.B.B. thank R. and K. Fay for their generous hospitality while we conducted this research at WHOI. Funding for the DPIV experiments was provided by NSF Grant No. IBN-0114148 and the Woods Hole Oceanographic Institution.

Adrian, R. J. (**1991**). "Particle imaging techniques for experimental fluid mechanics," Annu. Rev. Fluid Mech. **20**, 421–485.

Anderson, E. J., McGillis, W. R., and Grosenbaugh, M. A. (**2001**). "The boundary layer of swimming fish," J. Exp. Biol. **204**, 81–102.

Baker, C. F., and Montgomery, J. C. (**1999a**). "Lateral line mediated rheotaxis in the Antarctic fish *Pagnothenia borchgrevinki*," Polar Biol. **21**(5), 305–309.

Baker, C. F., and Montgomery, J. C. (**1999b**). "The sensory basis of rheotaxis in the blind Mexican cave fish, astyanax fasciatus," J. Comp. Physiol. **184**, 519–527.

Chagnaud, B. P., Bleckmann, H., and Engelmann, J. (**2006**). "Neural responses of goldfish lateral line afferents to vortex motions," J. Exp. Biol. **209**, 327–342.

Coombs, S., and Grossmann, G. (**2006**). "Mechanosensory-based orienting behaviors in fluvial and lacustrine populations of mottled sculpin (*Cottus bairdi*)," Mar. Freshwater Behav. Physiol. **39**, 113–130.

Coombs, S., and Janssen, J. (**1990**). "Behavioral and neurophysiological assessment of lateral line sensitivity in the mottled sculpin, *Cottus bairdi*," J. Comp. Physiol., A **167**, 557–567.

Coombs, S., Braun, C. B., and Donovan, B. (**2001**). "Orienting response of Lake Michigan mottled sculpin is mediated by canal neuromasts," J. Exp. Biol. **204**, 337–348.

Dijkgraaf, S. (**1963**). "The functioning and significance of the lateral-line organs," Biol. Rev. Cambridge Philos. Soc. **38**, 51–105.

Fish, F. E., and Lauder, G. V. (**2006**). "Passive and active flow control by swimming fishes and mammals," Annu. Rev. Fluid Mech. **38**, 193–224.

Hanke, W., Brücker, C., and Bleckmann, H. (**2000**). "The aging of the low-frequency water disturbances caused by swimming goldfish and its possible relevance to prey detection," J. Exp. Biol. **203**, 1193–1200.

Hanke, W., and Bleckmann, H. (**2004**). "The hydrodynamic trails of Lepomis gibbosus (Centrarchidae), Colomesus psittacus (Tetraodontidae) and Thysochromis ansorgii (Cichlidae) investigated with scanning particle image velocimetry," J. Exp. Biol. **207**(9), 1585–1596.

Hoekstra, D., and Janssen, J. (**1986**). "Lateral line receptivity in the mottled sculpin *Cottus bairdi*," Copeia **1**, 91–96.

Hughes, G. M., and Morgan, J. (**1973**). "The structure of fish gills in relation to their function," Biol. Rev. Cambridge Philos. Soc. **48**, 419–475.

Janssen, J., Coombs, S., Hoekstra, D., and Platt, C. (**1987**). "Postembryonic growth and anatomy of the lateral line system in the mottled sculpin, *Cottus bairdi* (Scorpaeniformes: Cottidae)," Brain Behav. Evol. **30**, 210–229.

Kanter, M., and Coombs, S. (**2003**). "Rheotaxis and prey detection in uniform currents by Lake Michigan mottled sculpin (*Cottus bairdi*)," J. Exp. Biol. **206**, 59–60.

Montgomery, J. C., Baker, C. F., and Carton, A. G. (**1997**). "The lateral line can mediate rheotaxis in fish," Nature (London) **389**, 960–963.

Müller, U. K., van den Heuvel, B. L. E., Stamhuis, E. J., and Videler, J. J. (**1997**). "Fish footprints: morphology and energetics of the wake behind a continuously swimming mullet (*Chelon labrosus* Risso)," J. Exp. Biol. **200**, 2893–2906.

Pohlmann, K., Grasso, F. W., and Breithaupt, T. (**2001**). "Tracking wakes: The nocturnal predatory strategy of piscivorous catfish," Proc. Natl. Acad. Sci. U.S.A. **98**, 7371–7374.

Sisneros, J. A., and Tricas, T. C. (**2002**). "Neuroethology and life history adaptations of the elasmobranch electric sense," J. Physiol. Paris **96**, 379–389.

Stamhuis, E. J., and Videler, J. J. (**1995**). "Quantitative flow analysis around aquatic animals using laser sheet particle image velocimetry," J. Exp. Biol. **198**, 283–294.

Webb, J. F. (**1989**). "Gross morphology and evolution of the mechanoreceptive lateral-line system in teleost fishes," Brain Behav. Evol. **33**, 34–53.

Webb, P., Gerstner, C., and Minton, S. (**1996**). "Station-holding by the mottled sculpin, *cottus bairdi* (teleostei: cottidae), and other fishes," Copeia **2**, 488–493.

Wolfgang, M. J., Anderson, J. M., Grosenbaugh, M. A., Yue, D. K. P., and Triantafyllou, M. S. (**1999**). "Near-body flow dynamics in swimming fish," J. Exp. Biol. **202**, 2303–2327.

Wilga, C. D., and Lauder, G. V. (**2001**). "Functional morphology of the pectoral fins in bamboo sharks, *Chilosycllium plagiosum*: Benthic vs pelagic station-holding," J. Morphol. **249**, 195–209.

Willert, C. E., and Gharib, M. (**1991**). "Digital particle imaging velocimetry," Exp. Fluids **10**, 181–193.

The influence of signal parameters on the sound source localization ability of a harbor porpoise (*Phocoena phocoena*)

Ronald A. Kastelein[a)]
Sea Mammal Research Company (SEAMARCO), Julianalaan 46, 3843 CC Harderwijk, The Netherlands

Dick de Haan
Wageningen IMARES (Institute for Marine Resources & Ecosystem Studies), P.O. Box 68, 1970 AB IJmuiden, The Netherlands

Willem C. Verboom
TNO Observation Systems, P.O. Box 96864, 2509 JG Den Haag, The Netherlands

(Received 9 December 2006; revised 11 May 2007; accepted 14 May 2007)

It is unclear how well harbor porpoises can locate sound sources, and thus can locate acoustic alarms on gillnets. Therefore the ability of a porpoise to determine the location of a sound source was determined. The animal was trained to indicate the active one of 16 transducers in a 16-m-diam circle around a central listening station. The duration and received level of the narrowband frequency-modulated signals (center frequencies 16, 64 and 100 kHz) were varied. The animal's localization performance increased when the signal duration increased from 600 to 1000 ms. The lower the received sound pressure level (SPL) of the signal, the harder the animal found it to localize the sound source. When pulse duration was long enough ($\approx$1 s) and the received SPLs of the sounds were high (34–50 dB above basic hearing thresholds or 3–15 dB above the theoretical masked detection threshold in the ambient noise condition of the present study), the animal could locate sounds of the three frequencies almost equally well. The porpoise was able to locate sound sources up to 124° to its left or right more easily than sounds from behind it. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2747202]

PACS number(s): 43.80.Lb, 43.80.Ev, 43.80.Jz, 43.80.Nd [WWA] Pages: 1238–1248

I. INTRODUCTION

Many harbor porpoises (*Phocoena phocoena*) are caught in gillnets as fisheries bycatch (Read and Gaskin, 1988; Lowry and Teilmann, 1994; Jefferson and Curry, 1994; Palka *et al.*, 1996; Bravington and Bisack, 1996; Trippel *et al.*, 1996). This probably occurs because porpoises often do not detect gillnets, or detect them to late to avoid them (Kastelein *et al.*, 2000). One potential method to reduce harbor porpoise bycatch in gillnets is to deter the animals from the nets by using alarms producing aversive sounds (Lien *et al.*, 1995; Kraus *et al.*, 1997; Laake *et al.*, 1998; Trippel *et al.*, 1999; Gearin *et al.*, 2000; Culik, 2001; Anonymous, 2000).

Behavioral research on wild and captive harbor porpoises has given some indication of which signal types have a deterring effect (Cox *et al.*, 2001; Culik *et al.*, 2001; Kastelein *et al.*, 1995, 1997, 2000, 2001; Olesiuk *et al.*, 2002; Teilmann *et al.*, 2006). However, the optimal spatial distribution of the alarms on the nets is unknown. Optimal alarm arrangements depend on the source level of the alarms, the background noise, the water depth, propagation losses (which in turn depend on frequency), and the directionality of porpoise hearing (Kastelein *et al.*, 2005).

Net avoidance by porpoises is not only a function of acoustic detection (hearing the alarms), but also of sound source localization (knowing where they are). The spatial localization of sound by vertebrates is based on binaural cues: differences in the perception of acoustic wave forms impinging on the two spatially separated ears (e.g., intensity or time of reception differences). Interaural intensity differences are important for the localization of high-frequency sounds; interaural time/phase differences for the localization of lower frequency sounds (for humans $<$2000 Hz; Atema *et al.*, 1988; Popov and Supin, 1992; Supin and Popov, 1993; Popov *et al.*, 2006).

Underwater, vertebrates face special problems with sound localization because the speed of sound in water is almost five times greater than in air, thus increasing wavelength and reducing interaural differences in arrival time of sound. Interaural intensity differences, which arise as a function of intra-ear spacing and, in odontocetes, from medial shielding of the auditory bullae by air-filled sinuses, are also used for sound source localization; this use depends on the orientation of the axis of equivalent receiving apertures and on the degree of shadowing of each ear from sounds coming from the opposite side. However, in odontocetes the auditory organs are largely isolated from the skull, enhancing sound source localization underwater (Dudok van Heel, 1962; Fleischer, 1980; Oelschläger, 1986a, b). *In vivo* anatomical inspections of bottlenose dolphins (*Tursiops truncatus*) using computed tomography has demonstrated the degree to which the tympano-periotic complex is shielded on the dorsal, medial and posterior surfaces by air-filled sinuses (Houser *et al.*, 2004): coverage by air is most complete on the dorsal and

a)Author to whom correspondence should be addressed. Electronic mail: researchteam@zonnet.nl

 0001-4966/2007/122(2)/1238/11/$23.00

medial surfaces. The air-filled sinuses and the tympanoperiotic complex provide a reflective barrier to the passage of sounds, including impulsive sounds generated during echolocation, between the ears. The arrangement of the sinuses contributes to time of arrival differences, amplitude differences, and spectral differences resulting from shadowing (Houser *et al.*, 2004), all of which contribute to an animal's ability to localize sound sources.

Sound source localization ability is usually studied by determining the species' minimum audible angle (MAA), which, when measured in front of an animal, is defined as the smallest angle at which a sound source is recognizable as being off the midline of an animal's long axis. The MAAs of the harbor porpoise for low-frequency sounds in the horizontal plane were determined in two studies with unrestrained animals (Andersen, 1970; Dudok van Heel, 1959, 1962 [the latter two references deal with the same study]). One porpoise did reasonably well at 2 kHz (average MAA=3° deviation from the median plane; Andersen 1970), but the other's performance at 3.5 kHz (MAA=22°) and 6 kHz (MMA=16°; Dudok van Heel, 1959; 1962) was poorer than that of the bottlenose dolphin, which has MAAs of 2–3° for tones between 10 and 80 kHz and around 4° for 6 and 90–100 kHz tones. Broadband clicks, resembling echolocation clicks, were spatially resolved with even greater precision than tones by a bottlenose dolphin (MAA=0.7–0.9° for clicks centered at 64 kHz; Renaud and Popper, 1975). The MAAs of the harbor porpoises have not been determined for the high-frequency sounds that they use for echolocation (120–140 kHz; Møhl and Andersen, 1973).

Because most presently commercially available alarms used in fisheries produce signals above 10 kHz, and an alarm could be at any angle to an animal when the first signal is received (not just in front of it), the abovementioned MAA studies give little insight into the abilities of harbor porpoises to localize acoustic alarms in all directions around their body. Therefore, more information is needed about the sound source localization ability of the harbor porpoise and its relationship to signal parameters. The aim of this study was to determine the effect of signal duration, received level, and frequency on the ability of a harbor porpoise to localize underwater sound sources, by emitting tonal sounds within a 360° circle around its head in the horizontal plane.

II. MATERIALS AND METHODS

A. Study animal

The study was done with a rehabilitated male harbor porpoise (PpSH052). The animal was stranded in November 1998 on the Dutch island of Ameland at the approximate age of 10 months. The cause of stranding was probably separation between mother and calf, and not illness or trauma. During the study, the animal was healthy and aged from two to three years. He weighed 20 kg in the 1999 study period and increased in weight from 23 to 29 kg in the 2000 study period. His standard length (straight line between tip of upper jaw and notch of tail fluke) during the study period was 110 cm in 1999 and increased from 113 to 119 cm in 2000. The animal was fed fish six times a day at 0900 h (sometimes during research), 1030 h (during research), 1230 h (during research), 1430 h (during research), 1545 h (sometimes during research), and 1700 h.

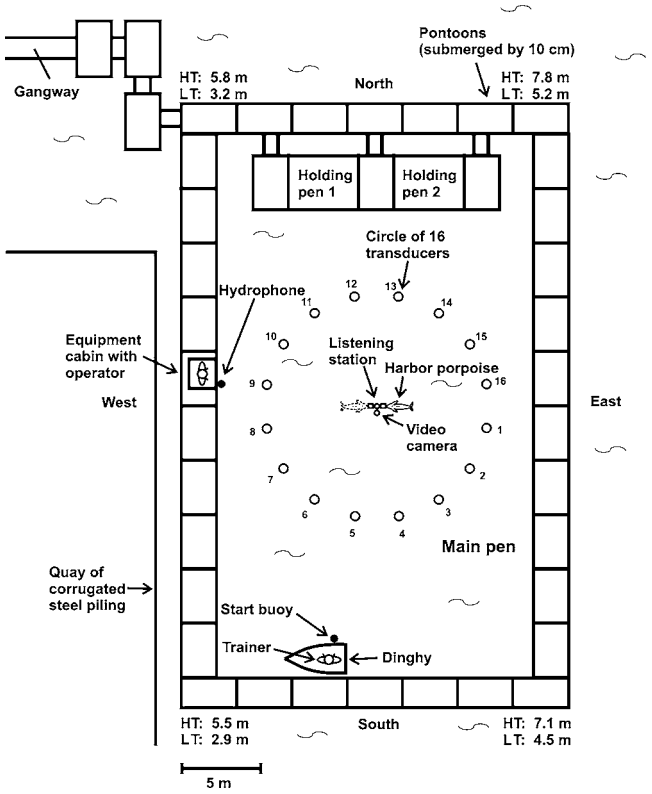


FIG. 1. Top view of the study area, showing the 16 transducers in a 16-m-diam circle around the listening station, the hydrophone, the underwater video camera, the equipment cabin housing the operator, and the trainer in the dinghy with the start buoy. The water depths at the four corners of the pen at high (HT) and low tides (LT) are also shown.

B. Study area

The study was conducted in two seasons (March–October of the years 1999 and 2000) when the porpoise was housed in a large floating pen (34 m×20 m; 3.5 m deep at the sides and 4–6 m deep in the center depending on the tide; Fig. 1). The bottom 10 cm of the surrounding pontoons (plywood boxes [3 m×2 m; 1 m high] filled with Styrofoam and coated with fiberglass) was below the water surface. The net at the bottom of the pen was made of nylon and the net of the sides was made of polypropylene. Both types of net had a stretched mesh size of 9 cm and a twine diameter of 3 mm. The twine of the net was covered with algae, and seawater flowed freely through the net. The floating pen was in a harbor that was horseshoe shaped (500 m ×280 m) with the entrance to the northeast. The harbor is in the southwest Netherlands, at Neeltje Jans (51°37′N,03°40′E). The harbor entrance is near the inside of the Oosterschelde surge barrier, which is only closed during exceptionally high tides and storms, but was open during the two seasons in which the study was conducted. Therefore, the tidal range inside the harbor was similar to that in the nearby North Sea. No shipping occurred within 2 km of the study area throughout the two study seasons. No activities occurred in, or near, the harbor during the study

TABLE I. Stimuli characteristics used in the study. Received SPLs (RL) were averaged over the 16 transducers (dB *re*: 1 μPa). Also shown are the basic audiogram hearing thresholds of the harbor porpoise for the three frequencies used in the present study (dB *re* 1 μPa; Kastelein *et al.*, 2002).

		1999					2000				
Center freq. (kHz)	Harbor porpoise hearing threshold (dB)	Total signal duration (ms)	Plateau (ms)	Rise time (ms)	Fall time (ms)	Average RL (dB)	Total signal duration (ms)	Plateau (ms)	Rise time (ms)	Fall time (ms)	Average RL (dB)
16	44	1000	720	120	160	89	1000	500	200	300	86
16	44	⋯	⋯	⋯	⋯	⋯	1000	500	200	300	94
64	46	600	320	120	160	82	1000	500	200	300	80
64	46	900	620	120	160	82	⋯	⋯	⋯	⋯	⋯
64	46	1000	720	120	160	82	⋯	⋯	⋯	⋯	⋯
100	32	⋯	⋯	⋯	⋯	⋯	1000	500	200	300	69

period, so no unexpected sounds occurred (this was verified during the sessions with an audio listening system). The harbor was sheltered, and no white crested waves occurred in the harbor while wind speeds were up to 5 on the Beaufort scale. In addition, the pontoons surrounding the pen were 90 cm high above the water surface, and thus sheltered the water surface in the pen.

The sea floor below the pen was flat and covered with sandy silt. The salinity in the pen was measured weekly and varied between 3.1 and 3.6%. The mean monthly water temperature varied between 14.2 and 19.6 °C. The underwater visibility (as determined by using a Secchi disk) was measured at noon of each test day and varied between 1.6 and 3.5 m. Water depth at the location of the pen, estimated wind speed and direction were recorded at the start of each session. Sessions were not carried out in wind speeds over 4 on the Beaufort scale or during rain. During the experiments the study animal was kept in the main pen and two other porpoises were kept in the holding pens (Fig. 1).

C. Acoustic stimuli

The sound localization ability was tested for three narrowband frequency-modulated (FM) signals with center frequencies (the frequency around which the modulation occurs) of 16, 64, and 100 kHz. The frequency modulation range was ±1% of the center frequency (resulting ranges: 15.84–16.16 kHz, 63.36–64.64 kHz, and 99.00–101.00 kHz) and the sinusoid modulation frequency was 100 Hz. FM signals were used as test stimuli to reduce standing wave effects that could occur as sine waves reflected off surfaces. The frequencies were chosen for the following reasons: 16 kHz is an octave frequency near the fundamental frequency of some commercially available alarms, 100 kHz was the maximum frequency that could be produced with a sufficient received level (RL) by the transducers (130 kHz would have been preferred as it is closer to the peak frequency of harbor porpoise echolocation clicks; Møhl and Andersen, 1973; Verboom and Kastelein, 1995, 1997, 2003), and 64 kHz was an octave frequency between the other two selected frequencies. The number of frequencies tested was determined by the time available for data collection.

The narrowband FM signals were generated by a wave form generator (Hewlett-Packard, model 33120A). The resolution was 12 bits and sample rate was 40 Mega samples/s. The generator was connected to a custom-built selector box that consisted of a signal shaper, a driver and a rotational switch to select the desired transducer (and therefore direction of the sound relative to the porpoise). The selector box drove transducers with a cylindrical piezo crystal element (LabForce 1 BV, model 90.02.01), which were omnidirectional in the horizontal plane. The correct setting of the voltage levels (amplitude) and the frequency of the stimulus were checked each session with an oscilloscope (Philips, model PM 3233).

Just before a session, 16 transducers (equally spaced between 0 and 360°, at 22.5° intervals), that were suspended 85 cm above the water surface from ropes, were lowered to a depth of 1.5 m in a 16-m-diam circle around a central listening station. The wire of each transducer went through a (10 cm × 6 cm) plastic float. Thus, above each transducer a float was visible at the water surface. The station consisted of a water-filled Polyvinylchloride tube (in 1999) or stainless steel tube (in 2000) with a 5-cm-long, 3-cm-wide rubber knob attached horizontally near the end. During trials, the porpoise was stationed with the tip of its upper jaw against the knob in the length axis. The knob pointed towards the east side of the pen in 1999, and alternated daily between pointing east and west in 2000 (Fig. 1). The knob was, like the transducers, 1.5 m below the water surface. After a session, the station and all 16 transducers were lifted out of the water with pulley systems. Next to each pulley, the number of the transducer below it was shown on a board visible to the operator.

In the 1999 season, signal durations were 600, 900, and 1000 ms including a 120 ms rise time and a 160 ms fall time (Table I). The rise and fall times were caused by the electronic components and the difference was not created intentionally. Based on the experiences in 1999, the total signal duration in the 2000 season was set at 1000 ms including a 200 ms rise time and a 300 ms fall time (Table I). Signals of over 1000 ms duration were not used, since during the sound production of longer signals the animal might have been able to direct its receiving beam by moving its head towards the active transducer (Kastelein *et al.*, 2005). A previous psychoacoustic study, in which the go/no-go paradigm was used, showed that a harbor porpoise of approximately the same age

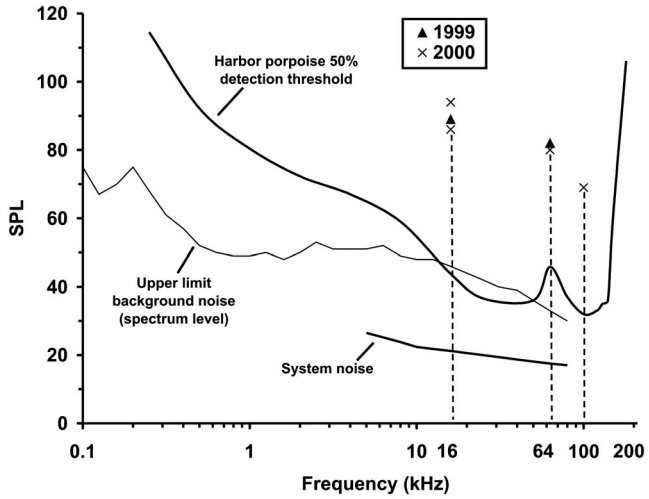


FIG. 2. The average RLs (line level in dB re 1 μPa rms) at the porpoise's listening station of the three test frequencies, in relation to the upper limit of the background noise range and the system (input) noise (spectrum level in dB re 1 μPa/$\sqrt{}$Hz), as well as the basic (unmasked) hearing threshold curve for a harbor porpoise (line level in dB re 1 μPa rms; Kastelein *et al.*, 2002).

as the animal in the present study usually needed 80 ms to react to the onset of a sound, which had a 150 ms rise time (Kastelein *et al.*, 2002). Depending on the frequency and RL, the animal in that study needed 880–1900 ms to move its rostrum 20 cm to the side. A head movement of 20 cm to the side during the production of a 1000 ms signal would have caused only a minor angular displacement of the subject relative to the transducers (~1.4°: less than the maximum variation in position allowed in the stationing animal when initiating a trial). Thus the animal reacting to the sound before its offset would not significantly change the transducer's angle of projection relative to the porpoise.

The selected SPLs at the porpoise's head (RLs) were based on the species' audiogram (Kastelein *et al.*, 2002), the background noise spectrum level at the three test frequencies, and the acoustic characteristics of the transducers. The frequencies tested in the present study were within the hearing range of harbor porpoises and the levels were well above (34–50 dB) the unmasked audiogram threshold levels (Kastelein *et al.*, 2002; Fig. 2 and Table I). However, detection depends not only on the basic audiogram, but also on the background noise level. The theoretical masked detection thresholds (MDTs) for the three signals in background noise conditions similar to those during the experiments are logically calculated as follows:

$$\mathrm{MDT} = L_{\mathrm{sp}} + \mathrm{CR} - \mathrm{DI}, \tag{1}$$

in which MDT is the theoretical masked detection threshold for pure tones (in dB re 1 μPa), L_{sp} is the (equivalent) spectrum level of the background noise (in dB re 1 μPa/$\sqrt{}$Hz), CR is the critical ratio of the hearing system (in dB re 1 μPa/$\sqrt{}$Hz), and DI is the receiving directivity index (in dB re 1 μPa). The directionality of hearing of the harbor porpoise has been studied electro-physiologically (Voronov and Stosman, 1983) and behaviorally (Kastelein *et al.*, 2005). The DI of harbor porpoises is 4.3 dB for 16 kHz, 6.0 dB for 64 kHz, and 11.7 dB for 100 kHz tonal signals (Kastelein *et al.*, 2005). No information is available on the CR of harbor porpoises. A guideline CR can be estimated based on human data and the measured CR of bottlenose dolphins (Johnson, 1968b). Assuming a CR of 36.5, 45, and 48 dB re 1 μPa/$\sqrt{}$Hz for harbor porpoises at 16, 64, and 100 kHz, respectively, the frequency-dependent factors (CR-DI) are estimated to be 32 dB re 1 μPa/$\sqrt{}$Hz at 16 kHz, 39 dB at 64 kHz and 36 dB at 100 kHz. Based on the information above, all signals produced in the present study were audible to the porpoise, although with a varying RL/MDT ratio (Fig. 2 and Table II).

D. Calibration

To measure the RL by the animal during the tests, a hydrophone (Brüel & Kjær [B&K], model 8101) was placed in the position of the porpoise's head while it was at the listening station (1.5 m below the water surface). The hydrophone was connected to a conditioning amplifier (B&K, model Nexus 2690). The output of the conditioning amplifier was connected via a BNC-2090 (National Instruments) coaxial module to a computer (Dell, model XPS D200) with a data-acquisition card (National Instruments, DAQ model PCI-MIO-16E-1, 12 bits resolution). To calculate the SPLs, a pistonphone (B&K 4223) was used to refer each recorded data sample to this calibrated voltage/SPL ratio, providing a correction for analogue and digital system errors. The analysis of SPLs was carried out using a custom-made analysis module based on LABVIEW 4.1 software (National Instruments). Each study season the signals were calibrated twice

TABLE II. The relationship between received SPL (RL, rms) at the listening station and calculated masked detection threshold (MDT). Below a 10 dB difference between RL and MDT, the porpoise's performance may have been influenced by the limited RL/MDT ratio.

Year	Frequency (kHz)	RL (dB re 1 μPa)	MDT (dB re 1 μPa)	Difference (RL/MDT ratio) (dB)	Possible influence of RL/MDT ratio on results
1999	16	89	79	+10	No
2000	16	86	79	+7	Yes
2000	16	94	79	+15	No
1999	64	82	70	+12	No
2000	64	80	70	+10	No
2000	100	69[a]	66	+3	Yes

[a]Maximum producible level due to the characteristics of the transducers.

TABLE III. The average (N=2) RLs per transducer of the 16, 64 and 100 kHz test signals by the harbor porpoise while at the listening station in the middle of the circle of 16 transducers (for the locations of the transducers see Fig. 1). The values are for the 2000 study season.

	Frequency					
	16 kHz		64 kHz		100 kHz	
Transducer	Average RL (dB re 1 μPa)	Deviation from average over all transducers	Average RL (dB re 1 μPa)	Deviation from average over all transducers	Average RL (dB re 1 μPa)	Deviation from average over all transducers
1	95	+1	81	+1	67	−2
2	93	−1	80	0	73	+4
3	96	+2	83	+3	75	+6
4	93	−1	79	−1	66	−3
5	96	+2	81	+1	68	−1
6	94	0	81	+1	66	−3
7	95	+1	81	+1	71	+2
8	92	−2	80	0	70	1
9	96	+2	84	+4	70	1
10	96	+2	81	+1	73	+4
11	94	0	80	0	66	−3
12	91	−3	78	−2	72	+3
13	96	+2	79	−1	69	0
14	94	0	79	−1	72	+3
15	96	+2	81	+1	70	+1
16	96	+2	79	−1	64	−5
Average	94		80		69	

(at the start and end of the season). The average RLs (rms) at the listening station for each of the three test frequencies from each transducer in the 2000 study season are shown in Table III. Differences between the two SPL measurement sessions of one year were <3 dB per frequency per transducer.

E. Background noise

The equipment used to measure the background noise consisted of a broadband hydrophone (B&K 8101, 0–100 kHz), a voltage amplifier system (TNO-TPD, 0–300 kHz) and an analyzer system (Hewlett-Packard 3565, controlled by a notebook computer; frequency range 0–100 kHz, sample frequency 260 kHz, df=31 Hz, fast Fourier transform measurement converted to 1/3-octaves bandwidths). The system was calibrated with a pistonphone (B&K 4223) and a white noise signal into the hydrophone preamplifier input. Measurements were corrected for the frequency sensitivity of the hydrophone and the frequency response of the measurement equipment.

Background noise levels were determined in the range 100 Hz–80 kHz and were converted to "spectrum level" (dB re 1 μPa/$\sqrt{}$Hz). Figure 2 shows the background noise upper limit in the pen. Also shown in this figure is the noise produced by the measurement system. The background noise in the pen was measured under conditions similar to those under which actual sessions were carried out.

The B&K 8101 transducer is omnidirectional in the horizontal plane. The directivity of the ambient noise thus could not be tested. If the background noise had been directional, the signal to noise ratio would have been different for each transducer. However, because no specific sounds were produced in the harbor, or within a 2 km radius outside the harbor during the study, the sound field in the pen was assumed to be uniform.

F. Acoustic monitoring

Before sessions, the signal-producing equipment was checked by using a hydrophone (LabForce 1 BV, model 90.02.01) connected to a heterodyne bat detector (Batbox III, Stag Electronics). The hydrophone was 1.5 m below the water surface, next to the equipment cabin (Fig. 1). To monitor the audible part of the underwater background noise during the sessions, the same hydrophone was connected to a charge amplifier (B&K 2635) of which the output was connected to an amplified loudspeaker.

G. Test procedure

Before a session began, a few transducers were randomly activated so that the study animal, which was swimming in the pen, became familiar with the test frequency to be used during the session. A session began when the trainer called the animal to the start buoy connected to the dinghy by tapping the buoy (Fig. 1). Following a hand signal from the trainer, the animal then swam towards the listening station and positioned itself correctly on the rubber knob (which pointed either east or west; Fig. 1). The direction was changed by turning the rubber knob at the end of the station between test days. This switching was done during the 2000 season because some of the results from 1999 showed an asymmetry in the animal's performance. By switching the

direction of the animal each day in 2000, it was possible to test whether this asymmetry was due to the acoustic characteristics of the environment (reflections), or to the hearing characteristics of the animal in the horizontal plane.

The animal's position at the station was filmed from above (the camera was attached to the station), and images were monitored and recorded by the operator in the equipment cabin. A maximum difference of only 2° between the animal's body axis and the stationing knob's axis was accepted in both horizontal directions. Trials were canceled when the animal was not in the correct position. To enhance the contrast of the dark dorsum of the porpoise and the dark background, a stripe of zinc ointment was sometimes put on the porpoise's back between the blowhole and the dorsal fin.

Three seconds after the animal was stationed properly, a signal was produced from one of the 16 transducers. The animal reacted in one of three ways:

1) A hit. The animal swam to the transducer that had been active and touched the float above it. A bridge tone was whistled, and the animal returned to the start buoy where it received a fish reward.
2) A miss. The animal swam towards a transducer that had not been active and touched the float above it. The trainer called the animal back to the start buoy. No reward was given.
3) No choice. The animal swam away from the station to the trainer without going to any float. No reward was given.

The operator and trainer communicated via headset radios. The trainer did not know which transducer had been activated, and heard from the operator whether the animal's response was correct or incorrect. The operator recorded the responses of the animal (i.e., the transducer indicated by the animal, or a no-choice response).

During a session the signal parameters (frequency, amplitude, and signal duration) were kept constant. The only variable that changed within a session was the transducer which was activated. Twenty-four trials were conducted per session. Each transducer was activated three times, in random order, over two consecutive sessions (in a session each transducer was activated at least once and at most twice). In each session two catch trials occurred (the first trial and one about halfway through) in which the animal was rewarded if it remained at the station in the correct position and did not make contact with any of the sources (no signal was broadcasted). A whistle was blown when the animal was in the correct position at the station, after which he returned to the trainer and received a fish.

To check for potential cues caused by the operator and the equipment settings, the equipment was set, the transducer connector was detached and a trial was run. This was done each week. The animal did not react to the activation of the equipment and responded as if it was a signal-absent trial.

Daily, three to five sessions (each lasting about 20 min) were conducted, each testing one of the three frequencies (16, 64, and 100 kHz) in random order, and in 1999 the three signal durations (600, 900, and 1000 ms) of the 64 kHz signals were tested in random order as well. Data collection sessions were conducted in September and October 1999 and between June and October 2000.

III. RESULTS

During the study, no sessions were canceled due to uncooperative behavior by the animal. After the signal onset, the animal took the shortest route between the station and the transducer he selected. This trip took on average 4.5 s (around 4 s to reach transducers in front of the animal and around 5 s for transducers behind him). After touching the float above the selected transducer, the porpoise always swam directly towards the trainer.

The animal's responses per direction relative to its body axis in the 2000 study season, when pointing east and west, did not differ statistically (Kendall Rank Correlation Test, $p < 0.05$). The average correct response rates (over 16 directions) were similar for all three test frequencies. Therefore the correct responses in the same directions relative to the body length axis when the animal was pointing east and west were averaged. This increased (usually doubled) the sample size per direction relative to the body length axis.

A. Influence of signal duration

In the 1999 study season, 64 kHz signals were offered at one RL (82 dB re 1 μPa, 12 dB above the calculated MDT), but with three signal durations (600, 900 and 1000 ms). The animal's average correct response rate (over all 16 directions) increased with increasing signal duration (Figs. 3(a)–3(c)). The animal showed a clear asymmetry in its localization ability; it was better at determining the location of sound sources in front of it and on its left than those behind it and on its right. When the total signal duration was only 600 ms, the animal's average correct response level was very low (29%). During those sessions the animal became visibly frustrated due to its low success rate. Therefore only five sessions were conducted with this short signal duration.

B. Influence of signal level

The 16 kHz signal was offered at one signal duration (1000 ms), but at three RLs. The average RLs were 86, 89 and 94 dB re 1 μPa, rms, respectively 7, 10 and 15 dB above MDT (Figs. 4(a)–4(c)). The animal's average correct response rate (over all 16 directions) increased as the RL increased. At average RLs of 86 and 89 dB re 1 μPa, the animal's ability to locate the sound sources was asymmetrical around the length axis of its body (Figs. 4(a) and 4(b)), but at an average RL of 94 dB re 1 μPa, the porpoise's performance became more symmetrical (Fig. 4(c)).

C. Influence of signal frequency

The porpoise's sound source localization abilities for the 16 kHz signals (at an average RL of 94 dB re 1 μPa, 15 dB above MDT) and 64 kHz signals (at an average RL of 80 dB re 1 μPa, 10 dB above MDT) were similar (Figs. 5(a) and 5(b)). The porpoise's performance was symmetrical around its body length axis. For those signals the animal had a very high performance level (84–99%) in directions up to 124° to his left and right, but performance was less for sound sources

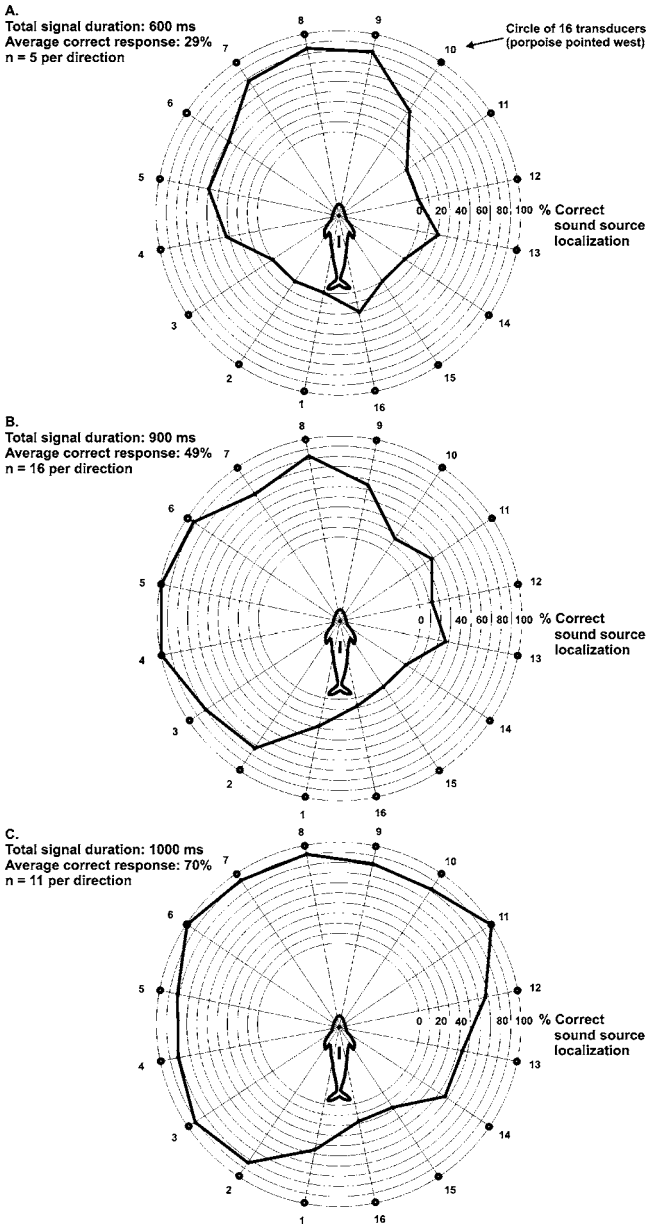


FIG. 3. Effect of signal duration on localization performance. The porpoise's average correct response levels per transducer for 64 kHz signals with the same average RL (82 dB) and RL/MDT ratio of 12 dB, but with total signal durations of 600 ms (A), 900 ms (B), and 1000 ms (C). Year: 1999. The animal pointed to the west.

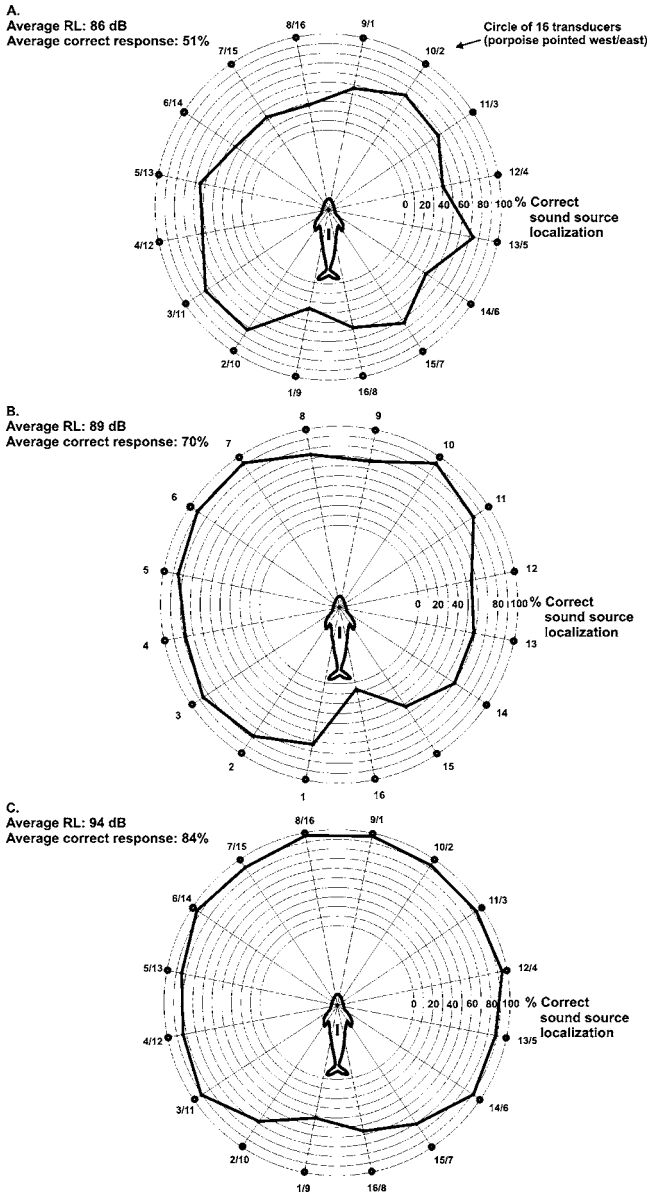


FIG. 4. Effect of average RL on localization performance. The porpoise's average correct response levels per transducer for 16 kHz, 1000 ms test signals with RL of 86 dB and RL/MDT ratio of 7 dB (A), with RL of 89 dB and RL/MDT ratio of 10 dB (B), and with RL of 94 dB and RL/MDT ratio of 15 dB (C). Year: 2000. The animal pointed to the east and west. N=44 per direction.

directly behind him. The animal's average correct response rate for 100 kHz signals (at an average RL of 69 dB re 1 μPa, only 3 dB above MDT) was slightly lower (79%; Fig. 5(c)). The level of the 100 kHz signal was limited by the characteristics of the available transducers.

D. Incorrect responses

Incorrect responses consisted either of misses or no choices. In misses, the porpoise usually chose a transducer next to the activated one, and only in a few cases, two or three transducers from the activated transducer (Fig. 6). The numbers of no-choice responses for the 16 and 64 kHz signals were lower than for the 100 kHz signals, but most occurred after activation of transducers caudal of the animal (Fig. 7).

IV. DISCUSSION

A. Evaluation of the data

Some caution is needed when interpreting the data from the present study. Only one animal was available, so it may be risky to extrapolate the findings of the present study to all members of the species. Age, sex, location, experience, health history and many other factors may influence the hearing and behavior of individuals.

The sound localization experiment in the present study consisted of two phases: detection and localization. Signal

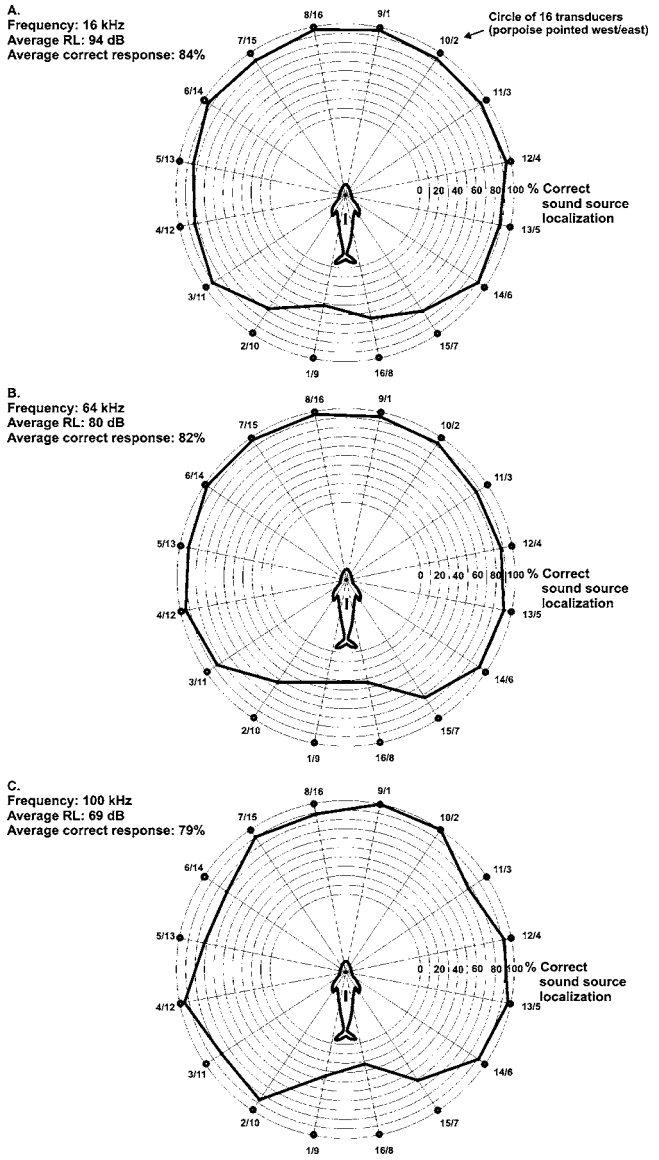


FIG. 5. Effect of signal frequency on localization performance. The porpoise's average correct response levels per transducer for 16 kHz test signals at RL of 94 dB and RL/MDT ratio of 15 dB (A), 64 kHz test signals at RL of 80 dB and RL/MDT ratio of 10 dB (B), and for 100 kHz test signals at RL of 69 dB and RL/MDT ratio of 3 dB (C). All signals had a total duration of 1000 ms. Year: 2000. The animal pointed to the east and west. N=44 per direction.

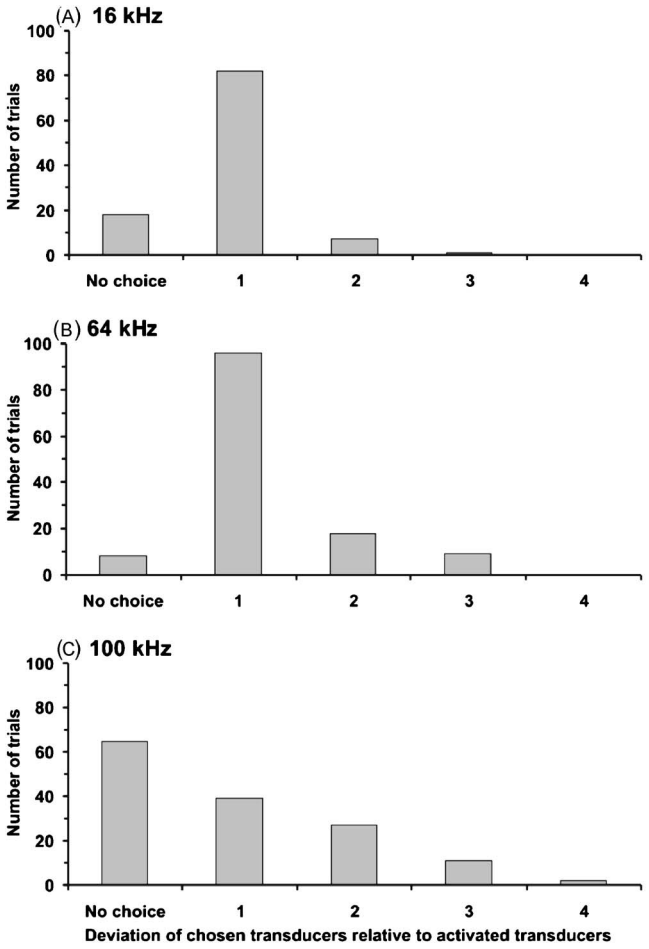


FIG. 6. The number of trials in which the porpoise made an incorrect choice and the deviation of the chosen transducer relative to the activated transducer, and the number of trials the animal made no choice, for 16 kHz signals (A), 64 kHz signals (B) and 100 kHz signals (C). All signals had a duration of 1000 ms (same trials that Fig. 5 is based on).

detection depends on signal properties (duration, spectrum, the RL), the hearing of the animal (sensitivity, directionality, critical ratio, hearing integration time), and the background noise (causing a masked detection threshold). In addition, the attentiveness of an animal for sounds at the moment a signal is being produced is of importance.

The RL from each of the 16 transducers was slightly different. However, it is unlikely that the animal used intensity discrimination to select the activated transducer, because the signal levels were much above the ambient noise level, and because the ambient noise level fluctuated (causing fluctuation in RL/MDTs). The results also suggest that the animal did not use intensity discrimination, as it was not better at pointing out the transducers producing RLs above average than transducers producing below average RLs.

Two previous studies have investigated the sound source localization abilities of harbor porpoises in the horizontal plane, but using very different methods than the present study. Andersen (1970) and Dudok van Heel (1959) attempted to determine the minimum audible angle (MAA) in harbor porpoises. Two potential sound sources were offered in front of the animals. The animals had to choose from which transducer a low-frequency sound came from. Dudok van Heel (1959) tested pure tones with durations of 660 ms, produced by two transducers 18 m in front of the animal. For 6 kHz tones the MAA was 16° and for 3.5 kHz tones it was 22° (the Source Level or RL were not reported). Andersen (1970) tested a 2 kHz pure tone with a signal duration of 500 ms, produced by two transducers 8 m in front of the animal. The RL was 120 dB re 1 μPa. The 50% correct MAA was asymmetrical and was around 5° on the left quadrant, and 0.6° on the right quadrant. In the present study the porpoise also showed an asymmetry in sound source localization, especially when the signal duration was short or the RL low. Both Andersen (1970) and Dudok van Heel (1959) used short signal durations (660 and 500 ms), thus, based on the findings in the present study, making it harder for the animal to localize the sound source and increasing the asym-

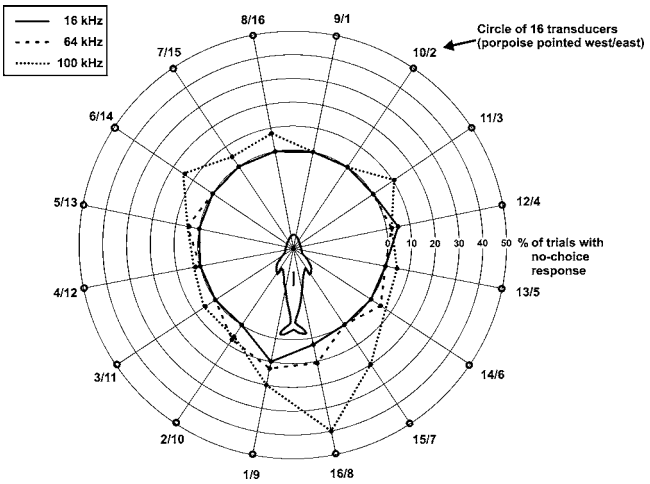


FIG. 7. The average percentage of trials (when the porpoise pointed to the east and to the west) per direction in which the porpoise selected no transducer after a sound signal, but swam immediately towards the trainer (no-choice response). The data come from same trials that Figs. 5 and 6 are based on.

metry in the responses. In contrast to the MAA studies of Andersen (1970) and Dudok van Heel (1959), in the present study the porpoise could not focus its attention in one general direction (as it could focus on the front in the MAA studies), but was completely unaware of the direction from which the sound would come. This probably made the task more difficult.

B. Signal parameters

1. Signal duration

The animal's correct response level was considerably higher for 900 and 1000 ms signal durations than for 600 ms durations (64 kHz). In bottlenose dolphins, the effect of pulse duration on the animal's hearing threshold is frequency dependent, but generally thresholds were influenced when signals were shorter than ~500 ms (Johnson, 1968a). Assuming that the same is true for harbor porpoises, the low localization performance with 600 ms signals might have been related to the shorter time the animal could focus its attention on the direction the sound came from. The porpoise probably heard the signal clearly, but needed more time to pinpoint the sound source accurately.

2. Signal to noise ratio

If the RL/MDT was less than 10 dB, there appeared to be a reduction in the correct response percentage (Fig. 4). For 16 kHz, three values for the relationship between correct response percentage and RL/MDT ratio have been determined. Assuming a linear relationship between these observations, extrapolation of this relationship implies that for 1000 ms, 16 kHz signals, a 100% score would be reached above a RL/MDT ratio of 18 dB (though at other frequencies this relationship may differ). Figure 5 also illustrates this trend (RL/MDT 15, 10 and 3 dB; 84, 82 and 79 correct response %), although frequency is a confounding effect.

The relatively high percentage of no-choice responses to sound coming from the caudal direction (Fig. 7) is probably due to the porpoise's hearing directivity pattern; a relatively low hearing sensitivity in the caudal direction causes a low RL/MDT ratio. Directional hearing data, derived from the same animal in the same environment (Kastelein *et al.*, 2005), show that the "fore-aft ratio" (=difference between the threshold level in the frontal direction and the threshold level in the caudal direction) at 16 kHz is around 10 dB, and at 64 and 100 kHz around 15 dB.

C. Significance for ecology and alarm signal design

The present study was designed to test how well a harbor porpoise could determine the location from which single short-duration sound signals emanated, without turning its head. In a natural setting, if sounds are of longer duration, or if multiple sounds come from one source in succession, harbor porpoises could vary the orientation of their heads or their position relative to the sound source (Richardson *et al.* 1995) to capitalize on binaural cues and therefore increase the chance of localizing the sound (Kastelein *et al.*, 2005). Some commercial acoustic alarms for gillnets ("pingers") emit sound signals in succession and it seems likely that wild harbor porpoises would utilize the cues they have at their disposal. The situation described in the present study could occur if acoustic alarms on nets were triggered by the echolocation signals of nearby porpoises (i.e., in interactive pingers, presently being developed). There is a danger that interactive pingers will be triggered too late for a porpoise to avoid a collision with a net (Kastelein *et al.*, 2000). In the case of an alarm sound it is of great importance that the porpoises can immediately localize the sound source, identify the geometry of the net, and determine the correct flight direction. Based on the results of the present study, the signal duration of echolocation-activated pingers should be $\geqslant$1 s. The random timing of pingers to avoid habituation is another point for discussion. If pingers had a synchronized firing rate, animals would have an increased ability to detect the rigging and shape of a gillnet. The present study showed that localization was efficient over a broad frequency range, so random frequencies (Kastelein *et al.*, 2001, 2007) can be used in pingers to reduce habituation.

The present study suggests that given sufficient signal duration ($\geqslant$1 s) of an acoustic alarm and sufficient RL, porpoises can localize the sources of 16, 64 and 100 kHz sounds with similar accuracy. Therefore, as far as localization of alarms is concerned, frequency, within the range tested in the present study, seems to be unimportant when designing the optimal acoustic alarm. Attention should be focused on other design parameters, such as the effects of pinger sounds on other marine fauna (such as fish, other odontocetes, and pinnipeds), deterrent effect of sounds, and battery life (determined by the duty cycle, amplitude, and frequency spectrum).

Future work to determine which signal durations and RLs result in 100% correct performance for all directions would aid in more effective acoustic alarm design.

ACKNOWLEDGMENTS

We thank Niek Straver (Labforce BV, The Netherlands) for making the transducers, John Nijsse and Helmi ter Horst for helping with the data collection and Rob Triesscheijn for drawing the graphs. We thank Nancy Jennings (dotmoth.co.uk), Kate Evans (University of Bristol, UK), Bill T. Ellison (Marine Acoustics Inc., USA), Alexander Supin (Institute of Ecology and Evolution, Moscow, Russia), Dorian Houser (Biomimetica, San Diego, USA) and two anonymous reviewers for their constructive comments on the manuscript. Funding for this project was obtained from The North Sea Directorate (DNZ, through Wanda Zevenboom, Contract Nos. 76/317581, IBO 14.1; 1999 and 76/318701, IBO 4.2; 2000) of the Directorate-General of the Netherlands Ministry of Transport, Public Works and Water Management (RWS), and the Netherlands Ministry of Agriculture, Nature and Food Quality (Fisheries Directory, through Edwin Meeuwsen, Contract no. TRCDK-DH/07/1049). The porpoise's training and testing were conducted under authorization of the Netherlands Ministry of Agriculture, Nature Management and Fisheries, Department of Nature Management. Endangered Species Permit FEF27 06/2/98/0184.

Andersen, S. (**1970**). "Directional Hearing in the Harbour Porpoise *Phocoena phocoena*," in *Investigations on Cetacea*, edited by G. Pilleri (The Brain Anatomy Institute, University of Bern), Vol. **2**, pp. 260–263.

Anonymous (**2000**). "Annex I, Report of the Sub-Committee on Small Cetaceans," J. Cetacean Res. Manage. **2**, 235–245.

Atema, J., Fay, R. R., Popper, A. N., and Tavolga, W. N. (**1988**). *Sensory Biology of Aquatic Animals* (Springer-Verlag, New York), p. 936.

Bravington, M. V., and Bisack, K. D. (**1996**). "Estimates of harbour porpoise bycatch in the Gulf of Maine sink gillnet fishery, 1990–1993," Rep. Int. Whal. Comm. **46**, 567–574.

Cox, T. M., Read, A. J., Solow, A., and Tregenza, N. (**2001**). "Will harbour porpoises (*Phocoena phocoena*) habituate to pingers?" J. Cetacean Res. Manage. **3**, 81–86.

Culik, B. M., Koschinski, S., Tregenza, N., and Ellis, G. (**2001**). "Reactions of harbour porpoises (*Phocoena phocoena*) and herring (*Clupea harengus*) to acoustic alarms," Mar. Ecol.: Prog. Ser. **211**, 255–260.

Dudok van Heel, W. H. (**1959**). "Audio-direction finding in the porpoise (*Phocaena phocaena*)," Nature (London) **183**, 1063.

Dudok van Heel, W. H. (**1962**). "Sound and cetaceans," Netherlands J. Sea Res. **1**(4), 407–507.

Fleischer, G. (**1980**). "Low-frequency receiver of the middle ear in mysticetes and odontocetes," in *Animal Sonar Systems*, edited by R.-G. Busnel and J. F. Fish (Plenum, New York), pp. 891–893.

Gearin, P. J., Gosho, M. E., Laake, J., Cooke, L., DeLong, R. L., and Hughes, K. M. (**2000**). "Experimental testing of acoustic alarms (pingers) to reduce bycatch of harbour porpoise, *Phocoena phocoena*, in the state of Washington," J. Cetacean Res. Manage. **2**, 1–9.

Houser, D. S., Finneran, J., Carder, D., Van Bonn, W., Smith, C., Hoh, C., Mattrey, R., and Ridgway, S. (**2004**). "Structural and functional imaging of bottlenose dolphin (*Tursiops truncatus*) cranial anatomy," J. Exp. Biol. **207**, 3657–3665.

Jefferson, T. A., and Curry, B. E. (**1994**). "A global review of porpoise (*Cetacea: phocoenidae*) mortality in gillnets," Biological Conserv. **67**, 167–183.

Johnson, S. C. (**1968a**). "Relation between absolute threshold and duration of tone pulse in the bottlenoses porpoise," J. Acoust. Soc. Am. **43**, 757–763.

Johnson (**1968b**). "Masked tonal thresholds in the bottlenosed porpoise." J. Acoust. Soc. Am. **44**, 965–967.

Kastelein, R. A., Au, W. W. L., and Haan, D. de (**2000**). "Detection distances of bottom-set gillnets by harbor porpoises (*Phocoena phocoena*) and bottlenose *dolphins* (*Tursiops truncatus*)," **49**, 359–375.

Kastelein, R. A., de Haan, D., Goodson, A. D., Staal, C., and Vaughan, N. (**1997**). "The effects of various sounds on a harbour porpoise (*Phocoena phocoena*)," in *The Biology of the Harbour Porpoise*, edited by A. J. Read, P. R. Wiepkema, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp. 367–383.

Kastelein, R. A., Goodson, A. D., Lien, J., and de Haan, D. (**1995**). "The effects of acoustic alarms on harbour porpoise (*Phocoena phocoena*) behaviour," in *Harbour Porpoises, Laboratory Studies to Reduce Bycatch*, edited by P. E. Nachtigall, J. Lien, W. W. L. Au, and A. J. Read (De Spil, Woerden, The Netherlands), pp. 157–167.

Kastelein, R. A., Rippe, H. T., Vaughan, N., Schooneman, N. M., Verboom, W. C., and de Haan, D. (**2000**). "The effects of acoustic alarms on the behavior of harbor porpoises (*Phocoena phocoena*) in a floating pen," Marine Mammal Sci. **16**, 46–64.

Kastelein, R. A., de Haan, D., Vaughan, N., Staal, C., and Schooneman, N. M. (**2001**). "The influence of three acoustic alarms on the behavior of harbour porpoises (*Phocoena phocoena*) in a floating pen," **52**, 351–371.

Kastelein, R. A., Bunskoek, P., Hagedoorn, M., Au, W. W. L., and de Haan, D., (**2002**). "Audiogram of a harbor porpoise (*Phocoena phocoena*) measured with narrow-band frequency-modulated signals," J. Acoust. Soc. Am. **112**, 334–344.

Kastelein, R. A., Janssen, J., Verboom, W. C., and de Haan, D. (**2005**). "Receiving beam patterns in the horizontal plane of a harbor porpoise (*Phocoena phocoena*)," J. Acoust. Soc. Am. **118**, 1172–1179.

Kastelein, R. A., van der Heul, S., van der Veen, J., Verboom, W. C., Jennings, N., de Haan, D., and Reijnders, P. (**2007**). "Effects of commercially-available acoustic alarms, designed to reduce small cetacean bycatch in gillnet fisheries, on the behaviour of North Sea fish species in a large tank," Mar. Environ. Res. (in press).

Kraus, S., Read, A. J., Solow, A., Baldwin, K., Spradlin, T., Williamson, J., and Anderson, E. (**1997**). "Acoustic alarms reduce porpoise mortality," Nature (London) **388**, 525.

Laake, J., Rugh, D., and Baraff, L. (**1998**) *Observations of Harbor Porpoise in the Vicinity of Acoustic Alarms on a Set Gill Net*. NOAA Technical memorandum NMFS-AFSC-84. U.S. Dept. of Commerce. 40 pp.

Lien, J., Hood, C., Pittman, D., Ruel, P., Borggaard, D., Chisholm, C., Wiesner, L., Mahon, T., and Mitchell, D. (**1995**). "Field tests of acoustic devices on groundfish gillnets: Assessment of effectiveness in reducing harbour porpoise by-catch," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp. 349–364.

Lowry, N., and Teilmann, J. (1994). "Harbour porpoise (*Phocoena phocoena*) in Danish waters. Status, bycatch and possibilities for bycatch reduction," in Report of the Danish Institute for Fisheries Technology and Aquaculture. North Sea Centre, P.O. Box 59, DK-9850 Hirtshals, Denmark, 46 pp.

Møhl, B., and Andersen, S. (**1973**). "Echolocation: High-frequency component in the click of the harbour porpoise (*Phocoena ph. L.*)," J. Acoust. Soc. Am. **53**, 1368–1372.

Oelschläger, H. A. (**1986a**). "Comparative morphology and evolution of the otic region in toothed whales (Cetacea, Mammalia)," Am. J. Anat. **177**, 353–368.

Oelschläger, H. A. (**1986b**). "Tympanohyal bone in toothed whales and the formation of the tympano-periotic complex (Mammalia: Cetacea)," J. Morphol. **188**, 157–165.

Olesiuk, P. F., Nichol, L. M., Sowden, M. J., and Ford, J. K. B. (**2002**). "Effect of the sound generated by an acoustic harassment device on the relative abundance and distribution of harbor porpoises (*Phocoena phocoena*) in retreat passage, British Columbia," Marine Mammal Sci. **18**, 843–862.

Palka, D. L., Read, A. J., Westgate, A. J., and Johnston, D. W. (**1996**). "Summary of current knowledge of harbour porpoises in US and Canadian Atlantic waters," Rep. Int. Whal. Comm. **46**, 559–565.

Popov, V. V., and Supin, A. Ya. (**1992**) "Electrophysiological study of the interaural intensity difference and interaural time-delay in dolphins," in *Marine Mammal Sensory Systems*, edited by J. Thomas, R. A. Kastelein, and A. Ya. Supin (Plenum, New York), pp. 257–267.

Popov, V. V., Supin, A. Ya., Klishin, V. O., and Bulgakova, T. N. (**2006**). "Monaural and binaural hearing directivity in the bottlenose dolphin: Evoked-potential study," J. Acoust. Soc. Am. **119**, 636–644.

Read, A. J., and Gaskin, D. E. (**1988**). "Incidental catch of harbor porpoises by gill nets." J. Wild. Man. **52**, 517–523.

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).

Supin, A. Ya., and Popov, V. V. (**1993**). "Direction-dependent spectral sen-

sitivity and interaural spectral difference in a dolphin: Evoked potential study," J. Acoust. Soc. Am. **93**, 3490–3495.
Teilmann, J., Tougaard, J., Miller, L. A., Kirketerp, T., Hansen, K., and Brando, S. (**2006**). "Reactions of captive harbor porpoises (*Phocoena phocoena*) to pinger-like sounds," Marine Mammal Sci. **22**, 240–260.
Trippel, E. A., Wang, J. Y., Strong, M. B., Carter, L. S., and Conway, J. D. (**1996**). "Incidental mortality of harbour porpoise (*Phocoena phocoena*) by the gill-net fishery in the lower Bay of Fundy," Can. J. Fish. Aquat. Sci. **53**, 1294–3000.
Trippel, E. A., Strong, M. B., Terhune, J. M., and Conway, J. D. (**1999**). "Mitigation of harbour porpoise (*Phocoena phocoena*) by-catch in the gillnet fishery in the lower Bay of Fundy," Can. J. Fish. Aquat. Sci. **56**, 113–123.
Verboom, W. C., and Kastelein, R. A. (**1995**). "Acoustic signals by Harbour porpoises (*Phocoena phocoena*)," in *Harbour Porpoises, Laboratory Studies to Reduce Bycatch*, edited by P. E. Nachtigall, J. Lien, W. W. L. Au., and A. J. Read (De Spil, Woerden, The Netherlands), pp. 1–40.
Verboom, W. C., and Kastelein, R. A. (**1997**). "Structure of harbour porpoise (*Phocoena phocoena*) click train signals," in *The Biology of the Harbour Porpoise*, edited by A. J. Read, P. R. Wiepkema, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp. 343–362.
Verboom, W. C., and Kastelein, R. A. (**2003**). "Structure of harbour porpoise (*Phocoena phocoena*) acoustic signals with high repetition rates," in *Echolocation in Bats and Dolphins*, edited by J. A. Thomas, C. Moss, and M. Vater (University of Chicago Press, Chicago), pp. 40–43.
Voronov, V. A., and Stosman, I. M. (**1983**). "On sound perception in the dolphin *Phocoena phocoena*," Zh. Evol. Biokhim Fiziol **18**, 352–357(Transl. from Zh. Evol. Biokhim. Fiziol. **18**, 499–506 **1982**).

Assessing temporary threshold shift in a bottlenose dolphin (*Tursiops truncatus*) using multiple simultaneous auditory evoked potentials

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

Carolyn E. Schlundt
EDO Professional Services, 3276 Rosecrans Street, San Diego, California 92110

Brian Branstetter
U.S. Navy Marine Mammal Program, Space and Naval Warfare Systems Center, San Diego, Code 2351, 53560 Hull Street, San Diego, California 92152

Randall L. Dear
Science Applications International Corporation, 4065 Hancock Street, San Diego, California 92110

(Received 2 April 2007; revised 15 May 2007; accepted 21 May 2007)

Hearing sensitivity was measured in a bottlenose dolphin before and after exposure to an intense 20-kHz fatiguing tone in three different experiments. In each experiment, hearing was characterized using both the auditory steady-state response (ASSR) and behavioral methods. In experiments 1 and 2, ASSR stimuli consisted of seven frequency-modulated tones, each with a unique carrier and modulation frequency. The tones were simultaneously presented to the subject and the ASSR at each modulation rate measured to determine the effects of the sound exposure at the correspond ing carrier frequency. In experiment 3 behavioral thresholds and ASSR input-output functions were measured at a single frequency before and after three exposures. Hearing loss was frequency-dependent, with the largest temporary threshold shifts occurring (in order) at 30, 40, and 20 kHz. ASSR threshold shifts reached 40–45 dB and were always larger than behavioral shifts (19–33 dB). The ASSR input-output functions were represented as the sum of two processes: a low threshold, saturating process and a higher threshold, linear process, that react and recover to fatigue at different rates. The loss of the near-threshold saturating process after exposure may explain the discrepancies between the ASSR and behavioral threshold shifts. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2749447]

PACS number(s): 43.80.Lb, 43.80.Nd [WWA] Pages: 1249–1264

I. INTRODUCTION

Auditory evoked potentials (AEPs) are small voltages automatically generated by the brain in response to acoustic stimuli. Because AEP measurements do not require the specific training needed for traditional behavioral tests, AEPs are becoming increasingly popular for investigating auditory functions in marine mammals (Ridgway *et al.*, 1981; Popov *et al.*, 1992; Dolphin *et al.*, 1995; Supin and Popov, 1995; Dolphin, 1996; Szymanski *et al.*, 1999; Nachtigall *et al.*, 2005; Popov *et al.*, 2005; Yuen *et al.*, 2005; Cook *et al.*, 2006; Mooney *et al.*, 2006). AEPs may be generated in response to various sound stimuli; however, for frequency-specific auditory testing in marine mammals, the use of sinusoidal amplitude-modulated (SAM) tones has been common (e.g., Supin *et al.*, 2001; Nachtigall *et al.*, 2005; Popov *et al.*, 2005; Yuen *et al.*, 2005; Finneran and Houser, 2006). SAM tones generate a particular type of AEP, called the auditory steady-state response (ASSR) or envelope following response (EFR), formed when stimuli are presented at a sufficient rate that individual transient responses overlap and form a steady-state signal (Stapells *et al.*, 1984; Supin *et al.*, 2001). The term ASSR is used here to remain consistent with the large body of human literature (reviewed by Picton, 2007). The primary characteristic of the ASSR is a fundamental frequency at the SAM tone modulation frequency; thus, ASSRs may be analyzed in the frequency domain and statistical techniques may be applied to objectively determine the presence or absence of the evoked response (Dobie and Wilson, 1989, 1996).

Although the ASSR is normally used to test a single stimulus frequency, multiple SAM tones, each with a unique modulation frequency, may be used to simultaneously test hearing at multiple frequencies (Picton *et al.*, 1987; Regan and Regan, 1988; Lins *et al.*, 1995; Lins and Picton, 1995; Dolphin, 1996; Popov *et al.*, 1997, 1998; Finneran and Houser, 2007). The evoked response to each SAM tone occurs at the corresponding modulation rate and, if sufficient frequency separation exists, significant interactions do not occur and the hearing thresholds estimated with the multiple SAM tones match those obtained with single stimuli (Lins and Picton, 1995; John *et al.*, 1998, 2002). Although the use of the multiple ASSR technique does not result in a one-to-one increase in testing speed, multiple ASSR measurements

improve testing speed compared to measurements with single SAM stimuli (John *et al.*, 2002). This makes the use of the multiple ASSR technique particularly attractive for use in marine mammals, where access to subjects is often limited.

The increased speed of the multiple ASSR technique also makes it a potentially useful approach for assessing frequency-dependent patterns of temporary hearing loss in animals, since hearing thresholds at multiple frequencies could be measured in a relatively short amount of time. To date, measurements of temporary hearing loss, or temporary threshold shift (TTS), in marine mammals have used both behavioral (Kastak *et al.*, 1999; Finneran *et al.*, 2000; Schlundt *et al.*, 2000; Finneran *et al.*, 2002, 2005) and AEP approaches (Nachtigall *et al.*, 2004) but have tested single frequencies. To provide meaningful comparisons of the effects of an exposure on the hearing ability at multiple frequencies, exposures have been repeated on subsequent days, so each frequency can be tested at equivalent times after the exposure (e.g., Schlundt *et al.*, 2000). Comparing the effects at different frequencies has therefore required the subjects to be exposed multiple times to a potentially aversive stimulus and has introduced potential confounding effects caused by day-to-day variability within the subject and acoustic environment. The use of the multiple ASSR technique could reduce these problems, since hearing could be simultaneously tested at multiple frequencies immediately after exposure.

This paper describes three experiments designed to evaluate the feasibility of assessing TTS in a bottlenose dolphin using the multiple ASSR technique. In the first experiment, AEP hearing thresholds at seven frequencies were simultaneously measured before and after the subject was exposed to an intense, single-frequency tone to determine the effects of the sound exposure at nearby frequencies. Although the hearing test signals possessed relatively large frequency bandwidth and insufficient frequency spacing to prevent interactions, thresholds measured before the exposure agreed closely with thresholds previously obtained behaviorally. The postexposure data revealed clear patterns of frequency-dependent hearing loss and some of the largest amounts of TTS yet observed in marine mammals. Experiment 2 replicated the first experiment with the addition of behavioral threshold measurements before and after the exposure at all seven frequencies of interest. Discrepancies between the behavioral and AEP results led to a third experiment in which ASSR and behavioral measurements were conducted at a single frequency before and after a series of exposures to investigate the effects of the exposure on the resulting thresholds and the ASSR input-output (I/O) functions, which describe the relationship between ASSR amplitude and the stimulus sound-pressure level (SPL).

II. METHODS

A. Subject

The subject (BLU) was a female bottlenose dolphin (41 years, 200 kg) with experience in cooperative psychophysical tasks, including auditory detection tasks and TTS experiments. The subject's daily food intake was a 23-lb mixture of capelin, squid, and herring, of which about 80% was consumed during the test sessions. Prior testing revealed significant hearing loss above 40–50 kHz (Finneran and Houser, 2006; Finneran *et al.*, 2007a; Finneran and Schlundt, 2007), which is not uncommon for dolphins of this age (Houser and Finneran, 2006b). The subject was housed in floating netted enclosures (9×9 to 12×24 m) located in San Diego Bay, California and trained to slide out of the water onto a foam mat for transport to the test pool (see below). The subject was transported from the test pool back to the San Diego Bay enclosures after each test session.

B. Experimental apparatus

Experiments were conducted in a vinyl-walled, aboveground pool approximately $3.7\times6\times1.5$ m and filled with seawater. The pool was located within a larger room whose walls and ceiling were treated with sound-absorbing foam to reduce reverberation. Ambient noise spectral density levels in the pool were less than 40 dB $re:1\ \mu\text{Pa}^2/\text{Hz}$ for frequencies above 3 kHz (Finneran *et al.*, 2007a). A wooden deck located at one end of the pool supported a submerged frame containing an underwater sound projector and a neoprene-covered, plastic "biteplate" upon which the subject was trained to position. For AEP measurements the biteplate and projector depths were approximately 14 cm—shallow enough that the top of the subject's head and blowhole remained above water while positioned on the biteplate. For the majority of behavioral tests (see Test sequence) the biteplate and projector were located at mid-depth.

C. Fatiguing stimuli

AEP and behavioral hearing tests were conducted before and after the subject was exposed to a fatiguing sound intended to induce TTS. Table I summarizes the fatiguing sound exposures for each experiment. The fatiguing sound was a 20-kHz tone with rise/fall times of 100 ms for all experiments. Experiments 1 and 2 featured a single exposure with 64-s duration. During experiment 3, three 16-s exposures were conducted, separated by 11 and 13 min.

The fatiguing tone was digitally generated, converted to analog using a multifunction data acquisition board (National Instruments NI PCI-6070E, 12-bit resolution, 200-kHz update rate), filtered (SRS SR–560), amplified (Crest 10001), and presented via a spherical piezoelectric sound projector (ITC 1001). The subject wore two calibrated hydrophones (B&K 8105) mounted with suction cups near the left and right external auditory meatuses. The hydrophone signals were amplified and filtered (B&K 2635), digitized (PCI-6070E), and analyzed in the frequency domain to estimate the received (rms) SPL during each exposure (Table I). The average SPL from the two hydrophones over the 64-s duration was 186 and 185 dB $re:1\ \mu\text{Pa}$ for experiments 1 and 2, respectively. The corresponding sound-exposure levels (SELs) were 204 and 203 dB $re:1\ \mu\text{Pa}^2\,\text{s}$, respectively. For experiment 3, the mean SPL from the three exposures was 193 dB $re:1\ \mu\text{Pa}$ (SD=0.8 dB) and the total SEL (accumulated over all three exposures) was 210 dB $re:1\ \mu\text{Pa}^2\,\text{s}$.

TABLE I. Fatiguing sound properties for experiments 1–3.

Experiment	Frequency (kHz)	SPL (dB *re*: 1 μPa)	Duration (s)	SEL (dB *re*: 1 μPa^2 s)
1	20	186	64	204
2	20	185	64	203
3	20	193	16	205
		194	16	206
		193	16	205
		Mean=193 (0.8 SD)		Cumulative=210

D. AEP measurements

AEP measurements were based on the detection of ASSRs to sinusoidal frequency-modulated (SFM) tones. Although SAM tones are normally used for ASSR stimuli, the small volume of the test pool and resulting complex standing wave patterns made the use of SAM tones problematic—small changes in subject position could result in large changes to the received sound waveform and AM modulation depth, which would affect the resulting ASSR amplitude (Supin and Popov, 1995). Prior measurements in the test pool showed dramatic improvement to the sound field by using frequency-modulated (FM) stimuli (Finneran and Schlundt, 2007), which create a more uniform overall sound pressure since the effects of multipath interference tend to cancel over a number of frequencies. Supin and Popov (2000) previously demonstrated that FM stimuli would produce the ASSR in dolphins. Finneran *et al.* (2007a) previously achieved good results in the same test pool as the present study using combination AM/FM stimuli; however, evoked responses to these stimuli featured multiple sidebands related to the FM rate, making it very difficult to separate responses to multiple stimuli. For these reasons, SFM stimuli were used in this study, rather than SAM or combination AM/FM stimuli. Stimuli were continuously presented to the subject, with the ASSR measurements ignoring the first 61 ms after stimulus onset. This allowed the direct sound and reflected waves to reach steady-state conditions and the initial transient evoked response to be avoided.

During experiments 1 and 2, ASSR thresholds were measured in response to seven simultaneous SFM tones with carrier (and approximate modulation) frequencies of 10 (0.95), 20 (1.0), 30 (1.05), 40 (1.1), 50 (1.15), 60 (1.2), and 70 (1.25) kHz. Modulation rates were chosen to be near 1 kHz, which has been shown to be a good rate for production of the ASSR in dolphins (Dolphin *et al.*, 1995; Supin and Popov, 1995; Finneran *et al.*, 2007b), yet allow individual ASSRs to be identified using a 16.4-Hz frequency resolution (Finneran and Houser, 2007). The exact modulation rates were adjusted slightly to ensure an integral number of cycles of the modulating waveform within the 61-ms epochs used for analysis (e.g., 950 Hz was adjusted to 950.82 Hz). Threshold measurements were also made during experiment 2 using single SFM tones with carrier and modulation frequencies of 30 and 1.05 kHz, respectively. During experiment 3, ASSR amplitude was measured as a function of stimulus SPL to create I/O functions. The I/O function measurements used single SFM tones with carrier and modulation frequencies of 30 and 1.0 kHz, respectively.

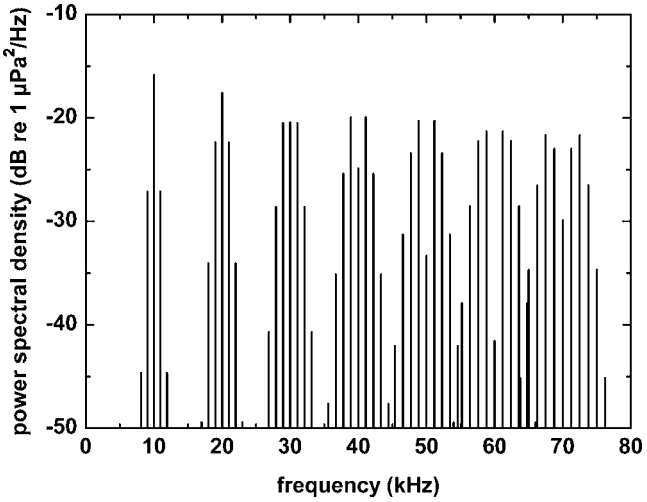


FIG. 1. Power spectral densities of the SFM tone components used for ASSR threshold measurements. The spectra are from signals with peak amplitudes of 1 V.

The SFM tone instantaneous voltage, $v(t)$, was described by

$$v(t) = \sum_{n=1}^{N} A_n \sin\left[2\pi\left(F_n t - \frac{DF_n}{2f_n}\right)\cos(2\pi f_n t)\right], \quad (1)$$

where N is the number of waveform components ($N=1$ or 7), n indicates the nth SFM component, A_n is the amplitude, F_n is the carrier frequency, f_n is the modulation frequency, and D is the modulation depth (the peak-peak frequency deviation divided by the carrier frequency, 10% for all components). For comparison to other definitions of SFM signals, the frequency deviation (Δf_n) and modulation index (β_n) of the nth component are calculated using $\Delta f_n = 0.5DF_n$ and $\beta_n = 0.5DF_n/f_n$, respectively. Modulation indices ranged from $0.5 < \beta_n < 3$, so the SFM components varied from narrow band to wideband, depending on the carrier frequency. At the higher frequencies (e.g., 50–70 kHz), there was spectral overlap between the stimuli (Fig. 1).

A rugged notebook computer with a multifunction data acquisition board (NI PCI-6251) was used to generate sound stimuli and record the evoked responses. SFM tones were digitally generated and converted to analog signals at a rate of 2 MHz with 16-bit resolution. The stimuli were then filtered (Krohn-Hite 3C series, bandpass 0.2 to 150 kHz) and passed through a custom programmable attenuator (0–65-dB attenuation) and applied to a sound projector (ITC 1032) located 45 cm in front of the biteplate.

AEPs were measured using 10-mm gold cup surface electrodes embedded in silicon rubber suction cups and attached to the subject using conductive paste. Three electrodes were used: a noninverting electrode placed on the subject's head, just behind the blowhole near the midline, an inverting electrode placed along the midline approximately 30 cm behind the noninverting electrode, and a common electrode placed on the dorsal fin. A biopotential amplifier (Grass ICP-511) was used to amplify ($\times 10^5$) and filter

(0.3–3 kHz) the voltage between the noninverting and inverting electrodes before digitization at 10 kHz with a 16-bit resolution (NI PCI-6251). The digitized signal was divided into 61-ms epochs for analysis. The first epoch and any epochs with peak voltage exceeding 12 μV were excluded from analysis. Stimulus presentation occurred until 500 epochs were acquired (about 31 s for each measurement).

Coherent averaging in the frequency domain was used to obtain 20 unique "subaverages," each created from 25 individual epochs, as well as a single "grand average" from all 500 epochs. The subaverages were created using sequential epochs, i.e., epochs 1–25 for the first subaverage, epochs 26–50 for the second subaverage, etc. The presence or absence of an evoked response at each modulation frequency was objectively determined by calculating the magnitude-squared coherence (MSC). MSC is a ratio of the power in the grand average to the average power of the subaverages and thus is a ratio of signal power to signal-plus-noise power (Dobie and Wilson, 1989). Critical values for MSC (MSC_{crit}), using $\alpha=0.01$, were obtained from Amos and Koopmans (1963) and Brillinger (1978). If the calculated MSC was greater than MSC_{crit}, the ASSR at the modulation frequency was considered to be detected.

Each ASSR measurement began with the trainer directing the subject to the biteplate. Stimuli were then presented and the evoked responses recorded. Testing continued for periods of approximately 2–10 min, after which a buzzer was sounded to signal the subject to return to the trainer for fish reward. The process was then repeated as necessary.

During ASSR threshold measurements, the SPL of each waveform component was independently adjusted using an up/down staircase technique: the SPL at a carrier frequency was decreased following a detected evoked response (at the corresponding modulation frequency) and increased if no response was detected. Testing continued until at least three reversals, defined as changes from a detection to a nondetection or vice versa, were obtained for each component. Custom software was used to generate stimuli, record, analyze, and display the evoked responses, and automatically adjust stimulus levels from one measurement to the next. During experiment 1, the SPL at each carrier frequency initially began 10 dB above the estimated threshold. SPLs were then adjusted from trial-to-trial using a fixed step size of 3 dB. Subsequent testing began at levels closer to the last previously estimated threshold. During experiment 2, testing always began at SPLs about 10 dB above the pre-exposure ASSR thresholds from experiment 1. SPLs were then adjusted using a variable step size that began at 10 dB but was reduced to 5 dB after the first reversal and 3 dB for the remaining reversals. This adaptive procedure was intended to reduce the time required to reach threshold.

The raw data from threshold testing consisted of the time at which each SFM tone was presented, the SPL of each SFM tone component, and whether the evoked response corresponding to each tone component was detected. These data were converted to a series of reversal points which were then averaged to yield threshold means (and SDs) corresponding to the specific times postexposure. For each experiment, the zero-time reference was the end of the (last) exposure. Threshold shifts were defined as the threshold at specific postexposure time minus the pre-exposure threshold. Only reversals occurring with a 3-dB step size were used for threshold estimates.

Single ASSR measurements were also made during experiment 3 to characterize the ASSR I/O functions. For these measurements, stimulus SPLs began at 130 dB $re:1$ μPa and were reduced by 10 dB for each successive measurement until two consecutive nondetections were observed. Additional SPLs were then tested as needed to fill in any large gaps in the function amplitude. Thresholds were defined as the mean of the SPLs corresponding to the lowest detected response and the next highest undetected response.

E. Behavioral measurements

Behavioral hearing test tones consisted of linear frequency modulated tones with instantaneous voltage,

$$v(t) = A \sin\left[2\pi\left(f_1 t + \frac{Bf_c}{2T}t^2\right)\right], \tag{2}$$

where A is the amplitude, f_c is the center frequency, $f_c=(f_1+f_2)/2$, f_1 and f_2 are the lower and upper frequencies, respectively, B is the bandwidth, $B=(f_2-f_1)/f_c$, and T is the sound duration. The bandwidth (B) was 10% and the duration (T) was 500 ms (including a 50-ms rise/fall) for all three experiments. Linear frequency-modulated tones were used for the hearing tests to remain consistent with previous behavioral measurements for BLU in the test pool (Finneran *et al.*, 2007a; Finneran and Schlundt, 2007).

Tones were digitally generated using a multifunction data acquisition card (NI PCI-6070E) within a personal computer, filtered (SRS SR-560), attenuated (TDT PA5), amplified (Hafler P4000), and output through a piezoelectric sound projector (ITC 1032) located 85 cm in front of the biteplate. Analog-to-digital conversion was performed at 12-bit resolution and 500-kHz update rate.

Behavioral testing followed procedures similar to those previously described by Finneran *et al.* (2005, 2007a, 2007b). The test sequence was divided into a number of trial blocks, each containing a variable number of trials. A light was turned on and off to indicate the occurrence of a trial. Trial durations were 2 s for experiment 1 and 1.5 s for experiments 2 and 3. Intertrial intervals were 4 s (experiment 1) and 3 s (experiments 2 and 3). The trial duration and intertrial interval were shortened after experiment 1 to allow a more rapid pace of testing. Fifty percent of the trials were signal present and 50% were signal absent. The order of signal present/absent trials was based on Gellerman (1933). The subject was trained to whistle in response to tones and to stay quiet otherwise. Test tones began 500 ms after the light onset. Tone SPLs began above previously measured thresholds and were then adjusted using an up/down staircase technique (Cornsweet, 1962) with an initial step size of 5 dB, followed by a 2-dB step size after the first reversal. Thresholds were based on the average SPL over the last five reversals occurring with a 2-dB step size.

Whistle responses occurring during signal-absent trials were categorized as false alarms. The false-alarm rate was

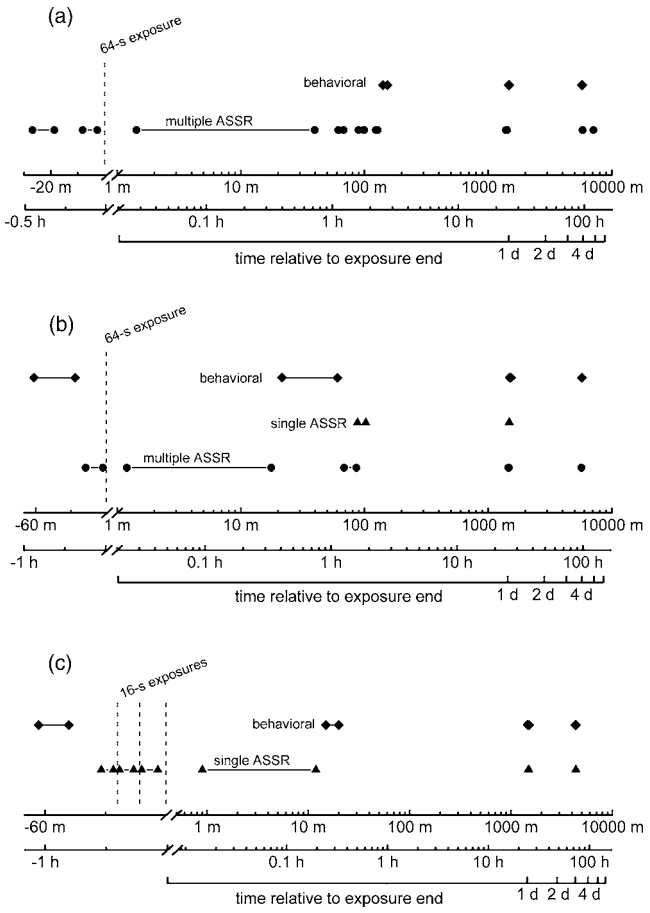


FIG. 2. Sequence of events for (a) experiment 1; (b) experiment 2; and (c) experiment 3. The vertical dashed lines indicate the end of each exposure. Circles, triangles, and diamonds indicate the starting and ending times for multiple ASSR, single ASSR, and behavioral measurements, respectively. Time zero coincides with the end of the last exposure. The time axis scale is linear before the last exposure and logarithmic afterwards.

calculated for each threshold measurement by dividing the number of false alarms by the number of signal-absent trials. The false-alarm rate calculation only considered trials occurring during the period over which the threshold was actually calculated (to obtain the false-alarm rate at threshold).

F. Test sequence

1. Experiment 1

Figure 2(a) illustrates the test sequence during experiment 1. Pre-exposure multiple ASSR thresholds were obtained over two separate test runs conducted from 3–27 min prior to the fatiguing sound exposure. Immediately following the exposure, multiple ASSR testing continued for approximately 40 min and then, after short pauses, resumed at approximately 60, 90, and 120 min postexposure. Multiple ASSR threshold testing was repeated 1, 4, and 5 days postexposure.

Behavioral threshold measurements were conducted at approximately 2.5 h, 1 day, and 4 days following the exposure. Hearing was tested at 30 and 40 kHz only, except at 4-days postexposure, where 4.5 kHz was also tested (for comparison to existing data). Because these measurements were made after extensive AEP tests, the subject had typically been in the pool participating in tests for 1–2 h, so declining food motivation and fatigue potentially affected the results (and actually prevented meaningful behavioral data from being collected at 1-day postexposure). In addition, the tests at 2.5 h were not originally planned, were conducted on the shallow biteplate (rather than the deeper biteplate normally used for behavioral tests), and the test tone levels were not specifically calibrated (though subsequent measurements suggested that thresholds were accurate to within ±5 dB). For these reasons, the 2.5-h behavioral thresholds should only be considered approximate and are merely intended as a rough comparison to the AEP results.

2. Experiment 2

Experiment 2 featured single and multiple ASSR threshold measurements as well as behavioral threshold measurements [Fig. 2(b)]. Pre-exposure hearing thresholds (and the variance occurring over successive days) were characterized with multiple ASSR and behavioral measurements 48 h, 24 h, and just prior to the exposure. Postexposure thresholds were made using multiple ASSRs over time periods of 1.2–18 min, 70–86 min, 1 day, and 4 days after the exposure. Single ASSR thresholds using a 30-kHz carrier frequency were also measured at approximately 97-min and 1-day postexposure. Postexposure behavioral thresholds were measured after ASSR measurements, at approximately 20–60 min, 1 day, and 4-days postexposure. Each behavioral session featured individual threshold measurements at 30, 40, 20, 50, 10, 60, and 70 kHz (in that order) and typically lasted 20–40 min.

3. Experiment 3

Figure 2(c) shows the test sequence for experiment 3. Single ASSR measurements (at 30 kHz) were conducted before the first exposure and then repeated after each of the three exposures. Behavioral thresholds were measured at 30 kHz before the first and after the last ASSR measurement. The test sequence was repeated 1 and 3 days after the exposures, with behavioral measurements occurring before and after single ASSR measurements each day.

III. RESULTS

A. Experiment 1

Example time waveforms and frequency spectra are shown in Figs. 3 and 4, respectively, for multiple ASSRs measured during experiment 1 over the time period from −27 to −19 min. Multiple ASSR spectral amplitudes were relatively small, normally under 50 nV (rms) at the stimulus levels employed. Noise levels were also relatively low, however, and ASSRs with amplitudes $>7-8$ nV were normally detected using MSC with 20 subaverages and $\alpha=0.01$.

Figure 5 shows the SPLs of each SFM frequency tone component presented during multiple ASSR testing and whether an evoked response was detected at the corresponding modulation rate. Immediate effects of the exposure occurred at 30 and 40 kHz, where SPLs increased substantially before responses were once again detected. A delayed in-

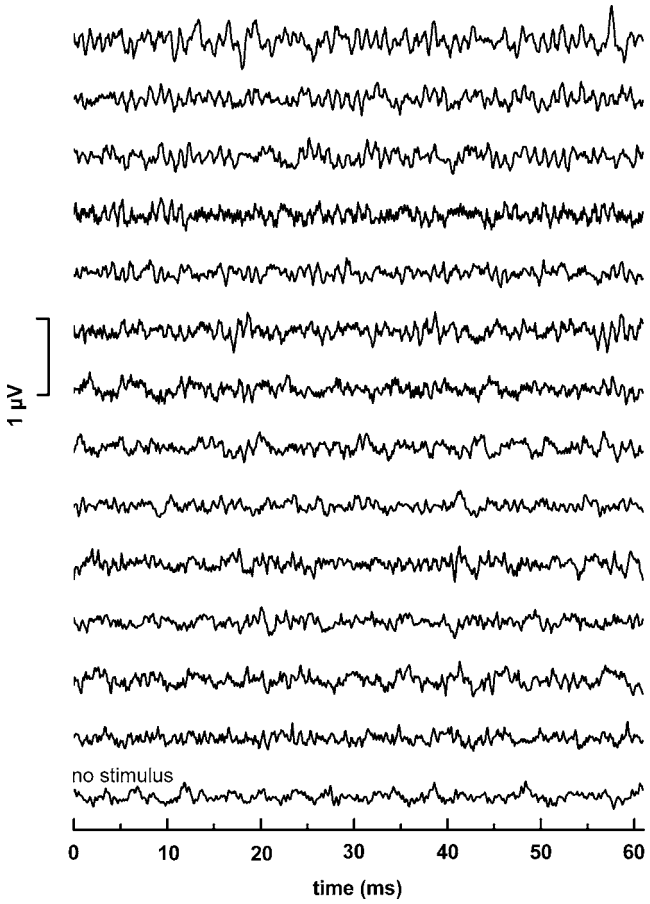


FIG. 3. ASSR instantaneous voltages measured during experiment 1 in response to seven simultaneous SFM tones. Measurements were made over the time range of −27 min (top trace) to −19 min (second trace from bottom) relative to the exposure. The amplitude of each SFM tone component was independently adjusted after each trial; SPLs of the seven stimulus waveform components corresponding to each response are shown in Fig. 4. The lowest trace was recorded in the absence of the sound stimulus as a measure of the background electrophysiological noise.

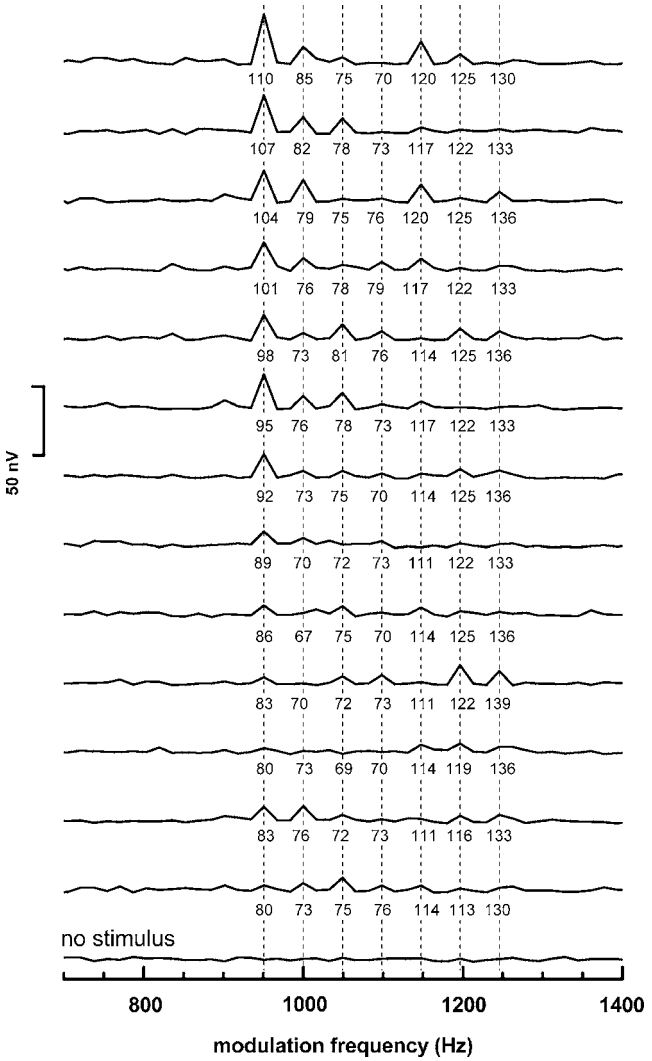


FIG. 4. ASSR (amplitude) spectra corresponding to the instantaneous voltages shown in Fig. 3. Stimuli consisted of seven simultaneous SFM tones, with modulation frequencies indicated by the dashed lines. The frequency resolution was 16.4 Hz. Measurements were made over the time range of −27 min (top trace) to −19 min (second trace from bottom) relative to the exposure. The amplitude of each SFM tone component was independently adjusted after each trial; the numbers below each trace indicate the SPL (dB *re*: 1 μPa) of each stimulus component. The lowest trace was recorded in the absence of the sound stimulus as a measure of the background electrophysiological noise.

crease in threshold occurred at 20 kHz (the exposure frequency). No similar effects were seen at 10, 50, 60, or 70 kHz.

Mean pre-exposure ASSR thresholds are shown in Fig. 6, along with behavioral thresholds for BLU previously obtained using linear FM tones with 10% bandwidth (Finneran *et al.*, 2007a; Finneran and Schlundt, 2007). The agreement between behavioral and multiple ASSR thresholds was good, with multiple ASSR thresholds typically 5–15 dB higher than the behavioral thresholds, which is consistent with previous AEP/behavioral comparisons in humans (Lins *et al.*, 1995; Rance *et al.*, 1995; Aoyagi *et al.*, 1999; Dimitrijevic *et al.*, 2002; Vander Werff and Brown, 2005) and marine mammals (Szymanski *et al.*, 1999; Yuen *et al.*, 2005; Cook *et al.*, 2006; Finneran and Houser, 2006; Houser and Finneran, 2006a; Finneran *et al.*, 2007a; Schlundt *et al.*, 2007).

Figure 7 shows the amount of TTS as a function of time postexposure for each of the SFM components. Threshold shifts at 10, 50, 60, and 70 kHz were generally within the range ±5 dB. At 30 and 40 kHz, initial TTSs were 30–40 dB and required 4 days to recover to within the baseline range. Little recovery occurred at 30 and 40 kHz within the first hour of testing. The amount of TTS over the time range of 60–6000 min (1 h to 4 days) was fit with a linear regression using the logarithm of time (Table II). Recovery rates were −6.0 and −4.8 dB per doubling of time at 30 and 40 kHz, respectively.

Figure 8 presents the hearing thresholds as a function of frequency for the pre-exposure (ASSR) or baseline [behavioral, Finneran *et al.* (2007a); Finneran and Schlundt (2007)] conditions, as well as various times postexposure (for clarity, only a subset of postexposure data is shown). Figure 9 presents the same data in terms of the amount of TTS as a function of the hearing test frequency. Both the ASSR [Figs. 8(a) and 9(a)] and behavioral [Figs. 8(b) and 9(b)] data show the same trends: the largest TTS occurred at 30 kHz, which is close to one-half octave above the exposure frequency (28.3 kHz). Smaller amounts of TTS were observed at 40

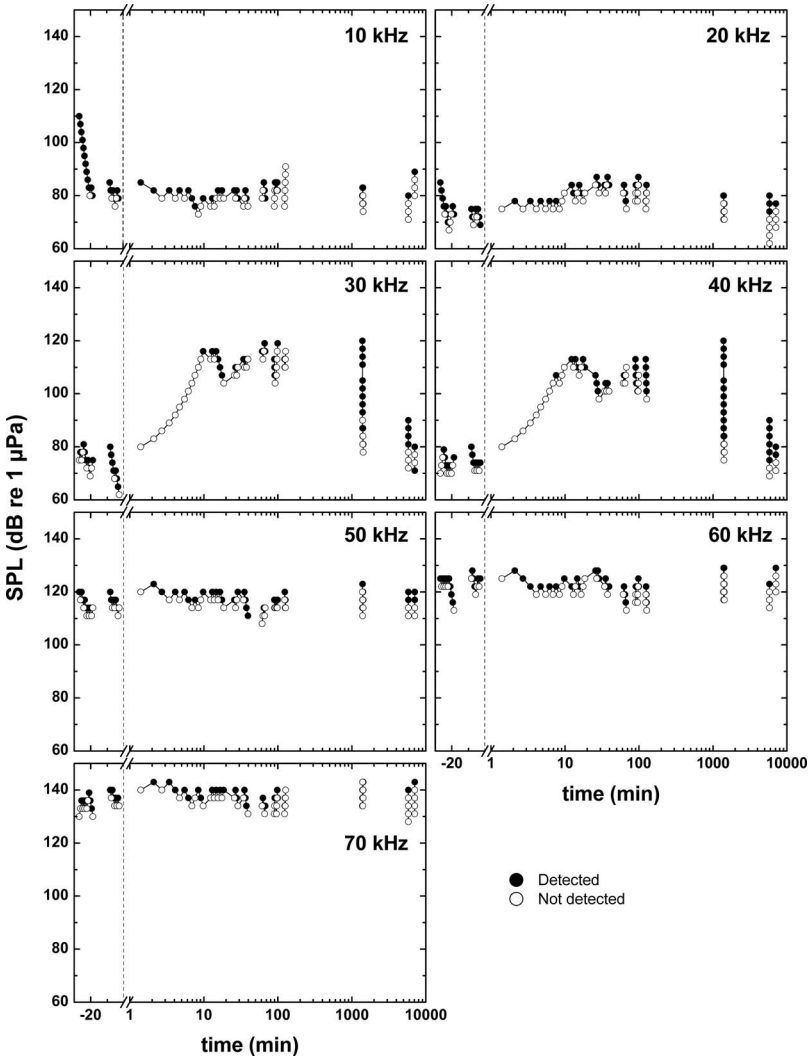


FIG. 5. AEP stimulus levels tested as a function of time relative to the exposure during experiment 1. Each panel shows the data from a specific carrier frequency. The filled symbols represent detected responses; the open symbols indicate no detected response at that SFM component frequency/SPL. The first data group, between time values of −27 to −19 min, corresponds to the waveforms and spectra shown in Figs. 3 and 4. The vertical dashed line indicates the exposure end (time zero). The time axis scale is linear before the exposure and logarithmic afterwards. The effects of the exposure are immediately seen at 30 and 40 kHz, where the SPLs had to be significantly increased to once again produce detectable responses. A delayed effect appeared at 20 kHz, while 10, 50, 60, and 70 kHz were not significantly affected.

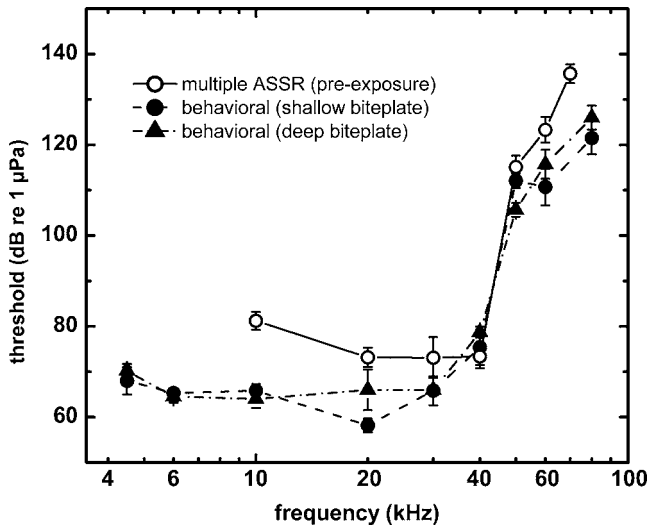


FIG. 6. Pre-exposure multiple ASSR and baseline behavioral hearing thresholds at each test frequency. Symbols represent the mean (±1 SD). ASSR thresholds were averaged over the pre-exposure data from Fig. 5. The behavioral thresholds for BLU were obtained with FM sweeps as described by Finneran *et al.* (2007a, 2007b), using the same shallow and deep biteplates described in the text.

and 20 kHz (ASSR) and no substantial effects were seen at 10, 50, 60, 70 kHz (ASSR) or 4.5 kHz (behavioral). The maximum amounts of TTS measured with multiple ASSRs at 30 and 40 kHz were 43 dB (at 60 min) and 34 dB (at 24 min), respectively. At 2-h postexposure, multiple ASSR TTSs were 38 and 26 dB at 30 and 40 kHz, respectively. Behaviorally measured TTSs at 2.5-h postexposure were 33 and 18 dB at 30 and 40 kHz, respectively. Analysis of the false-alarm rates was confounded to some extent by the few stimulus-absent trials, which ranged from 3–14 (a product of the low number of signal-present trials required to obtain five reversals). The subject rarely committed false alarms and there were no systematic differences between the pre- and postexposure data.

B. Experiment 2

Figure 10 shows the multiple ASSR results in terms of the amount of TTS as a function of time postexposure for each frequency (analogous to Fig. 7). As in experiment 1, immediate threshold shifts were seen at 30 and 40 kHz, while a delayed threshold shift was observed at 20 kHz. No substantial increases in thresholds were observed at 10, 50,

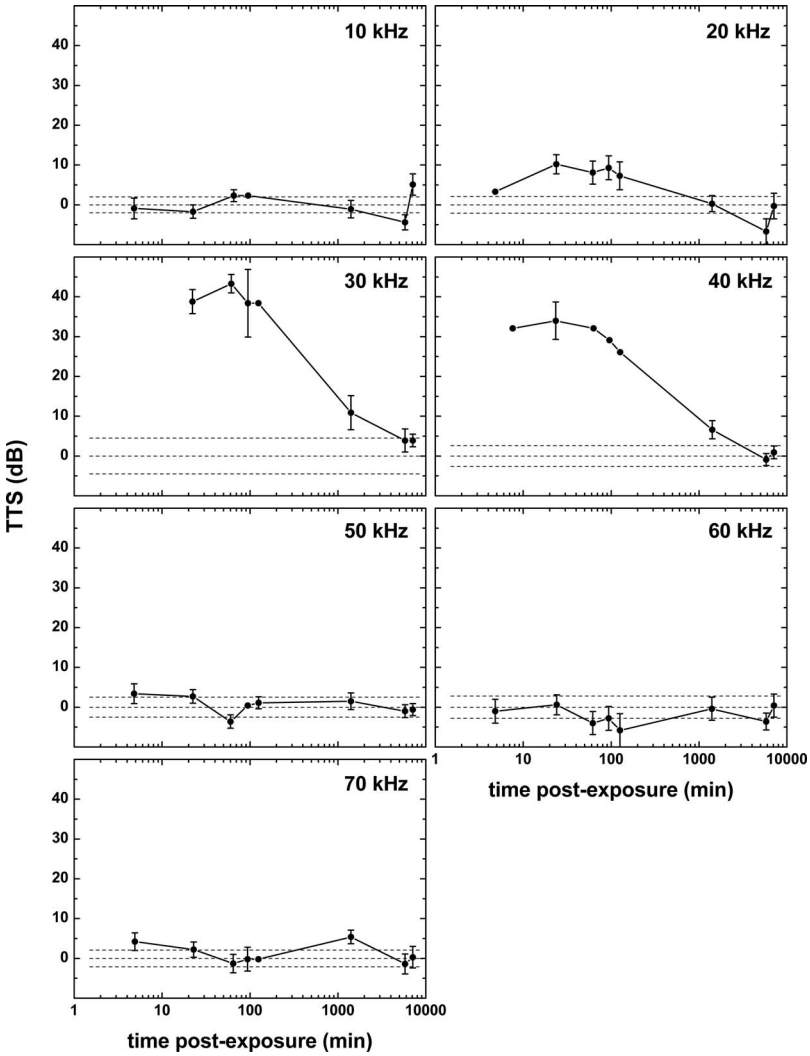


FIG. 7. Amount of TTS measured at each test frequency for experiment 1. Each symbol represents the mean threshold (±1 SD) at a particular time, calculated as the mean SPL and mean time over which a series of detection/nondetection reversal points were averaged. The dashed lines in each panel indicate 0 dB ± the pre-exposure SD. Thresholds at 10, 50, 60, and 70 kHz remained within approximately ±5 dB of the pre-exposure values while large TTSs, requiring up to 4 days for complete recovery, were observed at 30 and 40 kHz.

60, or 70 kHz. Measurement variability was higher than in experiment 1, especially at 10 kHz, where the pre-exposure SD was relatively high (7 dB).

Figures 11 and 12 present the hearing thresholds and the amounts of TTS, respectively, as functions of frequency for the ASSR [Figs. 11(a) and 12(a)] and behavioral [Figs. 11(b) and 12(b)] measurements. The ASSR results were similar to those of experiment 1: The largest TTS occurred at 30 kHz, followed by 40 and 20 kHz. At approximately 10-min post-exposure, TTSs at 30 and 40 kHz were 43 and 36 dB, respectively. There was good agreement (<3 dB) between the multiple ASSR data and the two single ASSR threshold measurements at 30 kHz. ASSR thresholds recovered to within the pre-exposure range after 4 days. Recovery data from 60–6000 min were fit with a linear function with the logarithm of time and the recovery rates were similar to those observed in experiment 1 (Table II).

Behaviorally measured TTSs at approximately 40-min postexposure were 19 and 13 dB at 30 and 40 kHz, respectively, much smaller than the multiple ASSR TTS measured at 10- and 78-min postexposure and the single ASSR TTS measured at 97 min postexposure. Behavioral thresholds returned to within 4 dB of pre-exposure values after 24 h. False-alarm rates were relatively high (up to 20%–50%) dur-

TABLE II. Slope and goodness of fit (r^2) for TTS recovery data. The functional form was $\mathrm{TTS}(t)=m\log_2(t)+b$, where m is the slope and b is the y intercept. Time values were constrained to $60\leqslant t\leqslant 6000$ min. The slope parameter has units of dB per doubling of time.

	Hearing test frequency					
	20 kHz		30 kHz		40 kHz	
Experiment	Slope (dB/$\log_2 t$)	r^2	Slope (dB/$\log_2 t$)	r^2	Slope (dB/$\log_2 t$)	r^2
1	−1.9	0.862	−6.0	0.982	−4.8	0.985
2	−1.8	0.999	−5.3	0.997	−5.1	0.999

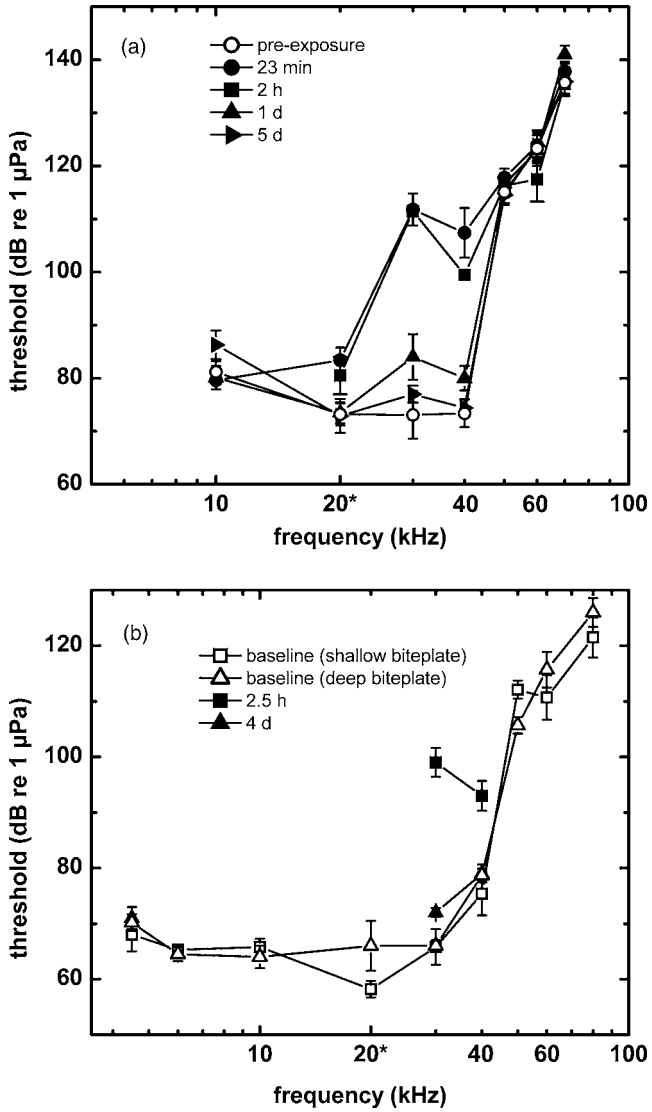


FIG. 8. Comparison of baseline, pre-, and postexposure audiograms obtained with the (a) multiple ASSR and (b) behavioral test methods during experiment 1. Symbols represent means ±1 SD. The asterisk indicates the exposure frequency (20 kHz). The exposure essentially shifted portions of the audiogram upward. Over time, the thresholds returned to normal and the audiogram returned to the pre-exposure or baseline patterns.

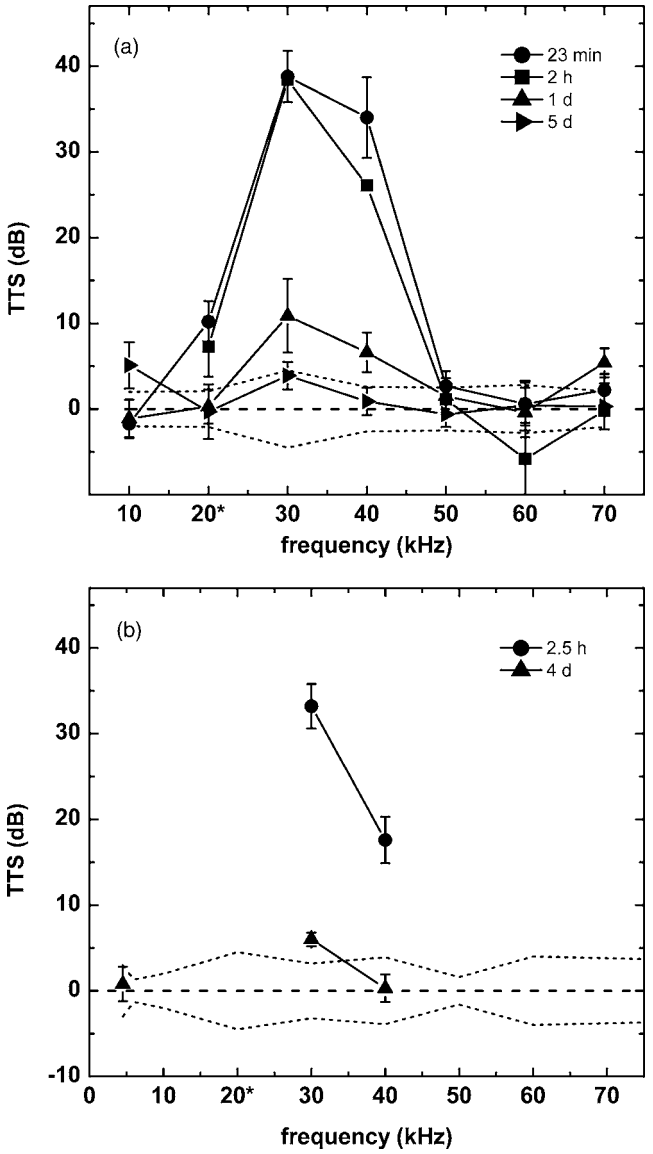


FIG. 9. TTS as a function of frequency for various postexposure times during experiment 1. Data are shown for the (a) multiple ASSR and (b) behavioral methods. The dotted lines show the SD for the baseline or pre-exposure thresholds. The asterisk indicates the exposure frequency (20 kHz). The maximum effect of the exposure occurred at 30 kHz, close to one-half octave above the exposure frequency.

ing the testing 48 h prior to the exposure. This may have reflected the subject having difficulty initially adapting to the changing test frequencies. The subject rarely committed false alarms on subsequent days (no false alarms during 91% of the threshold measurements) and the most during a single threshold test were 2/7. Again, the number of stimulus-absent trials was typically very small (mean=6.5).

C. Experiment 3

Figure 13 shows the ASSR I/O functions measured at various times during a sequence of three 16-s exposures. Only detected responses are shown. The I/O functions appeared progressively shifted downward and to the right after each exposure. The curves measured 1- and 3-days postexposure are shifted back to the left and approach the pre-exposure I/O function.

The pre-exposure curve followed a pattern often seen in ASSR I/O functions and consisted of three phases: a region near threshold where ASSR amplitude was approximately linear with SPL, a plateau at higher SPLs where increases in stimulus SPL produced little change in ASSR amplitude, and a region at high SPLs where the ASSR amplitude again increased [e.g., Fig. 13, Supin *et al.* (2001); Finneran and Houser (2006); Finneran *et al.* (2007a)]. This particular function can be represented as the sum of two individual processes: a saturating process existing at lower SPLs and a linear (with SPL) process at higher SPLs. This is demonstrated in Fig. 13 using the dashed lines to indicate the saturating and linear processes and the thick solid line to indicate their sum. The saturating process was modeled as $f(x) = C_1[1-\exp(C_2-x/C_3)]$. Nonlinear regression was first used to fit the pre-exposure data [Fig. 13(a)]; then, C_3 and the

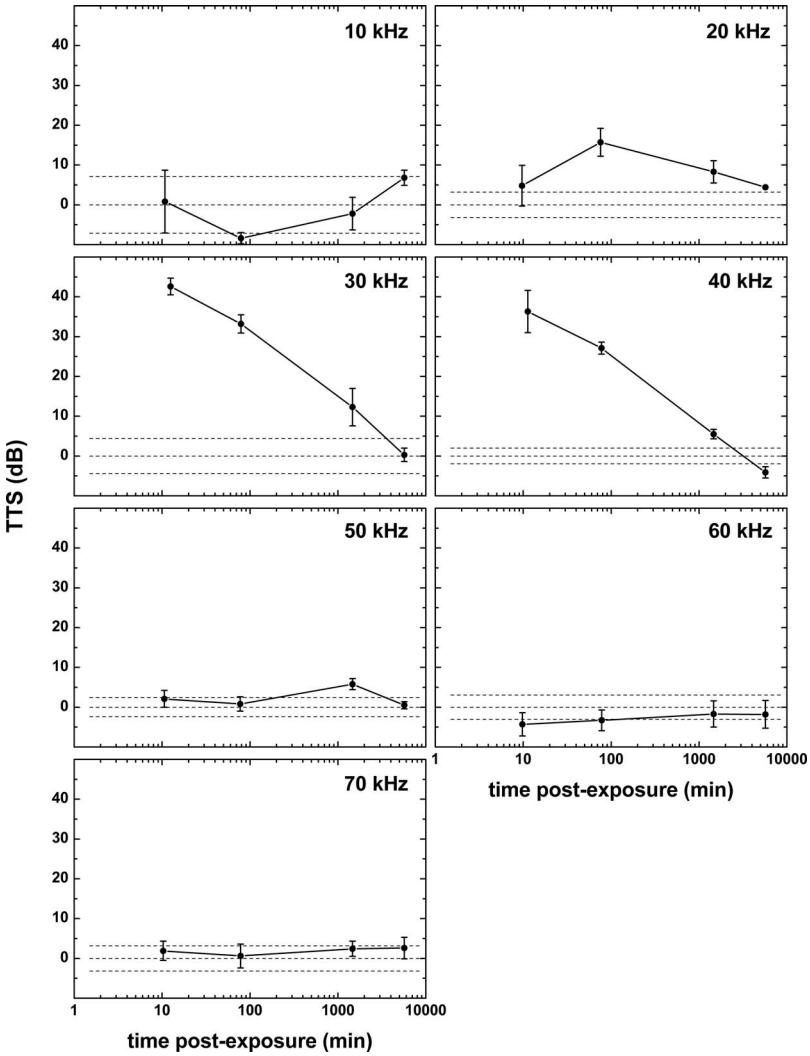


FIG. 10. Amount of TTS measured at each test frequency for experiment 2. Each symbol represents the mean threshold (±1 SD) at a particular time, calculated as the mean SPL and mean time over which a series of reversals points were averaged. The dashed lines in each panel indicate 0 dB± the pre-exposure SD. Thresholds at 10, 50, 60, and 70 kHz remained near the pre-exposure values. Large TTSs were again observed at 30 and 40 kHz with smaller, delayed TTS at 20 kHz.

slope of the linear process were held constant and nonlinear regression used to fit the postexposure data. The two-process model fit the pre- and postexposure data reasonably well. Immediately after each exposure [Figs. 13(b)–13(d)], the saturating process was shifted downward and to the right, while the linear process was initially shifted down [Fig. 13(b)] but changed little after the second and third exposures. After the third exposure the influence of the saturating process was negligible [Fig. 13(d)]. By 1–day postexposure, the linear portion had nearly recovered but the saturating process still showed small differences between the pre-exposure curve at 3–days postexposure.

Thresholds measured as functions of time postexposure are shown in Fig. 14. Vertical lines represent the time of each exposure. Following the third exposure the (single) ASSR TTS was approximately 45 dB. Behavioral measurements produced lower amounts of TTS: 26 dB at 17-min postexposure and within 4 dB of pre-exposure (1 dB of the previously measured baseline) after 24 h. The subject committed only one false alarm within the limited number of stimulus-absent trials (mean of 4.8 for each threshold measurement).

IV. DISCUSSION

A. Multiple ASSR threshold measurements

The multiple ASSR technique is commonly used in humans to simultaneously test up to four frequencies in each ear using SAM tones (Lins and Picton, 1995; John *et al.*, 1998). General guidelines prescribe a frequency separation of at least one octave between adjacent stimuli to avoid interactions (Lins and Picton, 1995; John *et al.*, 1998). Measurements in dolphins using a simultaneous SAM tone probe and pure-tone masker also suggest that a spacing of one octave would be sufficient to avoid significant interactions between nearby frequencies (Popov *et al.*, 1997, 1998). The current study clearly violates these conditions, as the SFM tone center frequencies are not always a full octave apart and the relatively wide bandwidths of the SFM tones produced some spectral overlap between sidebands for $F_n \geqslant 40$ kHz. Complex interactions between the SFM components therefore likely occurred. Since the SPL of each SFM component was independently adjusted and interactions depend, in part, on the relative SPLs of the components, the specific effects of the interactions are difficult to determine. In some cases the ASSR amplitude at a particular modulation rate may have been enhanced by the presence of an adjacent signal, in other

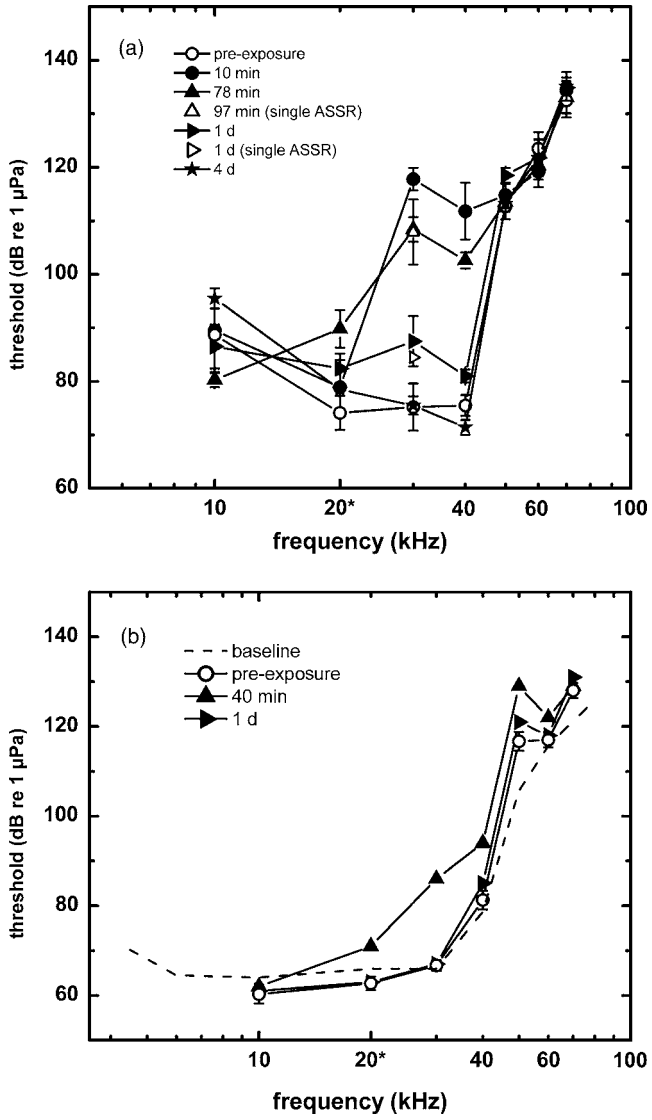


FIG. 11. Comparison of baseline, pre-, and postexposure audiograms obtained with the (a) ASSR and (b) behavioral test methods during experiment 2. Symbols are means ±1 SD. The asterisk indicates the exposure frequency (20 kHz).

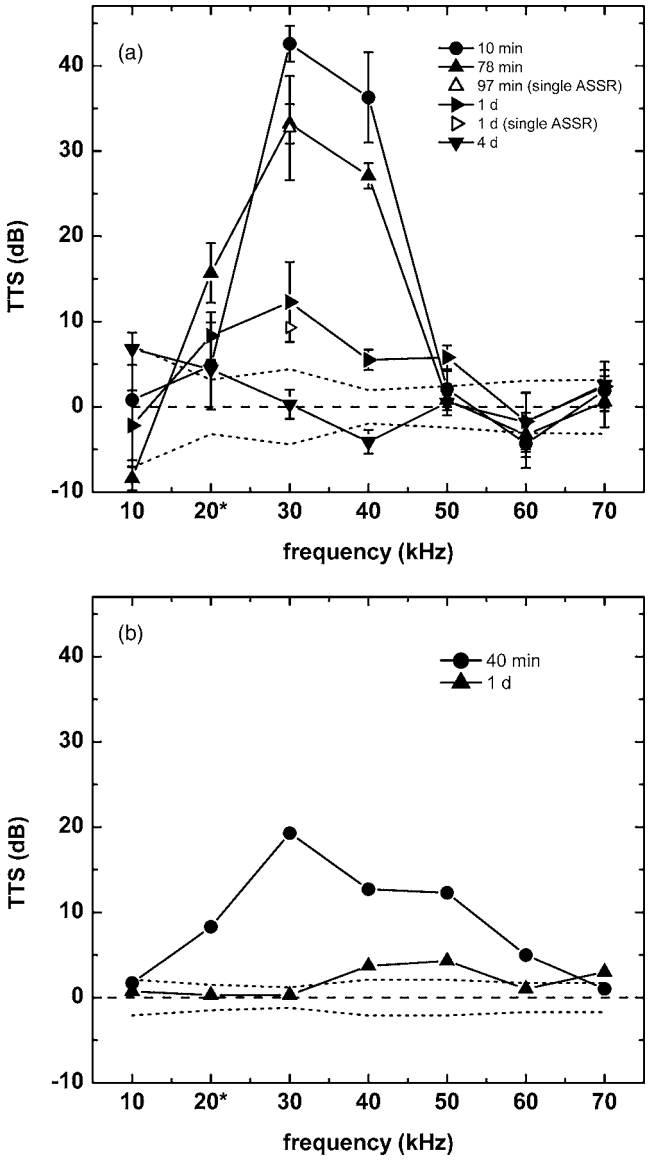


FIG. 12. Comparison of TTS as a function of frequency for various postexposure times during experiment 2. Data are shown for the (a) ASSR and (b) behavioral methods. The dotted lines show the SD for the pre-exposure thresholds. The asterisk indicates the exposure frequency (20 kHz). The maximum effect of the exposure occurred at 30 kHz, close to one-half octave above the exposure frequency.

cases it may have been reduced, and thus the detectability of the evoked responses would have fluctuated. This likely contributed to the variability seen over time in the detectability of the ASSRs (Fig. 5) and the resulting thresholds (Figs. 7 and 10). As testing continued, all the components tended to reach levels near their thresholds, which may have lessened the severity of the interactions. In the end, threshold variance was not unusually high—SDs in pre-exposure thresholds ranged from 2–3.5 dB, except at 20 kHz (4.4, 4.5 dB) and 10 kHz (7 dB, experiment 2). The agreement between pre-exposure thresholds and behavioral thresholds was similar to that observed using single SAM tones (Finneran *et al.*, 2007a). Also, the amounts of TTS measured using the single and multiple ASSR techniques (experiment 2) were very close (within 3 dB). So, for this particular subject and frequency range, the use of the multiple ASSR technique enabled the entire audiogram to be quickly estimated, rather than a threshold at a single frequency. In many applications, such as clinical hearing screening or the assessment of rehabilitated animals prior to release, threshold variability of ±5 dB observed over time would be acceptable since the upper cutoff frequency and any mild hearing loss would be discernible. In situations where this level of variability is not acceptable, the simultaneous components should be reduced in number and spaced further apart in frequency, or a single stimulus used instead.

The multiple ASSR technique was also successful in allowing TTS measurements at multiple frequencies immediately after a single fatiguing sound exposure. In experiment 1, multiple ASSR thresholds, based on at least three reversals, were obtained within 4 min of the exposure at all frequencies except 30 and 40 kHz (13 min), where the initial SPLs were well below levels necessary to produce detectable responses. During experiment 2, multiple ASSR thresholds, based on at least three reversals at the 3-dB spacing, were

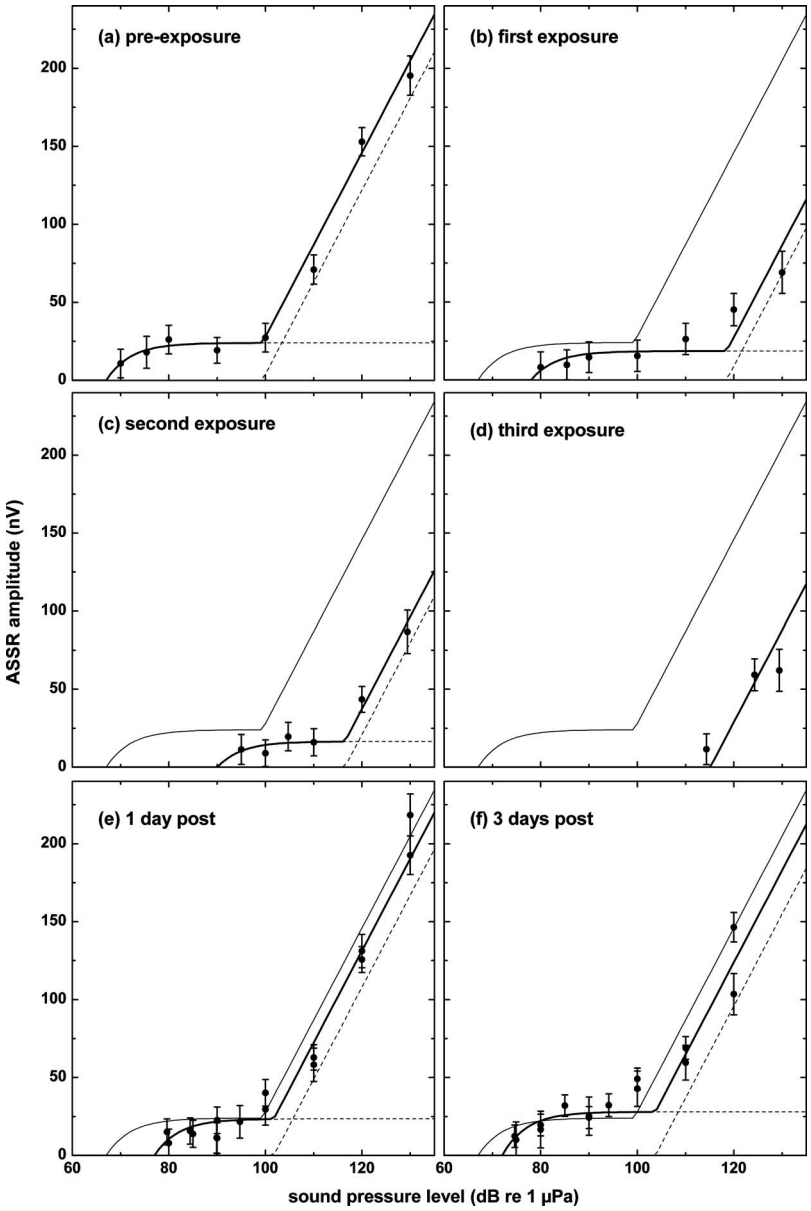


FIG. 13. Single ASSR input-output functions at 30 kHz recorded during experiment 3. The panels show the results of a series of measurements taken: (a) pre-exposure and after (b) the first 16-s exposure; (c) the second 16-s exposure; (d) the third 16-s exposure; (e) 1 d postexposure; and (f) 3 d postexposure. Symbols represent detected responses (mean ±95% CI). Two measurement series were conducted 1- and 3-day postexposure. Each curve can be represented as the sum (thick solid line) of two processes (dashed lines): a saturating process at lower SPLs and a linear process at higher SPLs. The pre-exposure curve fit is represented by a thin solid line in (b)–(f). The postexposure data may be fit by shifting the saturating process down and to the right without changing the linear process slope. After the third exposure the influence of the saturating process is negligible.

obtained at all frequencies within 11 min. Because of the variability in the ASSR thresholds, the multiple ASSR technique may not be suitable for studies involving very small amounts of TTS or investigations requiring fine-scale comparisons; however, the multiple ASSR method shows promise for initial evaluation of frequency-dependent effects—for broadband sources it would allow the most susceptible frequencies to be quickly identified. Follow-on work could then be conducted using single frequencies, if desired, to evaluate effects with greater resolution.

B. Behavioral threshold measurements

Behavioral approaches previously used to measure TTS in dolphins (Finneran *et al.*, 2005) were modified for experiment 2 so that thresholds could be obtained more rapidly. The most significant changes involved an accelerated signal presentation rate (an intertrial interval of 3 s instead of 4 s), a shorter response interval (1.5 s compared to 2 s), and calculating thresholds based on five reversals, rather than ten reversals as used previously (Finneran *et al.*, 2005). Also, multiple frequencies were tested before and after the fatiguing sound exposure; the first time this was performed with BLU. The procedural changes allowed audiograms from 10–70 kHz to be obtained during each session; however, there was generally more variability in the data than in previous behavioral measurements testing only a single frequency. Session times required to obtain thresholds at all seven frequencies ranged from 23 to 83 min with a mean time of 42 min, about 3–4 times longer than the time required for the multiple ASSR thresholds to be obtained but significantly faster than previous behavioral threshold measurements. The enhanced test pace and inclusion of multiple frequencies was not without consequences; for example, each time the frequency was changed the subject required a number of "warm-up" trials to acquire the stimulus frequency before reliable threshold data could be obtained. In some cases, e.g., at 50 kHz, thresholds remained unusually high over multiple sessions. The few stimulus-absent trials within a single threshold measurement made false-alarm

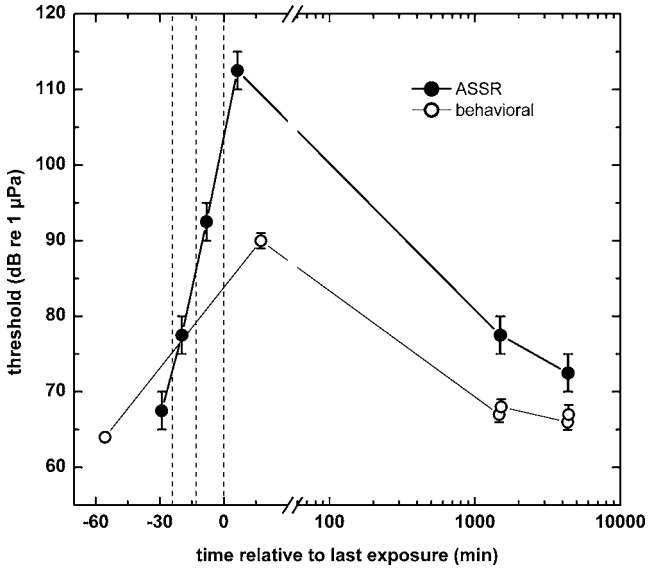


FIG. 14. Behavioral and ASSR thresholds at 30 kHz measured during experiment 3. ASSR thresholds were defined as the mean SPL of the lowest detection and the next highest nondetection. Error bars represent the 95% CI for the behavioral measurements and the SPL range between the lowest detection and next highest nondetection for the ASSR.

characterization problematic. Declining food motivation and satiation were also concerns as the total testing time became relatively long. Still, agreement between the pre-exposure thresholds in experiment 2 (obtained in 35 min) and baseline thresholds previously collected over a period of several weeks (Finneran *et al.*, 2007a; Finneran and Schlundt, 2007) was very good (within 4 dB, Fig. 11), except at 50 kHz, where the pre-exposure values were 11 dB above the baseline values. Just as the multiple ASSR technique may be particularly useful in situations where speed is of greater relative importance, a behavioral approach such as that used in experiment 2, with reinforcement only after a number of trials, a rapid pace of trial presentation, and thresholds based on a relatively small number of reversals, may be used to quickly estimate the audiogram. When time permits, more rigorous techniques may be applied to improve the measurement accuracy and reduce variability.

C. ASSR and behavioral measures of TTS

A consistent feature of the behavioral and ASSR data was the relatively large initial amounts of TTS. The ASSR measurements yielded TTSs up to 40–45 dB and the behavioral results indicated TTSs as large as 19–33 dB, substantially larger than TTSs previously reported for dolphins (Finneran *et al.*, 2000; Schlundt *et al.*, 2000; Finneran *et al.*, 2002; Nachtigall *et al.*, 2004; Finneran *et al.*, 2005). Another consistent feature was the relatively large amounts of TTS measured electrophysiologically compared to those measured behaviorally. The differences in postexposure measurement times do not provide a suitable explanation, since the ASSR measurements in experiment 2 at 78 and 97 min also resulted in much larger TTSs than the behavioral data at 40-min postexposure. The discrepancy cannot be attributed to the multiple ASSR technique, since the single and multiple ASSR measurements during experiment 2 yielded very close results and the same trends were seen in experiment 3, which only measured single ASSRs.

Behavioral reports of TTS in marine mammals have at times exhibited relatively large variances, with approximately equal exposures producing TTSs that differ by up to 15–25 dB (e.g., Kastak *et al.*, 2005), so some of the differences between the ASSR and behavioral results may have been caused by the inherent variability in the behavioral measurements. Although the ASSR and behavioral stimuli possessed the same frequency bandwidth, and previous behavioral measurements did not reveal substantial differences between linear FM and SFM thresholds with equivalent bandwidth/depth, differences between the stimuli may have also played a role. It may be possible that the relatively slow linear sweep used for the behavioral testing produced some perceptible artifact that was detectable, though this does not fully explain the increasing discrepancy between behavioral and ASSR thresholds with fatiguing sound exposure and seems unlikely to have resulted in such large differences.

It is more likely that the observed differences resulted from innate differences between the electrophysiological and behavioral techniques. Unfortunately, there have been few direct comparisons between electrophysiological and behavioral TTS. Evoked potential measurements in terrestrial mammals have typically relied on transient auditory evoked potentials or whole-nerve action potentials (APs) rather than ASSRs evoked from longer duration stimuli (cf. Schmiedt, 1984; Clark, 1991). The resulting comparisons between electrophysiological and behavioral TTS in terrestrial mammals have been equivocal: Mills *et al.* (1970) reported that asymptotic TTSs measured with AEPs approximated those measured behaviorally. Saunders and Rhyne (1970) measured the frequency following response in cats at the cochlear nucleus and observed less TTS compared to previous psychophysical measures; however, whether this was caused by the electrophysiological approach or to a reduced middle-ear reflex in the anesthetized subjects was unknown. Benitez *et al.* (1972) compared behaviorally measured TTS in chinchillas to that observed using APs and brainstem AEPs to clicks and tone pips, respectively. TTSs estimated from the AP I/O functions were much larger (40–50 dB) than those measured behaviorally, while the brainstem AEP and behavioral results were more consistent. A loss of synchrony in the neuronal responses, as opposed to a lack of responses, was suggested as an explanation. These findings did not agree with those of a later study showing good agreement between behavioral shifts and those measured with APs (Salvi *et al.*, 1979), but did agree with studies showing relatively good agreement between brainstem AEP and behavioral TTS [Henderson *et al.* (1983), in chinchillas; Borg (1982), in rats]. These comparisons between behavioral, AP, and brainstem AEP data are complicated by the peripheral nature of the AP (generated at the VIIIth nerve) compared to the more central nature of the brainstem AEP, differences in hearing test stimuli (e.g., clicks or tone pips), and differences in exposure durations and SPLs (Schmiedt, 1984; Abbas, 1988). Auditory fatigue not only affects sensitivity but also affects frequency selectivity and tuning curves, temporal summa-

tion, and temporal resolution, making it important to keep in mind the fundamental differences in the behavioral and evoked responses and the peripheral and central components each relies on.

The only previous TTS study to use the ASSR was conducted by Nachtigall *et al.* (2004). Although these tests were performed with a bottlenose dolphin, the focus was on smaller amounts of TTS and behavioral measurements were not made. The ambient noise levels were also relatively high, requiring relatively high SPLs at threshold even before the exposure (e.g., the subject's pre-exposure threshold at 8–10 kHz was approximately 25 dB higher than BLU's). This makes it difficult to rule out specific features of SAM or SFM tone ASSRs that would lead to increasing discrepancies between behavioral and ASSR thresholds at moderate to large amounts of TTS. For example, if the SFM tone ASSR depends on basilar-membrane amplitude changes produced as the excitation moves back and forth along the membrane, then changes to the basilar-membrane frequency response (such as a loss of sharp tuning) would affect the resulting ASSR amplitude. Similarly, if temporal resolution degraded and affected the ability to follow rapid changes in amplitude, ASSR amplitudes would be correspondingly reduced.

The discrepancy between ASSR and behavioral TTS could also be related to the particular shape of the ASSR I/O function, which often features a plateau where increases in stimulus SPL produce little change in ASSR amplitude. If the I/O function is shifted downward or compressed after noise exposure, i.e., the plateau occurs at a lower ASSR amplitude, then at some point the plateau may be obscured by the measurement noise floor and the measured ASSR threshold will rise dramatically. Measurements of small amounts of TTS would therefore be unaffected, but larger shifts would appear progressively higher than behavioral measures. The data in Fig. 13 tend to support this concept, where the I/O function was shifted downward after exposures and the plateau was no longer apparent after the third exposure.

Clearly, more data are needed to untangle the relationship between behavioral and ASSR measures of TTS. ASSR thresholds and I/O functions obtained with SAM and SFM tone stimuli, as well as behavioral threshold measures, are needed over a range of exposures to see if there is agreement at lower amounts of TTS and progressively larger discrepancies at larger TTSs and to track the changes in the I/O functions. Until the relationship between ASSR and behavioral TTS measurements is clarified, care should be exercised when interpreting large TTSs measured with the ASSR and in comparing and/or pooling ASSR and behavioral measures.

D. Comparison to TTS at 3 kHz

The behaviorally measured TTSs at 30 kHz from experiments 1 and 2 (33 and 19 dB, respectively, measured $\geqslant$40 min after exposure) were larger than the TTSs behaviorally measured at 4.5 kHz in BLU following 3-kHz exposures with 64-s durations and similar SPLs [e.g., TTS_4 of 6±4.7 dB (mean ±SD) for 180 dB *re*:1 μPa and TTS_4 of 14±2.8 dB for 193 dB *re*:1 μPa (Schlundt *et al.*, 2006)]. Behavioral thresholds at 3–4.5 kHz differ from those at 20–40 kHz by 5–10 dB, so while the increased sensation level of the 20-kHz exposure relative to the 3-kHz exposure may have played a role in the larger shift, it seems unlikely that this was the only cause. These data are similar to those previously obtained in humans, where the susceptibility to noise is higher within the region of best hearing (e.g., Mills, 1982), but there is no simple correlation between the amount of TTS and the pre-exposure thresholds (e.g., Ward *et al.*, 1959). The vast majority of TTS data in odontocetes is from relatively low-frequency sources ($\leqslant$10 kHz), below the frequency region of best hearing. TTS data at higher frequencies ($\geqslant$20 kHz) in odontocetes are very limited and consist of onset-TTS levels for a few individuals at 20 kHz and inconsistent data at 75 kHz (Schlundt *et al.*, 2000), pointing to a need for data regarding the onset and growth of TTS are at higher frequencies where sensitivities are better.

E. TTS recovery

In contrast to results seen at lower amounts of TTS, where recovery to short-duration exposures was complete within 10–30 min (Finneran *et al.*, 2002, 2005), the ASSR thresholds at 30 and 40 kHz showed little change with post-exposure time within the first hour and required 4 days to return to within the baseline range. Considering the time period from 60 to 6000 min, recovery rates were between 4 and 6 dB per doubling of time at 30 and 40 kHz. These rates are much higher than those measured by Nachtigall *et al.* (2004), who reported recovery rates of about 1.5 dB per doubling of time from TTS_5 of 5–7 dB after long-duration exposure to octave-band noise. The differences in recovery rates may be caused by different recovery processes, one occurring at higher levels of shift and proceeding more rapidly, the other occurring at lower amounts of shift and occurring more slowly. Recovery rates in the present study at 20 kHz, where the maximum shifts were only 10–15 dB, were approximately 1.8 dB per doubling of time, similar to those reported by Nachtigall *et al.* (2004). The fatiguing exposures from the two studies were also markedly different: Nachtigall *et al.* (2004) used 4–11-kHz noise for 30–55 min compared to the 64-s tones used in experiments 1 and 2.

F. Frequency patterns of TTS

The spectral pattern of TTS extended over a much larger bandwidth, roughly 25 kHz wide, compared to the narrowband fatiguing stimulus. The specific frequency pattern of TTS, with the maximum shift at 30 kHz, close to 1/2 octave above 20 kHz (28 kHz), agrees with existing marine and terrestrial mammal data indicating that the largest shifts to intense pure-tone fatiguing stimuli occur 1/2 octave to one octave above the exposure frequency, rather than at the exposure frequency itself (Ward, 1962; Schlundt *et al.*, 2000). Similar results have been observed in dolphins exposed to broadband noise (Nachtigall *et al.*, 2004). The behavioral data confirm these findings, with the largest TTSs also at 30 and 40 kHz for both experiments 1 and 2. The behavioral data from experiment 2 were somewhat unusual in revealing a relatively large shift at 50-kHz as well. It is not clear how

much importance to attach to this finding, however, since the 50-kHz measurements during experiment 2 were somewhat problematic and the subject seemed to have difficulty generalizing to the 50-kHz test frequency. The influence of the subject's high-frequency hearing loss on the spread of TTS with increasing frequency is unknown. Sensation levels at 50 kHz were about 40 dB higher than at 20–40 kHz, so it is possible that a broader spread of TTS would have occurred if the subject possessed normal high-frequency hearing.

G. I/O functions

The mechanisms governing the shape of the ASSR I/O curves remain speculative. However, the two-process mechanism proposed in the Results section (experiment 3) is reminiscent of basilar-membrane displacement amplitude curves, where active cochlear processes produce a low-level saturating response and passive cochlear mechanics produce a linear (with sound pressure) response at higher stimulus levels (Johnstone *et al.*, 1986). When acoustic trauma or overstimulation disrupts the active processes, the basilar-membrane response lacks the lower level saturating response and becomes more linear, though with a higher threshold. It is intriguing that the appearance and changes in the ASSR I/O functions also fit a similar framework featuring a low threshold, saturating process and a higher threshold, linear process, that react and recover to fatigue at different rates. This interpretation fits the data of Fig. 13 well, especially the preexposure function shape, the loss of the plateau after the third exposure, and the recovery of the linear portion before the plateau. If the two-process mechanism proves valid, the occurrence of the plateau or saturating response in the I/O function may prove to be a useful tool in assessing the overall health of the cochlea.

V. CONCLUSIONS

Multiple ASSR measurements may be used to simultaneously estimate hearing thresholds at a number of frequencies in dolphins. The resulting data provide good predictions of audiogram shape and reveal areas of mild (or worse) hearing loss, making this technique useful for rapid hearing assessment in clinical or field settings and revealing frequency-dependent patterns of hearing loss and recovery.

The amount of TTS measured with the ASSR may exceed that measured behaviorally, especially when the amount of TTS is relatively large. The specific reasons for the discrepancies are unknown, but may include systemic differences between AEP and behavioral methods, the nonmonotonic shape of the particular ASSR I/O functions, or the use of SFM stimuli.

The 20-kHz tone exposures affected a broad range of frequencies, with the maximum TTS occurring close to 1/2 octave above the exposure frequency. That the effects of a fatiguing exposure may extend to frequencies well removed from, and generally higher than, the actual frequency of exposure has important implications when predicting the effects of anthropogenic sound on animals.

ASSR I/O functions can be represented as the sum of two processes: a low threshold, saturating process and a higher threshold, linear process, that react and recover to fatigue at different rates.

ACKNOWLEDGMENTS

We thank Linda Green and Laura Lewis for animal training support and Dorian Houser for helpful discussions on the manuscript. Financial support was provided by the U.S. Office of Naval Research, Marine Mammal Science and Technology Program. Brian Branstetter was supported by a National Research Council postdoctoral research fellowship.

Abbas, P. J. (**1988**). "Electrophysiology of the auditory system," Clin. Phys. Physiol. Meas. **9**, 1–31.

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).

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.

Benitez, L. D., Eldredge, D. H., and Templer, J. W. (**1972**). "Temporary threshold shifts in chinchilla: Electrophysiological correlates," J. Acoust. Soc. Am. **52**, 1115–1123.

Borg, E. (**1982**). "Auditory thresholds in rats of different age and strain. A behavioral and electrophysiological study," Hear. Res. **8**, 101–115.

Brillinger, D. R. (**1978**). "A note on the estimation of evoked response," Biol. Cybern. **31**, 141–144.

Clark, W. W. (**1991**). "Recent studies of temporary threshold shifts (TTS) and permanent threshold shift (PTS) in animals," J. Acoust. Soc. Am. **90**, 155–163.

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.

Cornsweet, T. N. (**1962**). "The staircase method in psychophysics," Am. J. Psychol. **75**, 485–491.

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., 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. (**1996**). "Auditory evoked responses to amplitude modulated stimuli consisting of multiple envelope components," J. Comp. Physiol., A **179**, 113–121.

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.

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., and Houser, D. S. (**2007**). "Bottlenose dolphin (*Tursiops truncatus*) steady-state evoked responses to multiple simultaneous sinusoidal amplitude modulated tones," J. Acoust. Soc. Am. **121**, 1775–1782.

Finneran, J. J., and Schlundt, C. E. (**2007**). "Underwater sound pressure variation and bottlenose dolphin (*Tursiops truncatus*) hearing thresholds in a small pool," J. Acoust. Soc. Am. **122**, 606–614.

Finneran, J. J., Houser, D. S., and Schlundt, C. E. (**2007a**). "Objective detection of bottlenose dolphin (*Tursiops truncatus*) steady-state auditory evoked potentials in response to AM/FM tones," Aquat. Mammals **33**, 43–54.

Finneran, J. J., London, H. R., and Houser, D. S. (**2007b**). "Modulation rate transfer functions in bottlenose dolphins (*Tursiops truncatus*) with normal hearing and high-frequency hearing loss," J. Comp. Physiol. A (in press).

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., Schlundt, C. E., Dear, R., Carder, D. A., and Ridgway, S. H. (**2002**). "Temporary shift in masked hearing thresholds (MTTS) in odontocetes after exposure to single underwater impulses from a seismic watergun," J. Acoust. Soc. Am. **111**, 2929–2940.

Finneran, J. J., Schlundt, C. E., Carder, D. A., Clark, J. A., Young, J. A., Gaspin, J. B., and Ridgway, S. H. (**2000**). "Auditory and behavioral responses of bottlenose dolphins (*Tursiops truncatus*) and a beluga whale (*Delphinapterus leucas*) to impulsive sounds resembling distant signatures of underwater explosions," J. Acoust. Soc. Am. **108**, 417–431.

Gellerman, L. W. (**1933**). "Chance orders of alternating stimuli in visual discrimination experiments," J. Gen. Psychol. **42**, 206–208.

Henderson, D., Hamernik, R. P., Salvi, R. J., and Ahroon, W. A. (**1983**). "Comparison of auditory-evoked potentials and behavioral thresholds in the normal and noise-exposed chinchilla," Audiology **22**, 172–180.

Houser, D. S., and Finneran, J. J. (**2006a**). "A comparison of underwater hearing sensitivity in bottlenose dolphins (*Tursiops truncatus*) determined by electrophysiological and behavioral methods," J. Acoust. Soc. Am. **120**, 1713–1722.

Houser, D. S., and Finneran, J. J. (**2006b**). "Variation in the heåsensitivity of a dolphin population obtained through the use of evoked potential audiometry," J. Acoust. Soc. Am. **120**, 4090–4099.

John, M. S., Lins, O. G., Boucher, B. L., and Picton, T. W. (**1998**). "Multiple auditory steady-state responses (MASTER): Stimulus and recording parameters," Audiology **37**, 59–82.

John, M. S., Purcell, D. W., Dimitrijevic, A., and Picton, T. W. (**2002**). "Advantages and caveats when recording steady-state responses to multiple simultaneous stimuli," J. Am. Acad. Audiol **13**, 246–259.

Johnstone, B. M., Patuzzi, R., and Yates, G. K. (**1986**). "Basilar membrane measurements and the travelling wave," Hear. Res. **22**, 147–153.

Kastak, D., Schusterman, R. J., Southall, B. L., and Reichmuth, C. J. (**1999**). "Underwater temporary threshold shift induced by octave-band noise in three species of pinniped," J. Acoust. Soc. Am. **106**, 1142–1148.

Kastak, D., Southall, B. L., Schusterman, R. J., and Kastak, C. R. (**2005**). "Underwater temporary threshold shift in pinnipeds: effects of noise level and duration," J. Acoust. Soc. Am. **118**, 3154–3163.

Lins, O. G., and Picton, T. W. (**1995**). "Auditory steady-state responses to multiple simultaneous stimuli," Electroencephalogr. Clin. Neurophysiol. **96**, 420–432.

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.

Mills, J. H. (**1982**). "Effects of noise on auditory sensitivity, psychophysical tuning curves, and suppression," in *New Perspectives on Noise-induced Hearing Loss*, edited by R. P. Hamernik, D. Henderson, and R. J. Salvi (Raven, New York), pp. 249–263.

Mills, J. H., Gengel, R. W., Watson, C. S., and Miller, J. D. (**1970**). "Temporary changes of the auditory system due to exposure to noise for one or two days," J. Acoust. Soc. Am. **48**, 524–530.

Mooney, T. A., Nachtigall, P. E., and Yuen, M. M. L. (**2006**). "Temporal resolution of the Risso's dolphin, *Grampus griseus*, auditory system," J. Comp. Physiol., A **192**, 373–380.

Nachtigall, P. E., Supin, A. Y., Pawloski, J., and Au, W. W. L. (**2004**). "Temporary threshold shifts after noise exposure in the bottlenose dolphin (*Tursiops truncatus*) measured using evoked auditory potentials," Marine Mammal Sci. **20**, 673–687.

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.

Picton, T. W. (**2007**). "Audiometry using auditory steady-state responses," in *Auditory Evoked Potentials: Basic Principles and Clinical Application*, edited by R. F. Burkard, J. J. Eggermont, and M. Don (Lippincott Williams & Wilkins, Philadelphia), pp. 441–462.

Picton, T. W., Skinner, C. R., Champagne, S. C., Kellett, A. J., and Maiste, A. C. (**1987**). "Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone," J. Acoust. Soc. Am. **82**, 165–178.

Popov, V. V., Supin, A. Y., and Klishin, V. O. (**1992**). "Electrophysiological study of sound conduction in dolphins," in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Y. Supin (Plenum, New York), pp. 269–276.

Popov, V. V., Supin, A. Y., and Klishin, V. O. (**1997**). "Paradoxical lateral suppression in the dolphin's auditory system: Weak sounds suppress response to strong sounds," Neurosci. Lett. **234**, 51–54.

Popov, V. V., Supin, A. Y., and Klishin, V. O. (**1998**). "Lateral suppression of rhythmic evoked responses in the dolphin's auditory system," Hear. Res. **126**, 126–134.

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.

Regan, D., and Regan, M. P. (**1988**). "The transducer characteristic of hair cells in the human ear: a possible objective measure," Brain Res. **438**, 363–365.

Ridgway, S. H., Bullock, T. H., Carder, D. A., Seeley, R. L., Woods, D., and Galambos, R. (**1981**). "Auditory brainstem response in dolphins," Neurobiology **78**, 1943–1947.

Salvi, R. J., Henderson, D., and Hamernik, R. P. (**1979**). "Single auditory nerve fiber and action potential latencies in normal and noise-treated chinchillas," Hear. Res. **1**, 237–251.

Saunders, J. C., and Rhyne, R. L. (**1970**). "Cochlear nucleus activity and temporary threshold shift in cat," Brain Res. **24**, 339–342.

Schlundt, C. E., Dear, R. L., Carder, D. A., and Finneran, J. J. (**2006**). "Growth and recovery of temporary threshold shifts in a dolphin exposed to mid-frequency tones with durations up to 128 s," J. Acoust. Soc. Am. **120**, 3227(A).

Schlundt, C. E., Finneran, J. J., Carder, D. A., and Ridgway, S. H. (**2000**). "Temporary shift in masked hearing thresholds of bottlenose dolphins, *Tursiops truncatus*, and white whales, *Delphinapterus leucas*, after exposure to intense tones," J. Acoust. Soc. Am. **107**, 3496–3508.

Schlundt, C. E., Dear, R. L., Green, L., Houser, D. S., and Finneran, J. J. (**2007**). "Simultaneously measured behavioral and electrophysiological hearing thresholds in a bottlenose dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **122**, 615–622.

Schmiedt, R. A. (**1984**). "Acoustic injury and the physiology of hearing," J. Acoust. Soc. Am. **76**, 1293–1317.

Stapells, D. R., Linden, D., Suffield, J. B., Hamel, G., and Picton, T. W. (**1984**). "Human auditory steady state potentials," Ear Hear. **5**, 105–113.

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. Mammals **26**, 83–94.

Supin, A. Y., Popov, V. V., and Mass, A. M. (**2001**). *The Sensory Physiology of Aquatic Mammals* (Kluwer Academic, Boston).

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.

Ward, W. D. (**1962**). "Damage-risk criteria for line spectra," J. Acoust. Soc. Am. **34**, 1610–1619.

Ward, W. D., Glorig, A., and Sklar, D. L. (**1959**). "Temporary threshold shift from octave-band noise: Applications to damage-risk criteria," J. Acoust. Soc. Am. **31**, 522–528.

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.

Observations of potential acoustic cues that attract sperm whales to longline fishing in the Gulf of Alaska

Aaron Thode[a)]
Marine Physical Laboratory, Scripps Institution of Oceanography, San Diego, California 92093-0205

Janice Straley
University of Alaska Southeast, Sitka, Alaska 99835

Christopher O. Tiemann
Applied Research Laboratories, University of Texas at Austin, P.O. Box 8029, Austin, Texas 78713-8029

Kendall Folkert
P.O. Box 6497, Sitka, Alaska 99835

Victoria O'Connell
Alaska Department of Fish and Game, Sitka, Alaska 99835

(Received 13 October 2006; revised 15 May 2007; accepted 21 May 2007)

Sperm whales (*Physeter macrocephalus*) have learned to remove fish from demersal longline gear deployments off the eastern Gulf of Alaska, and are often observed to arrive at a site after a haul begins, suggesting a response to potential acoustic cues like fishing-gear strum, hydraulic winch tones, and propeller cavitation. Passive acoustic recorders attached to anchorlines have permitted continuous monitoring of the ambient noise environment before and during fishing hauls. Timing and tracking analyses of sperm whale acoustic activity during three encounters indicate that cavitation arising from changes in ship propeller speeds is associated with interruptions in nearby sperm whale dive cycles and changes in acoustically derived positions. This conclusion has been tested by cycling a vessel engine and noting the arrival of whales by the vessel, even when the vessel is not next to fishing gear. No evidence of response from activation of ship hydraulics or fishing gear strum has been found to date. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2749450]

PACS number(s): 43.80.Nd [WWA] Pages: 1265–1277

I. INTRODUCTION

Sperm whales (*Physeter macrocephalus*) are distributed throughout the world's oceans and are considered an endangered species in U.S. waters.[1–6] The current population in the North Pacific is unknown, although acoustic recordings from bottom-mounted recorders suggest a year-round presence.[7] While females and immature individuals are known to reside at low latitudes,[6] adult males are known to travel and forage at higher latitudes in both hemispheres.[6,8–12] The diet of these deep-diving animals primarily consists of various species of cephalopods, based on an analysis of stomach contents.[6,13–19] However, in certain regions fish seem to comprise part of the diet as well,[4,6,15,18,20] including the eastern Gulf of Alaska,[19] but it is unknown what fraction of this population's diet consists of fish.

The sperm whale is the largest marine mammal known to depredate on human fishing activities. While the vast majority of reports of cetacean depredation involves killer whales, pilot whales, and other smaller odontocete species,[21–24] depredation activities by sperm whales have received increasing coverage in scientific literature.[21–23,25–29] This species has been associated with fishing operations, particularly demersal longline operations, in a number of locations around the globe,[6,21,25–27] including Norway, Greenland, eastern Canada (Labrador and Newfoundland), Chile, and the Falkland Islands. Although quantitative data are not available, anecdotal accounts suggest an increasing trend in sperm whale depredation.

In the eastern Gulf of Alaska (GOA) an active longline fishery for sablefish *Anoplopoma fimbria* (also called blackcod and butterfish) continuously occurs from late February through mid-November. Sablefish occur on the continental slope and most commercial longliners fish for this species in water depths between 400 and 1000 m. The continental shelf off the Kruzof and Baranof islands is very narrow; consequently, these sablefish grounds are relatively close to shore, within 12 to 20 miles (Fig. 1). In the GOA, depredation of longline gear set for sablefish by sperm whales has been occurring since at least 1978 in the domestic U.S. fishery, and observers on Japanese longline vessels in the Gulf of Alaska reported depredation occurring in the mid 1970s. This fishery occurred year-round until the early 1980s, when fleet expansion resulted in a shortened season. By 1994, the entire quota was caught in 10 days. In 1995 individual fishing quotas were implemented, reducing overall effort while maintaining an 8.5-month open season. This extended season apparently provided more opportunities for sperm whales to depredate longline gear, and by 1997 reports of depredation had increased substantially. A domestic sablefish survey in

a)Electronic mail:athode@mpl.ucsd.edu

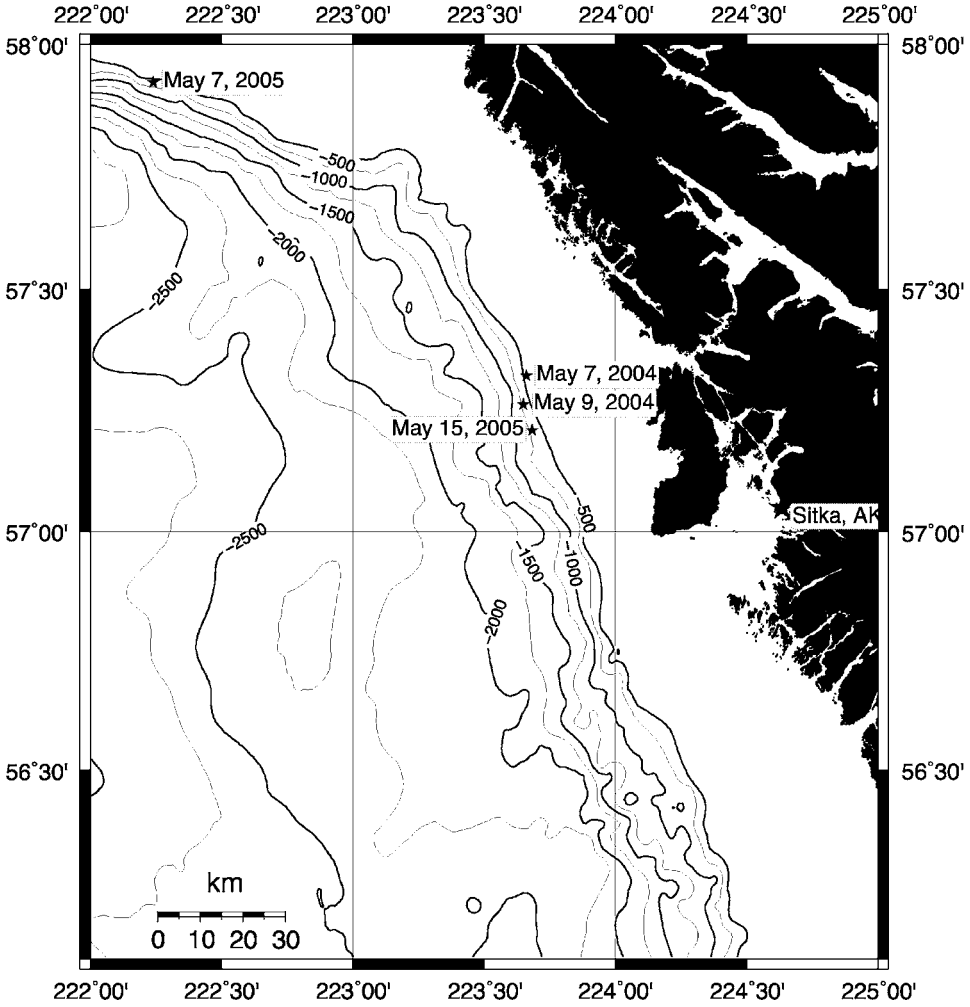


FIG. 1. Locations of four experimental sites discussed in the paper. All sites are along the continental slope off Sitka, AK. Bathymetry contours are in meters, with 250-m intervals.

the GOA looked at catch rates from 1999 to 2001 for all sets with sperm whales present; they compared boats with and without evidence of depredation and found a 5% lower catch rate in boats with depredation.[29]

While local Sitkan longliners have observed sperm whales following fishing vessels to deployment sites, they also often observe whales arriving after a haul begins, raising the question as to whether the animals are responding to distinctive visual or acoustic cues inadvertently produced by the activity. An example of a potential visual cue is the flocking of tens to hundreds of seabirds to a fishing haul site, and popular hypotheses for acoustic cues have included propeller cavitation, activation of auxiliary hydraulic systems to haul gear, echosounders, and strum noise produced by the vibration of the taut gear line as it is hauled out of the water. To our knowledge little to no acoustic monitoring has been conducted to observe or test potential acoustic cues for most marine mammal species, with the exception of Refs. 30 and 31.

While much of their foraging behavior cannot be observed directly, sperm whales are acoustically active underwater, and during a single dive one individual can make thousands of impulsive sounds called "clicks"[32–34] over a typical 45-min length dive. Measurements in other areas of the world have found that about 10–15 min before returning to the surface, an animal typically falls silent.[18,35] Thus passive acoustic monitoring of an animal's vocalizations can yield an estimate of the animal's dive cycle, even if the animal is not observed at the surface. Other statistics on the sounds' rhythm and internal characteristics can be collected as well. Furthermore, under certain circumstances these clicks generate multipath returns from the ocean surface and bottom that can be used to derive an animal's depth and range from the hydrophone, provided that the ocean depth is known. The technique has been previously used in the Gulf of Mexico to track the dive profiles of female sperm whales,[36] as well as in the Mediterranean Sea.[37] A companion paper discusses how this multipath can be used to track sperm whales off Sitka.[38]

This natural acoustic activity has provided an opportunity to observe correlations, and in some cases direct effects, of various types of potential acoustic cues on the acoustic activities of whales in the vicinity. Section II describes how demersal longline deployments can be converted into noninvasive listening posts by attaching compact autonomous passive acoustic recorders to the anchorlines of the deployment, and then discusses acoustic data analysis and tracking procedures. Section III describes the circumstances behind three independent encounters of sperm whales with instrumented longline gear sets between 2004–2005, of which two permitted some form of "controlled cue" hypothesis testing.

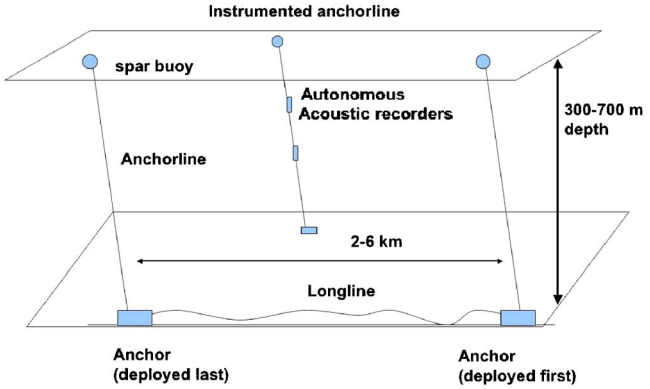


FIG. 2. (Color online) Schematic view of longline deployment, including instrumented anchorlines deployed separately from longline.

Finally, Sec. IV discusses the cumulative evidence for and against various hypothetical acoustic cues.

II. EXPERIMENTAL PROCEDURE

A. Equipment and deployment

Acoustic data presented here were collected from autonomous acoustic recorders, or "Bioacoustic probes," designed and built by Greeneridge Sciences Inc.[39] These instruments could sample acoustic data at sampling rates of 100 Hz to 20 kHz, using an HTI-96-MIN/3V hydrophone (typical sensitivity of −172 dB $re:1\ \mathrm{V}/\mu\mathrm{Pa}$) and storing the data to 1 GB of flash memory with 16-bit precision. For the data presented here, the data sampling rates varied between 8192 and 20105 Hz. The unusual sampling rates are a consequence of the low-level hardware requirements of the electronics. Additional auxiliary measurements of pressure, temperature, and acceleration on two axes were sampled once a second and also stored to memory. Four AAA batteries were found to provide sufficient energy to fill the memory. All components except for the hydrophone were inserted into a transparent acrylic pressure case with a Delrin end-plug, manufactured by Cetacean Research Technology in Seattle, WA. The resulting length and diameter of each recorder is 25 cm and 5 cm. The hydrophone is connected to the internal electronics via a Subconn underwater connector.

Figure 2 shows a schematic of a demersal longline deployment, once a vessel has left the area. The longline itself lies along the ocean bottom over a typical distance of a few nautical miles, typically at depths between 300 and 700 m. At each end of the longline a 35-kg anchor is used to fix the ends, and from each anchor an "anchorline" rises to the surface, attached to a spar buoy. To recover the line, the fishing vessel transits to the upstream buoy, and a deckhand pulls the anchorline over a set of rollers mounted on the side, wrapping the anchorline around a hydraulic winch, which then pulls the anchor and longline off the floor. As the hydraulic systems on these vessels are typically only activated just before a haul begins, the acoustic tones made by such a system have been a popular hypothesis for a potential acoustic cue. Once the anchor has been retrieved, the vessel attempts to drift with the current, while continuing to winch the longline aboard. Often the vessel captain has an auxiliary set of steering controls next to the rollers, which he/she will use to engage the engine during a haul in order to permit fine-scale control of the vessel.

During a typical instrumented deployment, two autonomous recorders are attached to a third anchorline, deployed before beginning the actual longline deployment, and recovered once the haul is complete. The instrumented anchorline is generally deployed within 1 km of the upstream anchorline, with recorder depths between 100–200-m depth, as far from the ocean surface as practical given the structural strength of the pressure cases. Given the large scope of the anchorlines, the actual deployment depths can vary considerably and must be logged from the pressure transducers. Flow noise was an initial concern, but it was found that continuous flow noise was only significant at frequencies below 50 Hz, although one significant exception will be discussed in Sec. III E. Visual observers record all major vessel and bird activities sighted from either the longlining vessel or from a small sport fishing vessel chartered for the day, and auxiliary acoustic data have been recorded from a hydrophone deployed 10–20 m beneath the bow of the fishing vessel.

B. Single-hydrophone analysis

Once acoustic data from an encounter have been transferred to hard disk in WAV format, three low-level acoustic analyses are performed on the data: sperm whale click detection, interclick interval (ICI) estimation, and source sound exposure level rate (SASELR) estimation of the fishing vessel acoustic output.

To detect sperm whale clicks the acoustic software analysis program ISHMAEL (Ref. 40) scans the record using the "energy detection" feature and activates a MATLAB script to process each detection. ISHMAEL computes the audio spectrogram, "equalizes" the spectrogram levels by subtracting a time-averaged background noise spectrum, and then integrates the squared modulus of the pressure spectrum between 100 Hz to 80% of the Nyquist frequency of a given recording.[7] Whenever this integrated value exceeds a threshold of 1.5, the MATLAB script logs the pulse time, amplitude, and duration. Upon completion of an ISHMAEL run, a second MATLAB script then consolidates the detection data into histograms of click rate. The effects of acoustic multipath were removed, to first order, by accepting only pulses that were not followed by another pulse within 0.2 s. Some acoustic multipaths can arrive later than 0.2 s after the main pulse, but the detection threshold would be set so that these weaker arrivals are generally not detected. However, some multipath arrivals are still accepted by the detector, so the click counts here may be biased toward an overcount.

Another useful parameter that could be automatically extracted from the raw acoustic data is the interclick interval, or the interval between two consecutive direct path click arrivals from the same sperm whale. The ICI is automatically estimated for each detected click by hypothesizing a range ICI values between 0.1 and 2 s, and then examining the subsequent time series to determine whether pulses are present at three predicted times after the click in question.[41] Best

estimates were obtained whenever multipath arrivals were first removed from consideration, as discussed previously. The ICI can be useful in distinguishing sperm whale clicks from other random pulsive sounds.

Finally, to characterize the acoustic output of the fishing vessel, the square modulus of the sound-pressure level is integrated between frequency ranges dominated by vessel noise when sperm whale clicks were absent, which in this paper will be between 250 and 1000 Hz. The integrated levels were then averaged over 5 s to produce an estimate of what is defined here as the "average sound exposure level rate" (ASELR), with units of uPa^2. In more common terminology the ASELR is the ensemble-averaged "power spectral density level" (PSDL),[42] integrated over a given frequency bandwidth. This term ASELR is used here because this quantity is not really an acoustic intensity or power measurement, as a true measurement of acoustic intensity requires an independent measure of the acoustic particle velocity.[43] Instead, if the ASELR is multiplied by a time interval, one obtains a bandlimited quantity defined as a "sound exposure level," (SEL) or "energy flux density," which has been argued to be a biologically significant metric of the acoustic field.[44]

As the GPS position of the fishing vessel relative to the instrumented anchorline is known to within 10 m (0.1% of typical vessel range), the received ASELR can be corrected for vessel slant range to produce an estimated "source level" SEL at 1-m range, or SASELR, with units of uPa^2 @ 1 m. The SASELR permits fundamental changes in the vessel acoustic signature to be separated from simple changes in vessel translational position. In all figures that follow the SASELR will be plotted, using a spherical spreading assumption if the slant range is less than the ocean depth, and using a cylindrical spreading assumption if the slant range is greater than the water depth. For the latter case the SASELR is defined to be $2\pi RD^*$ ASELR, where R is the vessel's horizontal range from the instrument and D is the local water depth.

C. Acoustic tracking procedures

Whenever feasible, one of two types of acoustic tracking is conducted. For situations where the bottom bathymetry is well characterized out to ranges of 2 km from the recorder, the relative arrival times of the acoustic multipath can be used to estimate the 3D position of the whale over time. This analysis, which is featured in Sec. III C, is the subject of a companion paper.[38]

Unfortunately, accurate bathymetry information is often not available, so in one of the 2005 deployments to be discussed, an alternative array geometry used two instrumented anchorlines deployed 4.9 km apart, at opposite ends of a longline deployment. The autonomous recorders were activated and time-synchronized before and after the deployment, and a linear clock drift was assumed to derive the time offset at all times in between. After processing each station's data stream using the pulse detection procedure outlined above, "direct path" detections were designated whenever the arrival in question is not preceded by another detection within a time $t_{\min}$. The selected direct-path arrivals were then matched between the stations by comparing the ICI patterns at both stations, using N direct-path detections following the pulse in question, and using a time tolerance of 25 ms for matching the arrivals.[45] For instruments spaced 4.9 km apart, good values of $t_{\min}$ and N were 0.4 s and 16, respectively. The resulting relative time-of-arrival (TOA) values fix the whale position to a locus of points that form an "isodiachron,"[46] which becomes a hyperboloid surface if a homogeneous sound speed is assumed throughout the water mass. Even if the sound of the vessel cannot be recorded on both stations, given a vessel's GPS position an "effective" vessel TOA can be computed and plotted against the TOA of whale clicks, and some information about changes in the animals' position relative to the vessel can be inferred. If two recorders are deployed at the same location, but at different depths, then false matches can be eliminated by comparing whether the TOA estimates from each recorder for a given click match to within 0.5 s. This latter technique is used in Sec. III E.

III. RESULTS

A. Overview

Passive acoustic measurements of sperm whale depredation activity began in 2004, with an initial goal of observing and identifying potential acoustic cues produced by hauling longliners. Sec. III B describes near-field measurements of acoustic signatures of the engine and hydraulic systems of a fishing vessel, taken on 7 May 2004, and Sec. III C discusses the first complete acoustic observations of two sperm whales arriving in the vicinity of a longline haul on 8 May 2004.

By 2005 potential acoustic cues had been identified, and the level of coordination and cooperation between the SEASWAP fishermen and researchers had reached a level where limited hypothesis testing became feasible during opportunistic encounters at sea. Sec. III C discusses how a potential hydraulic cue was tested during an 8-h sperm whale encounter on 7 May 2005, while Sec. III D describes the results of an engine cue test conducted on 15 May 2005, utilizing two instrumented anchorlines to permit crude localization estimates. All 2004 and 2005 encounters took place close to the continental shelf break near Sitka (Fig. 1).

B. Acoustic signatures of a fishing vessel

On 7 May 2004 the 58-ft. fishing vessel KELLEY MARIE volunteered to approach an instrumented anchorline during a time when no whales were present, engage and disengage the engine, and then activate the hydraulic system that is used to power the haul winch. The engine was a 6-cylinder diesel with 250 horsepower, and the propeller had five blades. Figure 3 shows a spectrogram (of the square modulus of acoustic pressure in units of power spectral density, or dB *re*: 1 uPa^2/Hz) of the propeller cavitation noise and winch hydraulic system as the vessel passed within 10-m horizontal range of a 100-m-deep hydrophone mounted on the anchorline, sampling at 15 019 Hz. At 10 s the vessel put the engine in neutral, and at 22 s the ship's hydraulics were activated, producing the tone visible at 190 Hz. The broadband cavitation signal from the ship's propeller is also clearly visible,

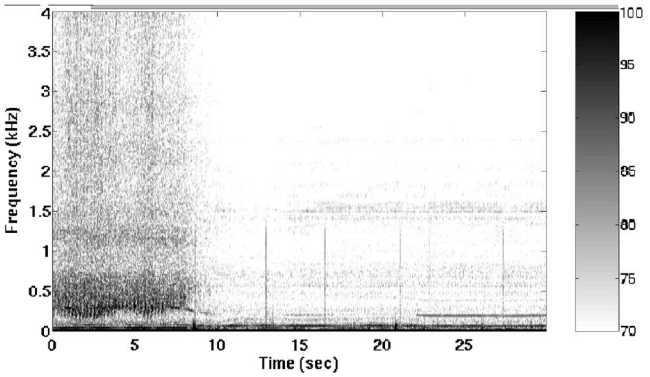


FIG. 3. Spectrogram of F/V KELLY MARIE, measured at 13:21:14, 7 May 2004, at a depth of ~100 m directly underneath the hull, using FFT size of 1024 with 25% overlap. The gray scale shows the square modulus of the acoustic pressure in units of power spectral density (dB *re*:1 uPa^2/Hz). Cavitation noise from the propeller is visible between 0 and 10 s, and the hydraulic system to power the hauling winches has been activated at 22 s, generating the 190-Hz tone visible in the spectrogram. The thin vertical lines between 0 and 1.25 kHz are not sperm whale clicks.

with the largest spectral density levels lying between 250 and 1000 Hz, but with significant detectable levels past 4 kHz. (The vertical lines between 0–1.25 kHz are not sperm whale sounds). The measured ASELR for the engine cavitation was 110 dB *re*:1 uPa^2 between 250 and 1000 Hz, and 95 dB *re*:1 uPa^2 for the hydraulic system between 150 and 250 Hz, yielding effective signal-to-noise ratios of 20 and 6 dB, respectively. A predicted spherical spreading transmission loss of 44 dB yields respective SASELR values of 153 and 139 dB *re*:1 uPa^2 @ 1 m. Interference effects from the Lloyd's mirror phenomenon have been ignored in the transmission loss computation, but simultaneous measurements by a second hydrophone at 195-m depth generates the same SASELR values to within 3 dB.

C. Acoustic measurements of sperm whales approaching a hauling vessel

The first acoustic measurements of sperm whales interacting with a hauling longliner began the morning of 9 May 2004, during a longline recovery by the F/V COBRA off a local promontory at the edge of the continental shelf. The previous day the COBRA had deployed the gear and departed to let the longline "soak" overnight. At 07:55 the next morning an instrumented anchorline was deployed in 460-m-deep water, 1.6 and 1.5 km from the two original anchorlines, with two recorders attached at 83 and 155 m depth. If one were east of the deployment looking west, one would see a similar deployment geometry as shown in Fig. 2. After slowly circling the area for an hour, the vessel retrieved the first anchorline buoy at 9:04, and by 9:16 the buoy anchor was on deck. The fish haul began immediately afterward, but sperm whales were not sighted until 10:08, when a sperm whale surfaced approximately 50 m away from the vessel, followed 3 min later by a second whale surfacing. Both animals dove around the vessel vicinity until the recovery of the second anchorline buoy at 11:00, and then proceeded to follow the vessel back toward the instrumented anchorline. The

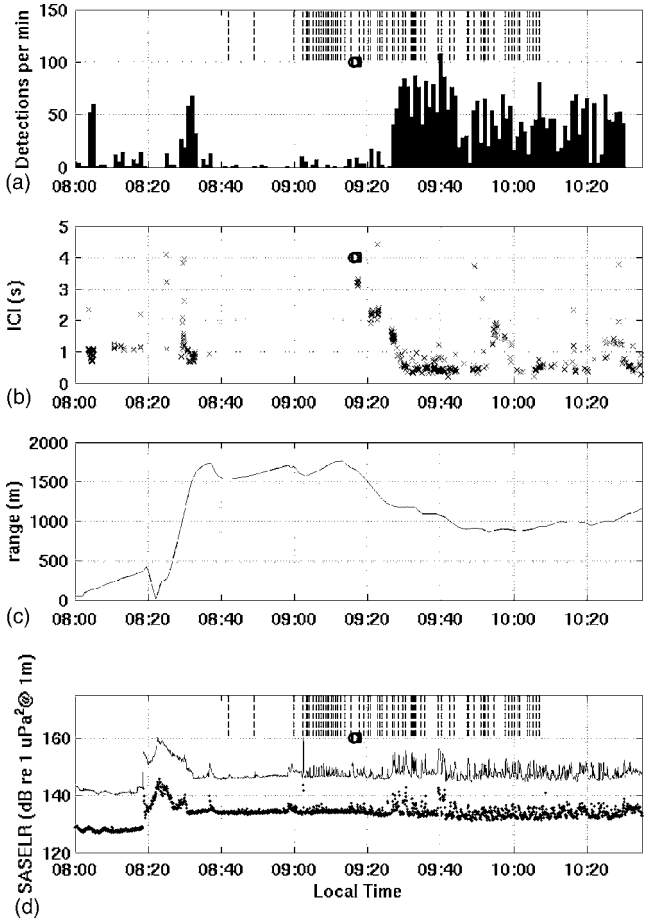


FIG. 4. Beginning of 9 May 2004, encounter, between 8:00 and 10:30 AM. (a) Histogram of pulsive sounds detected per minute. Vertical dashed lines indicate presence of acoustic signatures of an engine engaging and disengaging the propellor. The circle indicates time at which an anchor is dropped on deck (anchorline on board), and the square indicates the start of substantial sperm whale acoustic activity at 09:17:01; (b) ICI of sperm whale sounds detected on the instruments; (c) horizontal range of fishing vessel from instrumented anchorline buoy; (d) source-averaged sound exposure level rate (SASELR), in units of dB *re*:1 uPa^2 @ 1 m, averaged over 5-s intervals, integrated between 250 and 1000 Hz (solid line) and 150 and 250 Hz (dashed line, shifted −10 dB for clarity). Received levels have been adjusted by measured vessel slant range to produce effective source levels at 1-m range.

COBRA then drifted for 2 h before finally hauling the instrumented longline around 13:00.

The acoustic record of the beginning of the encounter, displayed in Fig. 4, shows substantial sperm whale acoustic activity. Subplot (a) shows a histogram of sperm whale clicks detected per minute, computed as described in Sec. II B, by integrating the squared pressure modulus between 500 and 7500 Hz (sampling frequency 15 019 Hz, 256-pt FFT, 1/16 overlap), and using a preset threshold of 1.5 in ISHMAEL.

Subplot (b) shows an estimate of the interclick interval (ICI) derived via the procedure in Sec. II B. From both the ICI and detection plots, it is clear that within minutes of the instruments' entering the water, sperm whale activity was detected in the area (8:05 AM). The sounds lasted for 2 min, had no acoustic multipath, and had a steady ICI of about 1 s. Five minutes of sperm whale clicks were also detected around 8:30, also with an ICI of about 1 s. These ICIs are typical of natural sperm whale foraging behavior found in

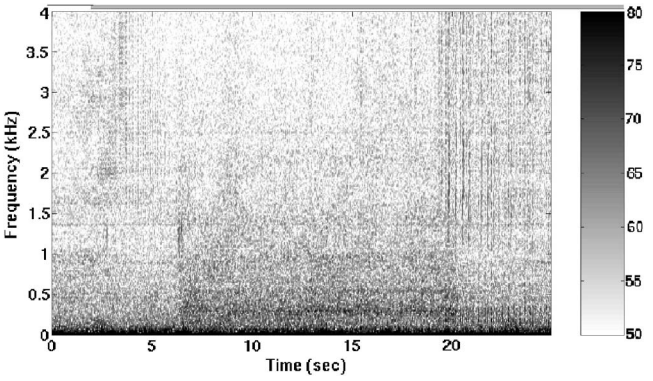


FIG. 5. Example of "engine cycling" as fishing vessel fine-tunes its position relative to the longline, during a time (9:51:03) that the vessel is closest to the acoustic recorder during the haul (900-m range). The engine is engaged at 6 s and disengaged at 20 s, generating broadband cavitation noise visible up to 4 kHz. Sperm whale clicks are visible between 1 and 4 kHz throughout the figure.

this area and at other high-latitude locations.[18,47] The lack of multipath indicates that the animals are greater than the few-kilometers range.

Subplot (c) shows the range of the vessel from the instrumented anchorline, and thus from the hydrophones' approximate position, while subplot (d) displays the estimated SASELR of the received acoustic field at 155-m depth, with the solid line representing integrated square pressure between 250 and 1000 Hz. The SASELR levels have been derived from the vessel range from the instrumented longline shown in (c), assuming a cylindrical spreading transmission loss, which seems to adequately model the propagation environment in that the final SASELR curve remains at a steady level between 9:00 and 10:20 even as the vessel range decreases from 1800 to 900 m. The dotted line represents the SASELR measured between 150 and 250 Hz, the region where a hydraulic winch tone would be expected. The exact time that the hydraulic system was switched on was not noted, but it was approximately 9:00 AM, a few minutes before the first anchorline retrieval, and the system remained on until the end of the haul. However, the SASELR over the hydraulic frequency band shows no sudden, permanent jump at this time, and a careful review of the acoustic record around 9:00 AM confirms the absence of any distinctive hydraulic signature at 1.6-km detection range.

However, an interesting feature in the SASELR appears as the vessel begins to haul the anchorline at 9:04. The 250-Hz–1-kHz curve displays a series of short-duration peaks that change the SASELR by 3–5 dB between 9:05 and 9:20. (The cycling continues after this time, but numerous sperm whale clicks contaminate the SASELR curve.) The short-term peaks beginning at 9:05 arise from a particular method of handling the vessel in order to keep the winched longline vertical. Generally a longliner tries to keep the engine in neutral and drift with the current while hauling the line. Often, however, due to snags, currents, or delays in gaffing fish, the line will begin to angle underneath the hull of the vessel. Under this circumstance the engine is briefly engaged for 5–10 s to swivel the vessel around the line, the result being a cavitation bubble cloud. Figure 5 shows an example of

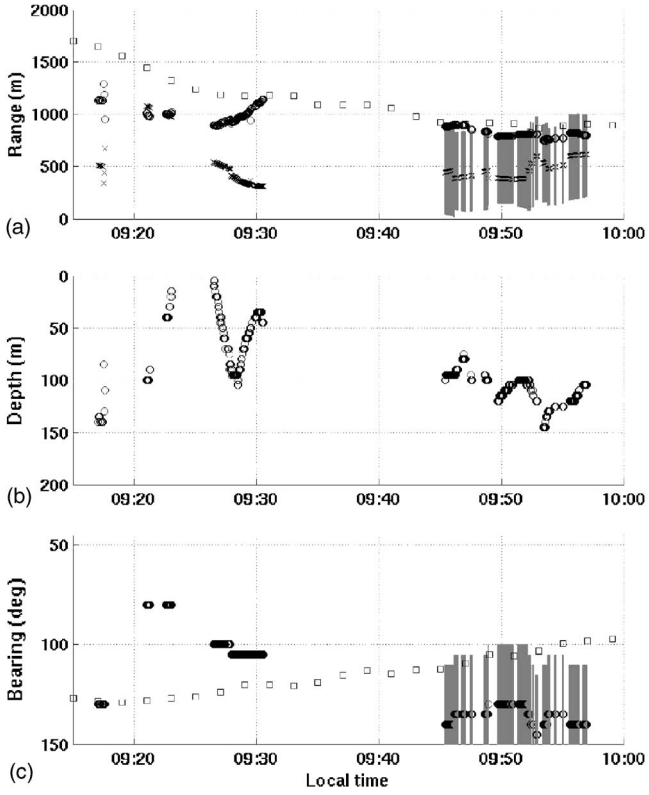


FIG. 6. (a) range; (b) depth, and (c) bearing of sperm whale clicks (circles) relative to instrumented anchorline, between 9:15 and 10:00. Bearings are with respect to true north. Subplots (a) and (c) also show the fishing vessel range and bearing (squares), respectively. Subplot (a) also displays the derived horizontal separation between sperm whale and vessel (crosses). Gray regions indicate bounds of uncertainty in whale azimuth and horizontal distance from vessel after 9:45, due to lack of bathymetry variation vs. azimuth in that region. The sperm whale clicks before 9:20 seem to be from an animal different than the one clicking after 9:20.

what a spectrogram of this signal appears like, taken at 9:51:03, or 34 min after substantial sperm whale activity has begun, and when the vessel is 900 m from the instrumented anchorline. Figures 4(a) and 4(d) use vertical dashed lines to mark discrete times when this activity occurs while hauling the anchorline.

At 9:16:14 the anchorline has been completely recovered, and the anchor that is attached to one end of the longline has been dropped on deck, producing an audible tone detected on the recorders 1800 m away (marked by a circle in the subplots). At 9:17:01 sperm whale clicks are again detected (black square in the subplots), but the clicks display two important differences from the sounds previously detected around 8:05 and 8:30.

First, there are considerable amounts of time-separated multipath present in the signal, enough to permit tracking of an animal in range and depth. Figure 6 shows the range, depth, and bearing of the clicks derived from this multipath, relative to the hydrophones, using data analyzed in Ref 38. The range and bearing of the F/V COBRA are also displayed, as well as the derived horizontal separation between whale and vessel. The clicks detected at 9:17 seem to arise from a different whale at a different spatial location than those made after 9:20, since the best-fit bearing for the former sounds is nearly 130 deg, and the latter 80 deg. If the sounds were

made by the same animal, it would have had to cover 790 m in 4 min, or 3.3 m/s (6-knot) mean speed, inconsistent with the speeds and directions derived after 9:20. However, later bearing estimates suggest that after 9:20 an animal is converging on the location of the fishing vessel, arriving about 300 m north of the vessel at 9:31, when the echolocation clicks stop for a few minutes. A second track obtained after 9:45 finds that the range of the whale and the fishing vessel lies within 50 m, but unfortunately the bathymetry profiles lying between an azimuthal arc of 100–150 deg are sufficiently similar such that the whale's azimuth can only be determined as being somewhere between the gray bars in subplots (a) and (c). Thus, while the whale may also be at the same azimuth as the vessel, the convergence of the whale and vessel positions cannot be proved.

The second difference between the clicks detected before 9:17 and those afterward is that the ICI for the latter is very high—3.4 s when the sounds begin [subplot (b) of Fig. 4]. Over the next 10 min, bouts of clicking are detected with minutes of silence in between, and the ICI steadily decreases, until at 9:26 continuous clicking starts and the ICI drops to 0.5 s or less for the rest of the encounter.

Finally, note that after 9:26 the sperm whale clicks contribute significant energy to the SASELR function in Fig. 4(d), at a level at least 3–6 dB greater than the cavitation sounds of the fishing vessel. The visual sighting of the whales by the vessel by 10:00 indicates that relative SASELR levels between vessel and whale can be justifiably compared in this manner. The fact that whales around a vessel can produce SASELR levels greater than the vessel itself has implications to be discussed in Sec. V.

D. Testing hydraulic and engine cues on a single whale around an anchorline

In 2005 potential acoustic cues began to be tested during opportunistic encounters. The first testing opportunity arose on 7 May 2005, when the F/V COBRA (the same vessel as in Sec. III C) traveled to the Spenser Spit, approximately 60 nautical miles northeast of Sitka (Fig. 1). After deploying a longline at 20:11, the COBRA moved about 1 km away to deploy an instrumented anchorline ("buoy 1") by 21:11. At 21:22 a sperm whale was sighted swimming directly toward the vessel, and within minutes was circling around the vessel at a radius of less than 50 m. The COBRA set its engines into neutral and began drifting, dropping an additional instrumented anchorline (buoy 2) at 22:10, 1.1 km from buoy 1. As night fell at 22:52 whale was observed to be swimming in circles around buoy 2, occasionally diving within 20 m of the spar buoy.

For several hours the vessel continued to drift away from the instrumented anchorlines. At 02:49 the following morning the COBRA finally engaged its engines to move toward buoy 2, and when the spar buoy was sighted in the sodium lamps, the engines were placed into neutral at 3:09. At 3:35 the whale was sighted in front of the vessel's sodium lamps, but visual contact was lost as the vessel drifted away from buoy 2 once again. The situation seemed auspicious for testing various acoustic cues, so at 4:19:25 the winch hydraulics were engaged for 3 min while leaving the engines in neutral. Finally, at 4:45:20 the propeller was re-engaged and the vessel moved back to buoy 2 to mimic a longline recovery.

Figure 7 summarizes the acoustic behavior of this lone animal throughout the night, as measured from a recorder sampling at 8.192 kHz on buoy 1. (Unfortunately, the recorder on buoy 2 failed to record.) Subplot (a) shows the number of acoustic detections per minute logged by the sensor, using a spectral integration range between 500 and 3.5 kHz.

Between 22:00 to shortly after 23:00 the time intervals of vocal activity versus silence are short and irregular. However, by 23:15 the animal had settled into a pattern of long intervals of vocal activity averaging 38 min, followed by an average of 16 min of silence. Shortly before 3:00, just after the COBRA's engines had been engaged and the vessel was moving back to buoy 2, the cycles of acoustic activity and silence become irregular once again. At 3:55 the animal seems to resume an extended cycle of acoustic activity, which is not interrupted by the activation of the vessel's hydraulic system at 4:19 (marked by black circle). From Fig. 7(c) one notes that transitions to "normal" dive cycle behavior (23:10 and 3:58) correspond to times when the vessel drifts more than 800 m away from buoy 2.

Subplot (b) quantifies this discussion by plotting the "acoustic cycles" of the sperm whale over this time period. An acoustic cycle is defined here as the time interval between acoustic gaps, where a gap in turn is defined as a period of time where the number of acoustic detections averaged over a 3-min interval is less than 30 clicks per minute (i.e., an ICI greater than 2 s). As mentioned in the Introduction, acoustic cycles are associated with the start of the dive cycle and an animal's foraging time at depth. The periods of silence correspond roughly with the animal's ascent and rest at the surface (e.g., Refs. 18, 35, and 48).

Subplot (b) also quantifies the variance in natural acoustic cycles measured from sperm whales in this area, using recordings of natural acoustic activity of sperm whales collected on 8 and 11 May 2004, and 25 April 2005, yielding measurements of 15 complete natural acoustic cycles. The median of this natural distribution is 25 min, and is plotted as a horizontal line in subplot (b), along with two dashed lines that indicate one standard deviation of 11 min above and below this median.

One sees that, after the initial encounter with the whale early in the evening, the observed acoustic cycles lie at or above the median acoustic times recorded under natural conditions until shortly after 3:00, the time at which the boat's engines have been engaged and disengaged. Once the engine has been disengaged and the vessel is within 100 m of the buoy, the animal displays five acoustic cycles of 10 min or less, before reverting to acoustic cycles consistent with both its earlier behavior and the other results from naturally foraging animals. Neither the hydraulic activation nor second engine engagement are associated with consistently short acoustic cycles.

The SASELR source level between the 250-Hz and 1-kHz band has been computed in Fig. 7(d). As the vessel's engines are disengaged most of the night, much of the acous-

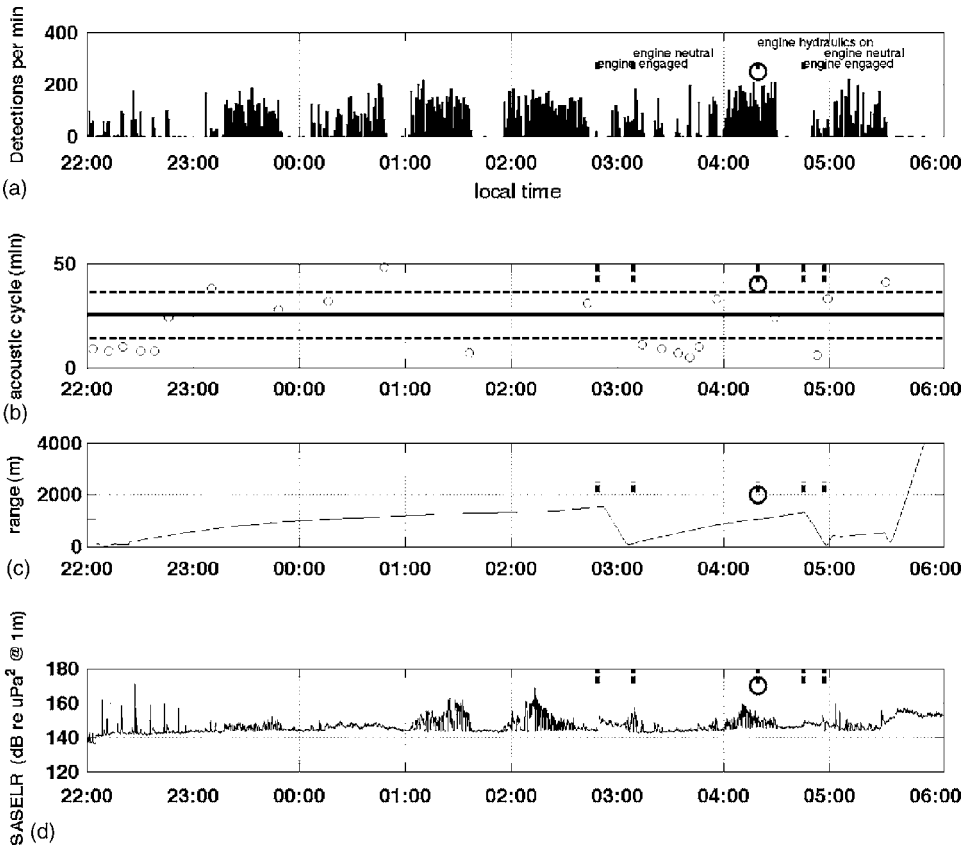


FIG. 7. 7 May 2005, overnight encounter, between 21:00 and 6:00 local time. Local time in hours and minutes is plotted on the *x* axis. (a) Histogram of pulsive sounds detected per minute. Vertical dashed lines indicate times at which boat engine was engaged and disengaged, as well as a time that the ship hydraulics were activated for 3 min, with the engine set in neutral (04:19:25). A black circle marks the time of the hydraulic system test. (b) Duration of sperm whale acoustic cycle in minutes (see the text for definition). Circles display the start time and duration of a cycle, the horizontal solid line indicates the median acoustic cycle culled from acoustic measurements of natural foraging behavior in the area, and the horizontal dashed line indicates the standard deviation of the natural acoustic cycles; (c) horizontal range of fishing vessel from instrumented anchorline buoy (buoy 2); (d) SASELR (dB *re*:1 uPa^2 @ 1 m) averaged over 5-s intervals, integrated between 250 and 1000 Hz (solid line) and 150 and 250 Hz (dashed line).

tic energy detected is actually associated with the sperm whale. However, it can be seen that at the two times when the boat engine is engaged the SASELR jumps 6 dB above the ambient noise background. As in Sec. III B, no hydraulic acoustic signal was detected at 1-km range either via the SASELR plot or visual or aural monitoring of the data.

E. Testing of engine cues on multiple whales with multiple acoustic sensors

On 15 May 2005, 8 days after the previous section, another whale encounter with the COBRA took place, providing an opportunity to test whether the vessel cavitation noise initially observed in Sec. III B is associated with changes in vocal behavior. Furthermore, two instrumented longlines were deployed simultaneously, permitting relative time-of-arrival measurements and thus providing a rudimentary acoustic tracking capability over long distances, as discussed in Sec. II C. Figure 8 shows a map of the deployment geometry off the continental shelf, in an area only a few kilometers away from the first encounter discussed in Sec. III B. At 10:50 an instrumented anchorline (buoy 1) was dropped with one recorder attached at 92-m depth, and from 11:15 to 11:55 a longline was deployed beginning from a location (anchor 1) 640 m south from buoy 1, and ending at anchor 2. The COBRA then deployed a second instrumented anchorline (buoy 2) 1.2 km east of anchor 2 at 12:40, with two recorders attached at approximately 100-and 200-m depth, respectively. Buoy 2 was thus 4.9 km NW from buoy 1. Both instrumented anchorlines had a lead weight attached beneath the recording instruments to prevent substantial inclination of the recorders, in order to permit more predictable deployment depths.

The COBRA then traveled 3 km to the NE to make a detailed bathymetry map of the area where the encounter in Sec. III B occurred, finishing by 15:30 (labeled "Survey 15:30" in Fig. 8). The vessel then traveled back to buoy 1 at 6 knots, passing within 400 m of buoy 1. At 16:08 the vessel put its engine in neutral and began to drift. Throughout the morning and afternoon no whales had been sighted, but a hydrophone dropped overboard at 16:10 detected sperm whale clicks at 16:13, and a decision was made to cycle the engine to simulate a haul, while keeping the hydraulic system off. The engine cycling began at 16:17:30 when the vessel was 1.1 km away from anchor 1 (labeled "Engine test" in Fig. 8). At 16:21:57 a sperm whale surfaced 20 m away from the vessel, and began a dive at 16:29:06. By 16:30 two whales had been sighted next to the vessel, and by 16:37 the first albatross were sighted approaching the vessel. The engine cycling continued until 16:48, when the vessel reengaged the engines and started to move toward anchor 1, now about 1.6 km away, to begin a haul of the anchorline. Over 100 albatrosses had settled by the vessel by 17:10, marking the first large aggregation of birds encountered during the test, and thus the first time a potential visual cue was available. Once the haul began at least three distinct sperm whales had been identified.

Figures 9 and 10 display the relative time-of-arrival (TOA) measurements of sperm whale acoustic pulses between buoys 2 and 1, using the methods discussed in Sec. II C. Mapped over the data is the modeled vessel TOA, which is not derived from the acoustic record, but computed

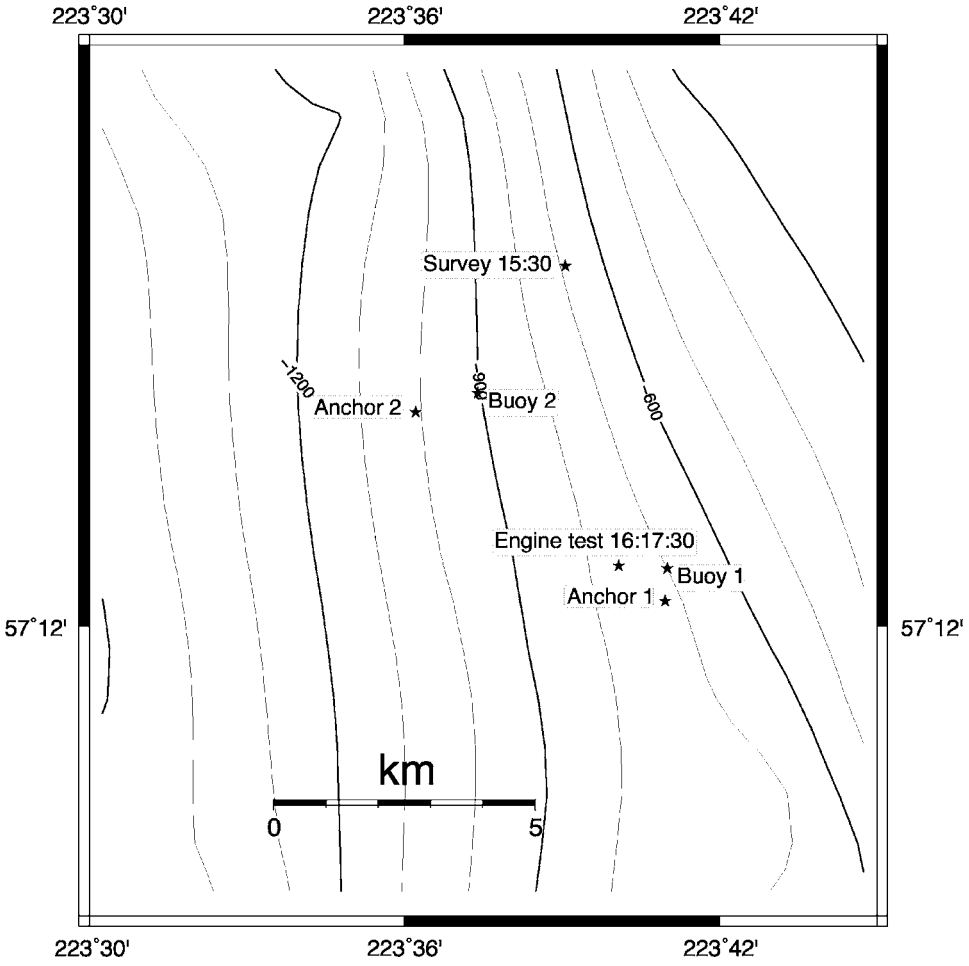


FIG. 8. Bathymetry map of region surrounding engine cycling test on 15 May 2005. A "buoy" marks an instrumented longline, and the "anchor" points mark the ends of the longline deployment. The publicly available bathymetry shown here is only accurate at 200-m depths or less due to the low spatial resolution of the data in this region.

from the COBRA's GPS log. Positive TOA values indicate that the sound source is closer in range to buoy 2, which lies north of buoy 1, and will be interpreted as a "northerly" location in Fig. 8.

While sperm whales were not visually sighted until 16:21, the raw detection data indicate that sperm whale clicks were detected on the southern buoy 1 by 11:41:26. After buoy 2 had been deployed at 12:40, consistent TOA measurements become possible and two whales are detected near buoy 1 (TOA of −2 to −3 s in Fig. 9), passing south of the buoy at 13:45, since only locations south of buoy 1 could

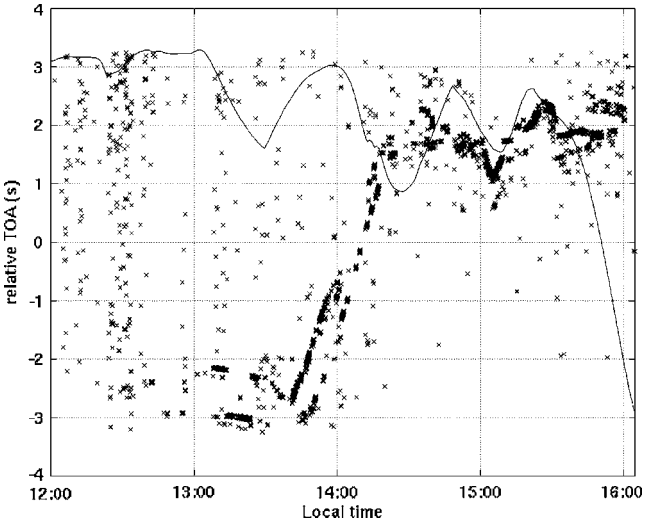


FIG. 9. Relative time-of-arrivals (TOA) of direct-path sperm whale clicks on buoy 1 relative to buoy 2. A positive TOA is defined as a signal that arrives on buoy 2 before buoy 1, i.e., a "northerly" bearing. The solid black line is the computed TOA for the fishing vessel.

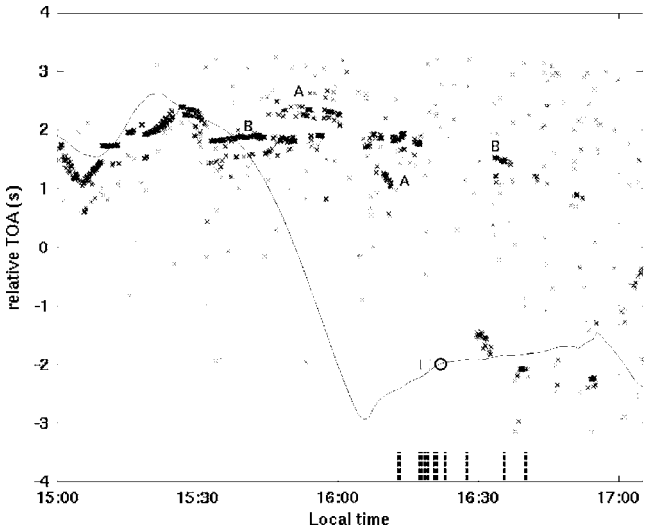


FIG. 10. Same as Fig. 9, but for times between 15:00 and 17:00. Vertical hashed lines represent deliberate cycling on the engine, with the square representing the first test, and the circle representing the first visual sighting of a whale next to the vessel. The gap in TOA activity seem between 16:00 and 16:10 is due to masking of sperm whale sounds by boat engine noise.

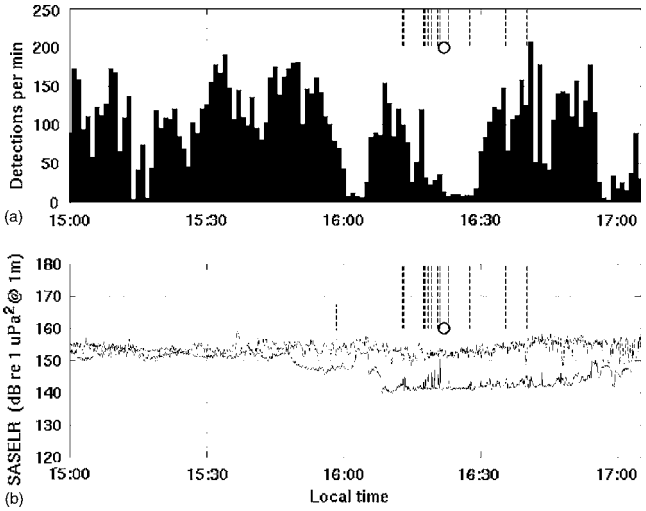


FIG. 11. 15 May 2005, test of engine cycling as an acoustic cue. (a) Sperm whale click detections per minute vs time—note the gap from 16:00 to 16:05 reflects masking of the sperm whale signals by vessel noise, not lack of acoustic activity. Dashed lines represent engine cycling events, while the black circle indicates time of first visual sighting of sperm whale 20 m from vessel; (b) SASELR corrected for vessel range, computed over the 250–1000-Hz frequency band for buoy 1 (black line) and buoy 2 (dashed line). The latter curve is substantially contaminated by acoustic "knocking" on the hydrophone.

produce TOA values of -3. The animals moved north at an estimated 2 m/s (4.7 km in 37 min) toward the vessel location, until the vessel and animal TOA merge at 14:50. From that time to 15:30, the whales' and vessel's TOA mirror each other, suggesting that the whales were either following or somehow coordinating their movements with the fishing vessel as it conducted its bathymetry survey, located at the label "Survey 15:30" in Fig. 8.

Figure 10 shows the TOA data from 15:00 through 17:00, and Fig. 11 shows a single-hydrophone analysis from buoy 1 over the same time period, viewed in terms of pulse count and COBRA SASELR. Unfortunately, there were substantial impulsive "knocking" sounds on buoy 2 that precluded automated detection analysis in Fig. 11(a), and substantially contaminates the SASELR curve for buoy 2 in Fig. 11(b).

At 15:00 the COBRA was still north of the set, traveling in a large circle mapping bathymetry, but shortly thereafter it left the area and traveled rapidly south to buoy 1. By 16:04 the vessel has passed close to buoy 1 at 6 knots, generating substantial acoustic noise, as can be seen from the SASELR plot in Fig. 11(b). Note that when the vessel was underway at full speed, its adjusted SASELR is 25 dB above background levels in the 250-Hz to 1-kHz range. The noise is sufficiently intense to mask sperm whale click detections over the time period from 16:00 to 16:05 in the pulse detection histograms in Fig. 11(a) and the TOA detections in Fig. 10. The reason that the absence of detections is known to be due to masking, and not absence of whale activity, is that sperm whale activity is still detected on buoy 2 during this time.

Once the COBRA's engine had been set to neutral by 16:08, the background noise subsided and buoy 1 detected sperm whale activity again. As the deliberate engine cycling began, all sperm whale acoustic activity ceased on both stations, as shown in Figs. 10 and 11(a). The commencement of the engine cycling can be seen as vertical lines in Fig. 10 and as 6–10-dB spikes in the SASELR curve in Figs. 11(a) and 11(b). The visual sighting of a whale next to the vessel is shown as a circle in both figures, and fluke shots confirm the presence of two individuals after this time. These whales must be different than the acoustic active whales near buoy 2 in Figs. 9 and 10, and thus four whales were in the area by this time. At the same time two additional animals were floating next to the vessel, with both animals diving at 16:29 and 16:33. The dive cycles were short—7 min or less, and the two animals' dive cycles are apparently staggered—thus, there are no clear gaps of silence in the pulse detection record.

The key observations from this encounter can be summarized as follows:

(1) Between 13:00 and 15:30 two whales traveled at least 5 km from the south and mirrored the fishing vessel's movements, bypassing the gear deployment.
(2) All sperm whale acoustic activity tapered off during the engine cycling test, as is visible in Figs. 10 and 11.
(3) Two initially nonvocalizing animals surfaced next to the vessel within 10 min of starting the engine cycling.

IV. DISCUSSION

The three encounters described above provide cumulative insight into what sperm whales do and do not respond to with regard to acoustic cues. Below we review the list of potential acoustic cues and summarize the evidence for and against each candidate.

A. Hydraulics

Before this study began, the narrow-band acoustic tones produced by the hydraulics were a popular candidate for a distinctive acoustic cue that could be exploited by whales, as the hydraulic system for the winch is typically never activated until shortly before a haul begins, and would thus be a distinctive signature. Indeed, as Fig. 3 illustrates, the signal can clearly be detected underwater when a vessel is 100 m away in calm ocean conditions.

During all of our actual acoustic encounters, however, no hydraulic signature was ever detected through either the automated SASELR computations or direct monitoring of the acoustic data from instruments as close as 1-km range, even though the flow noise levels of the instruments between 100–200 Hz were sufficiently low to presumably permit such detection. Furthermore, the activation of the hydraulic system without engaging the engine in Sec. III D prompted no apparent changes in the acoustic pulse rate or dive cycle of the lone sperm whale in the anchorline vicinity (Fig. 7). Finally, during the activity recorded in Sec. III E no hydraulic systems were activated until after 18:00, yet sperm whale positions are clearly mirroring the vessel's movements before then, and whales were visually sighted next to the vessel by 16:20. These combined observations suggest that the vessel hydraulic system is not a primary acoustic cue for attracting sperm whale attention.

B. Fishing gear strum and echosounders

Another hypothesis for an acoustic cue is that longline fishing gear would produce an acoustic signal as it "strums" while hauled under tension. Once again, direct monitoring of the acoustic record for all deployments indicates no evidence of a distinctive acoustic waterborne signature generated by fishing gear under tension. However, if the gear were producing sounds at very low frequencies, say 50 Hz or below, it is conceivable that such a signal could be buried in the flow noise recorded on the instrument. However, during at least two encounters in Secs. III D and III E, sperm whale reactions to the fishing vessel were noted even when the longline was not being hauled; indeed, the vessel was at least 1–2 km distant from the closest surface expression of the gear in both cases. Thus, acoustic strum from fishing gear seems to be an unlikely candidate for an acoustic cue for these encounters.

Similarly, echosounder signals are generally not detected in the data unless the vessel is less than 50 m from the vessel. Statistical analysis by the SEASWAP project has found no difference in encounter rate between vessels with and without echosounders. Since the frequency range of most echosounders lies in the kilohertz range, they would not be expended to propagate great distances compared to the other acoustic cues discussed here.

C. Birds and other visual cues

A large concern throughout this effort was distinguishing acoustic cues from potential visual cues such as the arrival of seabirds scavenging on the fishing haul. During a haul hundreds of birds can surround the vessel, including the northern fulmar (*Fulmarus glacialis*), the black-footed albatross (*Phoebastria nigripes*), and various species of gulls. In principle these bird flocks could be visually detected miles away. While the visual acuity of dolphins is excellent,[49] little is known about the visual capabilities of the sperm whale above water.

Fortunately, birds were not a significant confounding factor in two of the three encounters above. All the measurements in Sec. III D took place without the presence of birds as no longline was actually hauled, and most of the observations were recorded at night. Also, in Sec. III E the visual observers noted whales surfacing by the vessel at least 15 min before more than three birds had circled and settled by the COBRA. Thus, birds cannot be discounted as a potential visual cue for the animals, but in the context of these observations they were not a significant cue.

D. Cavitation noise from propeller

From the first acoustic observations conducted in 2004, it was apparent that the cavitation noise generated by changes in the propeller rotation speed produced a significant broadband acoustic signature that could be detected kilometers away. These changes occur via engaging the engine from neutral, or to a lesser extent via changes in vessel shaft speed. In all three encounters documented here, the act of engaging the propeller from a neutral state increased the SASELR by 6–10 dB between 250 to 1000 Hz and produced a detectable signal on a single hydrophone from 1 to nearly 2-km range, with a signal-to-noise ratio (SNR) of at least 6–10 dB. Even in a spherical-spreading environment, a worst-case propagation scenario, a signal with 10-dB SNR at 1 km would propagate 3 km before it merges with the measured ambient background noise spectrum. Measurements on buoy 2 from Sec. III E suggest that the cavitation signals do not propagate further than 5-km range in 600–700-m-deep water.

The act of engaging and disengaging the ship's propeller provides a distinctive acoustic cue for a longline haul, and sperm whale acoustic activity seems to alter in response to these cues. As Figs. 4(d) and 11(b) illustrate, the fact that a hauling vessel needs to engage and disengage its engine frequently makes a distinctive mark in the received acoustic data between 250 and 1 kHz. In the encounter described in Sec. III C, after an initial period of silence large amounts of sperm whale acoustic activity were detected 15 min after these signatures began, and derived acoustic tracks in Fig. 6 reveal that at least one whale was converging on the vessel location within 15 min of the start of the engine cycling activity. In Sec. III E deliberate engine cycling was associated with a complete cessation in acoustic activity from two sperm whales, and 4 min after the cycling began a third whale surfaced within 50 m of the COBRA. At this time the COBRA was over 1 km from the nearest anchorline spar buoy (anchor 1).

Section III C also suggests that engaging the engine to move the vessel from a drifting state produces an acoustic signature that is perceptually significant to sperm whales. The only observed disruptions in the animal's dive cycle that night, as inferred from the acoustic activity record, took place when the drifting vessel engaged its engines and traveled back to an instrumented anchorline. While the effect of the vessel's lights cannot be discounted, the lights should have been visible at a range of 1.5 km and thus would have been present as a constant stimulus the entire night. The TOA trajectories of the two whales before 15:30 in Sec. III E also suggest that vessel engine noise is sufficient to attract whales that are at least 5 km away.

V. CONCLUSION

Beginning with passive observation and then advancing to hypothesis testing, acoustic monitoring of depredating sperm whales off Sitka has gathered evidence that cavitation noise arising from the ship's propeller is the best candidate for a distinctive acoustic cue that causes changes in the behavior of sperm whales in the area, and hydraulic system and fishing gear signatures have at most a secondary role. In particular, the tendency of vessels to cycle their engine as they conduct a haul produces a distinctive signature that is projected to propagate 4–8 km under the conditions measured here, and this signature is associated with the interruption of sperm whale acoustic activity, the convergence of animals toward the vessel, and the surfacing of animals next to the vessel. Whales also seem to respond to situations when a vessel is transitioning from drifting to transiting.

A natural question to ask would be whether knowledge of acoustic cues could be practically applied to reduce depredation encounters with vessels. Given the well-known ability of marine mammals to habituate quickly to sounds intended to discourage depredation (e.g., Refs. 50 and 51), it would be easy to conclude that any change in fishing activity strategy to alter acoustic cues would be a temporary situation at best.

However, knowledge of acoustic cues opens up a variety of strategies, including reducing cue detection range, evaluating whether passive acoustic monitoring for sperm whales from fishing vessels is a viable avoidance measure, and faking cues to decouple the association of a cue with fishing activity. Even a set of actions that causes a delay in the response time of the animals can help reduce losses.

An animal cannot react to an acoustic cue that it does not hear, so any activity that reduces the intensity of a distinctive sound will reduce the volume of water over which an animal can detect a cue. A signal reduction of 6 dB translates into a factor of 4 reduction in intensity, or a halving of the detection radius under spherical spreading conditions, and greater reductions in less attenuating environments. Thus, reducing noise levels would potentially reduce the number of animals detecting the sound. Local fishermen have been advised not to linger in an area where gear has been deployed, and particularly not to drift in the same area as a haul, as well as to conduct "circle hauls" or other techniques that minimize the number of times engines need to be disengaged while fishing.

Figure 11(b) in Sec. III E also shows that the acoustic signature of a vessel is difficult to extract from a receiver 100–200-m depth at 5-km range, while the TOA plots in Figs. 9 and 10 demonstrate that sperm whale acoustic signals can propagate beyond that distance, a result consistent with previous observations of sperm whale detection distance during acoustic surveys.[52] There is thus a possibility that fishing vessels could acoustically monitor an area for the presence of sperm whales before deploying or retrieving gear. Practical experience in deploying cabled hydrophones indicates that, if an HTI-96 min hydrophone can be dropped to at least 20-m depth underneath an idling vessel, sperm whales can be detected to at least a couple of kilometers range in Beaufort 3 conditions. Further work would be needed to determine whether a fishing vessel could detect a sperm whale at a greater distance than a sperm whale could detect a fishing vessel.

Knowledge of acoustic cues also raises the possibility that they can be faked, thus introducing an element of risk in a whale's decision to expend time and energy investigating a cue. For example, at present if a sperm whale hears engine cycling from a fishing vessel, it is almost guaranteed to encounter a haul if it responds, which apparently more than compensates for the energy loss sustained in traveling to the site, and the opportunity cost of forgoing natural foraging activity during that transit time. From a game-theoretical perspective this is an optimal strategy,[53] and from the perspective of behavioral theory depredation behavior is strongly positively reinforced. Even if the cue changes, but still unambiguously indicates the presence of fishing activity, the animal would quickly habituate and continue its behavior.

Suppose, however, that the acoustic cue remains the same, but the consequences of responding to that decision are altered. For example, what if fishing vessels make a habit of cycling their engines at random, or cycling engines around "decoy" anchorlines? In this situation the presence of the acoustic cue no longer guarantees an encounter with a longline haul, and the whale faces a potential energy loss when responding to the action. From a game-theoretical perspective,[53] if the whale is a "rational" decision maker then one can see how a widespread adoption of faking cues by a fishing fleet might eventually disassociate the cue from an actual haul, if the negative consequences of responding to a cue, in terms of lost time and effort, are large enough. Of course whales, like people, may not be rational decision makers. In comparative psychology it has been long noted that intermittent schedules of reinforcement can condition stronger behavioral responses than a consistent reward schedule.[54,55] However, the effectiveness of "extinction" or "negative reinforcement" in deconditioning undesirable behaviors has also been well-documented in the same literature. The key unknown factor is what opportunity cost an animal faces when responding to a faked cue. If there is little to no "punishment" in terms of lost time and energy when responding to a faked cue, then from both a game theory and learning theory perspective the behavior may continue and even strengthen. However, if the cumulative punishment accrued from lost feeding opportunities were large enough, then the conditions for an extinction or negative reinforcement learning model might exist.

Thus, the ability of animals to habituate quickly to changes in acoustic stimuli does not negate the importance of identifying acoustic cues that attract the animals, because efforts to reduce the detection range of these cues and to produce "false cues" might be effective long-term strategies in reducing depredation, or at least delaying the response of animals to fishing activity.

ACKNOWLEDGMENTS

The North Pacific Research Board supported this research under Projects F0412, R0309, F0527, and F0626. The cooperation of fishermen with the Alaska Longline Fishermen's Association was fundamental to the effort, with special recognition to Jay Skordahl, Steve Weissberg, Carter Hughes, John Petraborg, and Linda Behnken. Jen Cedarleaf provided Sitka-based logistical assistance and bird identification, and student interns Nellie Warner, Morgan Hartley, and Patrica Ramon helped collect field data. Valeria Teloni provided essential background on sperm whale literature, and Robert Gisiner provided background on behavioral conditioning.

[1]H. Whitehead, "Estimates of the current global population size and historical trajectory for sperm whales," Mar. Ecol.: Prog. Ser. **242**, 295–304 (2002).
[2]T. Lyrholm and U. Gyllensten, "Global matrilineal population structure in sperm whales as indicated by mitochondrial DNA sequences," Proc. R. Soc. London, Ser. B **265**(1406), 1679–1684 (1998).
[3]Y. Cherel and G. Duhamel, "Antarctic jaws: Cephalopod prey of sharks in

Kerguelen waters," Deep-Sea Res., Part I **51**(1), 17–31 (2004).
[4]A Berzin, *Kashalot (The Sperm Whale)*. (U.S. Dept of Commerce, National Technical Information Service, Springfield, VA, 1971).
[5]M. E. Gosho, D. W. Rice, and J. M. Breiwick, "The sperm whale, *Physeter macrocephalus*," Mar. Fish. Rev. **46**(4), 54–64 (1984).
[6]D. W. Rice, in *Handbook of Marine Mammals*, edited by S. H. Ridgway and R. Harrison (Academic, London, 1989), Vol. **4**, pp. 177–233.
[7]D. K. Mellinger, K. M. Stafford, and C. G. Fox, "Seasonal occurrence of sperm whale (Physeter macrocephalus) sounds in the Gulf of Alaska, 1999–2001," Marine Mammal Sci. **20**(1), 48–62 (2004).
[8]P. B. Best, in *Behavior of Marine Animals*, edited by H. E. Winn and B. L. Olla (Plenum, New York, 1979), Vol. **3**, pp. 227–289.
[9]D. K. Caldwell, M. C. Caldwell, and D. W. Rice, in *Whales, Dolphins, and Porpoises*, edited by K. S. Norris (University of Califronia Press, Berkeley, 1966), pp. 677–717.
[10]N. Jaquet, "How spatial and temporal scales influence understanding of Sperm Whale distribution: A review," Mammal Rev. **26**(1), 51–65 (1996).
[11]H. Whitehead, M. Dillon, S. Dufault, L. Weilgart, and J. Wright, "Non-geographically based population structure of South Pacific sperm whales: Dialects, fluke-markings and genetics," J. Anim. Ecol. **67**(2), 253–262 (1998).
[12]T. Lyrholm, O. Leimar, B. Johanneson, and U. Gyllensten, "Sex-biased dispersal in sperm whales: contrasting mitochondrial and nuclear genetic structure of global populations," Proc. R. Soc. London, Ser. B **266**(1417), 347–354 (1999).
[13]M. B. Santos, G. J. Pierce, P. R. Boyle, R. J. Reid, H. M. Ross, I. A. P. Patterson, C. C. Kinze, S. Tougaard, R. Lick, U. Piatkowski, and V. Hernandez-Garcia, "Stomach contents of sperm whales *Physeter macrocephalus* stranded in the North Sea 1990–1996," Mar. Ecol.: Prog. Ser. **183**, 281–294 (1999).
[14]K. Evans and M. A. Hindell, "The diet of sperm whales (*Physeter macrocephalus*) in Southern Australian waters," ICES J. Mar. Sci. **61**(8), 1313–1329 (2004).
[15]T Kawakami, "A review of sperm whale food," Sci. Rep. Whales Res. Inst. **32**, 199–218 (1980).
[16]K. M. Fristrup and G. R. Harbison, "How do sperm whales catch squids?," Marine Mammal Sci. **18**(1), 42–54 (2002).
[17]K. Das, G. Lepoint, Y. Leroy, and J. M. Bouquegneau, "Marine mammals from the Southern North Sea: Feeding ecology data from delta C-13 and delta N-15 measurements," Mar. Ecol.: Prog. Ser. **263**, 287–298 (2003).
[18]H. Whitehead, *Sperm Whales: Social Evolution in the Ocean* (University of Chicago Press, Chicago, 2003).
[19]T. Okutani and T. Nemoto, "Squids as the food of sperm whales in the Bering Sea and Alaska Gulf," Tokai Regional Fisheries Laboratory, Report No. 18, 1964.
[20]M. R. Clarke and N. Macleod, "Cephalopod remains from sperm whales caught off Iceland," J. Mar. Biol. Assoc. U.K. **56**, 733–750 (1976).
[21]C. P. Nolan and G. M. Liddle, "Interactions between killer whales (*Orcinus orca*) and sperm whales (*Physeter macrocephalus*) with a longline fishing vessel," Marine Mammal Sci. **16**(3), 658–664 (2000).
[22]M. G. Purves, D. J. Agnew, E. Balguerias, C. A. Moreno, and B. Watkins, "Killer whale (*Orcinus orca*) and sperm whale (*Physeter macrocephalus*) interactions with longline vessels in the patagonian toothfish fishery at South Georgia, South Atlantic," Ccamlr Sci. **11**, 111–126 (2004).
[23]R. Hucke-Gaete, C. A. Moreno, J. Arata, and Blue Whale Ctr, "Operational interactions of sperm whales and killer whales with the Patagonian toothfish industrial fishery off Southern Chile," Ccamlr Sci. **11**, 127–40 (2004).
[24]E. A. Zollett and A. J. Read, "Depredation of catch by bottlenose dolphins (*Tursiops truncatus*) in the Florida king mackerel (*Scomberomorus cavalla*) troll fishery," Fish. Bull. **104**, 343–349 (2006).
[25]J. R. Ashford, P. S. Rubilar, and A. R. Martin, "Interactions between cetaceans and longline fishery operations around South Georgia," Marine Mammal Sci. **12**(3), 452–457 (1996).
[26]D. Capdeville, "Interaction of marine mammals with the longline fishery around the Kerguelen Islands (Division 58.5.1) during the 1995/96 cruise," Ccamlr Sci. **4**, 171–174 (1997).
[27]E. F. González, *presented at the XXI Congreso de Ciencias del Mar*, Chile, 2001.
[28]P. S. Hill, J. L. Laake, and E. Mitchell, "Results of a pilot program to document interactions between sperm whales and longline vessels in Alaska waters," U.S Department of Commerce, Report No. NOAA TM-NMFS-AFSC-108, 1999.
[29]M. F. Sigler, C. R. Lunsford, and J. M. Straley Marine Mammal Sci. (in press).
[30]G. R. McPherson, P. Turner, C. McPherson, and D. Cato, "Predation of large marine mammals (family *Delphinidae*) on longline and dropline target species. Phase I. Pilot study of the acoustic mechanism of predation, and development of a three dimensional acoustic tracking system." Australian Fisheries Management Authority and Eastern Tuna Management Advisory Committee, Department of Primary Industries, Queensland., Report No. QI 02105, 2002.
[31]G. R. McPherson, C. Clague, P. Turner, C. R. McPherson, A. Madry, I. Bedwell, and D. H. Cato, *presented at the Proceedings of Acoustics 2004*, Gold Coast, Australia, 2004.
[32]J. C. Goold and S. E. Jones, "Time and frequency-domain characteristics of sperm whale clicks," J. Acoust. Soc. Am. **98**(3), 1279–1291 (1995).
[33]L. V. Worthington and W. E. Schevill, "Underwater sounds heard from sperm whales," Nature (London) **180**, 291 (1957).
[34]W. A. Watkins, "Acoustic behaviors of sperm whales," Oceanus **20**, 50–58 (1977).
[35]L. A. Douglas, S. M. Dawson, and N. Jaquet, "Click rates and silences of sperm whales at Kaikoura, New Zealand," J. Acoust. Soc. Am. **118**(1), 523–529 (2005).
[36]A. Thode, D. K. Mellinger, S. Stienessen, A. Martinez, and K. Mullin, "Depth-dependent acoustic features of diving sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico," J. Acoust. Soc. Am. **112**(1), 308–321 (2002).
[37]W. M. X. Zimmer, M. P. Johnson, A. D'Amico, and P. L. Tyack, "Combining data from a multisensor tag and passive sonar to determine the diving behavior of a sperm whale (*Physeter macrocephalus*)," IEEE J. Ocean. Eng. **28**(1), 13–28 (2003).
[38]A. Thode, C. Tiemann, J. Straley, K. Folkert, and V. O'Connell, "Three-dimensional localization of sperm whales using a single hydrophone," J. Acoust. Soc. Am. **120**(4), 2355–2365 (2006).
[39]W. C. Burgess, "The bioacoustic probe: A general-purpose acoustic recording tag (A)," J. Acoust. Soc. Am. **108** (5, Pt. 2), 2583 (2000).
[40]D. K. Mellinger, "ISHMAEL 1.0 User's Guide," NOAA/PMEL Tech. Mem., Report No. PMEL-120, 2002.
[41]A. Thode, "Three-dimensional passive acoustic tracking of sperm whales (*Physeter macrocephalus*) in ray-refracting environments," J. Acoust. Soc. Am. **18**(6), 3575–3584 (2005).
[42]W. M. Carey, "Sound sources and levels in the ocean," IEEE J. Ocean. Eng. **31**(1), 61–75 (2006).
[43]National Research Council Ocean Studies Board, *Ocean Noise and Marine Mammals* National Academies, Washington, D.C., 2003.
[44]P. T. Madsen, "Marine mammals and noise: Problems with root mean square sound pressure levels for transients," J. Acoust. Soc. Am. **117**(6), 3952–3957 (2005).
[45]A. Thode, "Tracking sperm whale (*Physeter macrocephalus*) dive profiles using a towed passive acoustic array," J. Acoust. Soc. Am. **116**(1), 245–253 (2004).
[46]J. L. Spiesberger, "Geometry of locating sounds from differences in travel time: Isodiachrons," J. Acoust. Soc. Am. **116**(5), 3168–3177 (2004).
[47]M. Wahlberg, "The acoustic behaviour of diving sperm whales observed with a hydrophone array," J. Exp. Mar. Biol. Ecol. **281**(1–2), 53–62 (2002).
[48]S. Watwood, P. J. O. Miller, M. Johnson, P. T. Madsen, and P. Tyack, "Deep-diving foraging behaviour of sperm whales," J. Anim. Ecol. **75**(3), 814–825 (2006).
[49]R. L. Pepper, Jr., and V. Simmons, "In-air visual acuity of the bottlenose dolphin," Exp. Neurol. **41**(2), 271–276 (1973).
[50]D. A. Hanan, L. M. Jones, and R. B. Read, "California sea lion interaction and depredation rates with the commercial passenger fishing vessel fleet near San Diego," CalCOFI Report 30, 122–126 (1989).
[51]T. A. Jefferson and B. E. Curry, "Acoustic methods of reducing or eliminating marine mammal-fishery interactions: Do they work?," Ocean Coastal Manage. **31**(1), 41–70 (1996).
[52]J. Barlow and B. L. Taylor, "Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey," Marine Mammal Sci. **21** (3), 429–445 (2005).
[53]J. W. Bradbury and S. L. Vehrencamp, *Principles of Animal Communication* (Sinauer Associates, Sunderland, MA, 1998).
[54]John Hall, *Classical Conditioning and Instrumental Learning: A Contemporary Approach* (JB Lippincott, Philadelphia, 1976).
[55]http://www.animalbehaviour.net/PosReinforcement.htm, reviewed 15 May 2007.